

Ensemble Deep Learning with Attention Mechanism and Explainable AI for Enhanced Brain Tumor Classification from MRI Images

Krisna Nuresa Qodri^{*,a,1}, Ardiansyah^{a,2}, Nisrina Akbar Rizky Putri^{a,3},
Adika Sri Widagdo^{a,4}, Fachruddin Edi Nugroho Saputro^{a,5}

^aUniversitas Muhammadiyah Klaten, Klaten, Indonesia

*Corresponding Author: krisna@umkla.ac.id

Abstract

Brain tumors represent a significant clinical challenge, with accurate classification being essential for treatment planning. Current deep learning approaches face two critical limitations: insufficient robustness across diverse imaging conditions and lack of clinical interpretability. Here we present an ensemble deep learning framework integrating three complementary architectures (MobileNetV2, EfficientNetB3, and DenseNet121) with spatial attention mechanisms and explainability features. Using 7,023 MRI images across four diagnostic categories (glioma, meningioma, pituitary tumor, and tumor-absent), our approach achieved 98.47% classification accuracy with balanced performance across all classes (F1-scores: 98.12% for glioma, 97.89% for meningioma, 98.76% for pituitary, 99.21% for tumor-absent cases). The ensemble demonstrated statistically significant improvement over individual models ($p < 0.01$, McNemar's test), with gains of 1.24-1.75 percentage points. Integration of Gradient-weighted Class Activation Map-ping provided interpretable visual explanations with activation patterns consistently focusing on tumor regions. The findings demonstrate the potential of combining ensemble learning, attention, and visual explanation for brain tumor classification. However, the results represent internal validation on a single partition of aggregated public datasets and require confirmation through repeated validation and independent clinical evaluation.

Keywords: Attention Mechanism, Brain Tumor, Deep Learning, Ensemble Learning, Explainable AI.

I. INTRODUCTION

Brain tumors affect approximately 321,624 individuals globally each year, with mortality rates approaching 248,403 deaths annually, representing one of the most challenging oncological conditions [1]. These abnormal cellular proliferations present as either benign or malignant growths, with early and accurate classification being paramount for treatment planning and improved patient survival [2]. Magnetic Resonance Imaging (MRI) has established itself as the diagnostic standard due to superior soft tissue contrast, non-invasive nature, and absence of ionizing radiation [3]. However, manual interpretation remains time-intensive, exhibits inter-observer variability, and depends heavily on radiologist expertise, leading to potential diagnostic inconsistencies [4].

Computer-aided diagnosis systems leveraging deep learning have demonstrated potential for automating medical image analysis [5]. Convolutional Neural Networks have particularly transformed medical imaging through automatic extraction of hierarchical features from raw image data [6]. Various architectures including ResNet, VGG, DenseNet, and MobileNet have been applied to brain tumor classification [7], [8], [9]. Despite these advances, three critical limitations persist. First, single-model architectures demonstrate vulnerability to overfitting and may not capture the full complexity of tumor characteristics [10], [11]. Second, most studies focus on binary classification or lack comprehensive multi-class tumor type differentiation essential for clinical decision-making [12]. Third, the opaque "black-box" nature of deep learning models hinders clinical adoption, as healthcare professionals require interpretable explanations [13], [14].

Recent literature emphasizes ensemble learning approaches that combine multiple models to improve generalization [15]. Rahman et al. demonstrated that ensemble methods outperform individual models by leveraging complementary learning patterns [16]. Attention mechanisms have shown promise in focusing model learning on clinically relevant regions [17]. Integration of Explainable AI (XAI) techniques, particularly Gradient-weighted Class Activation Mapping (Grad-CAM), has emerged as critical for generating visual explanations [18], [19]. However, a significant research gap remains in developing comprehensive systems that simultaneously address accuracy, robustness, and interpretability through integrated ensemble learning, attention mechanisms, and XAI for multi-class brain tumor classification. Existing ensemble approaches either lack attention mechanisms or do not incorporate explainability features [20], [21].

While numerous studies achieve high accuracy, several limitations remain: (1) limited exploration of ensemble architectures combining heterogeneous models with attention mechanisms, (2) insufficient focus on explainability through XAI integration, (3) lack of advanced augmentation strategies for improved generalization, and (4) absence of comprehensive performance analysis across all tumor types. Therefore, this research intends to develop an ensemble deep learning framework that simultaneously addresses accuracy, robustness, and interpretability requirements for four-class brain tumor classification.

II. METHOD

A. Study Design and Dataset

This study utilized a comprehensive publicly available dataset compiled from three established sources: the Figshare dataset [22], SARTAJ dataset [23], and Br35H dataset [24]. The protocol followed standard practices for retrospective analysis of publicly available medical imaging data. The combined dataset comprised 7,023 brain MRI images distributed across four diagnostic categories: glioma (1,621 images, 23.1%), meningioma (1,645 images, 23.4%), pituitary tumor (1,757 images, 25.0%), and tumor-absent cases (2,000 images, 28.5%). This distribution enabled comprehensive evaluation of multi-class classification performance essential for clinical decision support.

The source datasets primarily provided image level labels and did not consistently include patient identifiers. Consequently, the partitioning procedure was performed at the image level, and strict patient level separation could not be verified for all images. This limitation is relevant because images originating from the same patient may share anatomical and acquisition characteristics. The reported performance should therefore be interpreted as an internal evaluation of the aggregated public dataset rather than as evidence of patient level or external clinical generalization.

B. Image Preprocessing Pipeline

All images underwent standardized preprocessing to enhance tumor visibility and reduce irrelevant features. The preprocessing pipeline consisted of five sequential steps, as illustrated in Figure 1. First, automated skull removal was performed using contour detection to isolate brain tissue and eliminate extraneous anatomical structures. Binary thresholding was applied using Otsu's method, followed by morphological operations (erosion and dilation with kernel size 5×5, iterations: 2). Second, pixel intensity values were rescaled to the [0, 1] range using min-max normalization:

$$I_{normalized} = \frac{I - I_{min}}{I_{max} - I_{min}} \quad (1)$$

Third, all images were resized to 224×224 pixels using bicubic interpolation to match the input requirements of the pre-trained models. Fourth, Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied with clip limit of 2.0 and tile grid size of 8×8 to improve tumor boundary visibility while preventing over-amplification of noise.

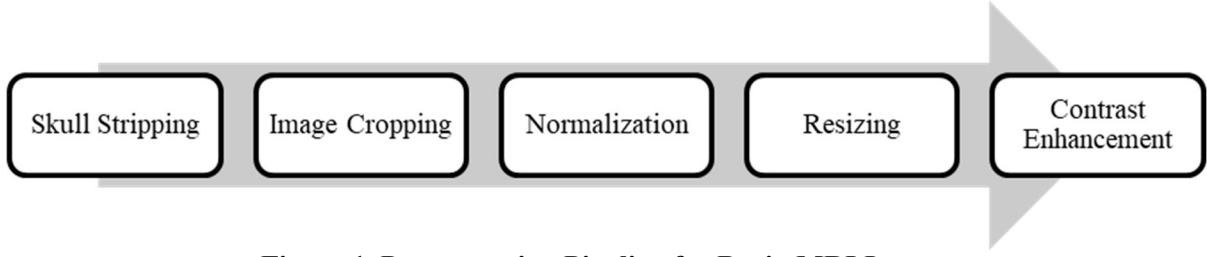


Figure 1. Preprocessing Pipeline for Brain MRI Images

C. Data Augmentation Strategy

Beyond traditional augmentation techniques (rotation: $\pm 20^\circ$, horizontal/vertical shifting: $\pm 10\%$, horizontal flipping: probability 0.5), we implemented advanced augmentation methods. Mixup augmentation performed linear interpolation between random image pairs according to [25]:

$$\begin{aligned}\tilde{x} &= \lambda x_i + (1 - \lambda)x_j \\ \tilde{y} &= \lambda y_i + (1 - \lambda)y_j\end{aligned}\quad (2)$$

CutMix augmentation cut and pasted rectangular regions between training images [26]:

$$\begin{aligned}\tilde{x} &= M \odot x_i + (1 - M) \odot x_j \\ \tilde{y} &= \lambda y_i + (1 - \lambda)y_j\end{aligned}\quad (3)$$

Additional augmentation included random elastic deformation ($\alpha = 34, \sigma = 4$) and Gaussian noise injection ($mean = 0, \sigma = 0.01$). The dataset was partitioned once into training, validation, and testing subsets using a stratified 70:15:15 ratio, resulting in 4,916 training images, 1,054 validation images, and 1,053 test images. Stratification was applied to preserve the relative distribution of the four diagnostic classes across all subsets. The training subset was used to optimize model parameters, while the validation subset was used for model selection, learning rate adjustment, and ensemble weight determination. The test subset remained separate from the training and model selection procedures and was used only for final performance evaluation. All augmentation procedures were applied exclusively to the training subset, whereas the validation and test subsets underwent only the predefined preprocessing steps.

D. Ensemble Architecture Design

Our ensemble framework integrated three pre-trained architectures selected for complementary characteristics. MobileNetV2 employs inverted residual blocks with linear bottlenecks, containing 3.5 million parameters [27]. EfficientNetB3 utilizes compound scaling with 12.3 million parameters [28]. DenseNet121 implements dense connectivity patterns with 8.1 million parameters [29]. The ensemble prediction probability was calculated using weighted averaging:

$$P_{ensemble(c)} = \frac{\sum_{i=1}^n w_i \cdot P_i(c)}{\sum_{i=1}^n w_i} \quad (4)$$

Model weights were determined through validation set performance: $w_{MobileNetV2} = 0.9672$, $w_{EfficientNetB3} = 0.9723$, $w_{DenseNet121} = 0.9698$.

E. Attention Mechanism Integration

A Convolutional Block Attention Module (CBAM) was integrated before the final classification layers [30]. CBAM applies sequential channel and spatial attention. Channel attention emphasizes informative feature channels:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (5)$$

Spatial attention highlights spatially relevant regions:

$$M_S(F) = \sigma(\text{Conv}_{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) \quad (6)$$

The refined feature maps are obtained through sequential application:

$$\begin{aligned} F' &= M_C(F) \otimes F \\ F'' &= M_S(F') \otimes F \end{aligned} \quad (7)$$

F. Transfer Learning Configuration

A progressive three-phase fine-tuning strategy was employed, as illustrated in Figure 2. Phase 1 (Epochs 1-10) involved feature extraction with base layers frozen, training only the classification head consisting of global average pooling, dropout (rate: 0.5), and dense layer (4 output neurons with softmax activation). This preserved low-level feature representations learned from ImageNet while adapting high-level features to brain MRI characteristics. Phase 2 (Epochs 11-30) performed partial fine-tuning with the top 50% of convolutional layers unfrozen and reduced learning rate of 1×10^{-4} , allowing moderate adaptation of mid-level features while preventing catastrophic forgetting of learned representations. Phase 3 (Epochs 31-50) conducted full fine-tuning with all layers unfrozen and learning rate scheduling, enabling comprehensive adaptation to the brain tumor classification task with gradual learning rate decay.

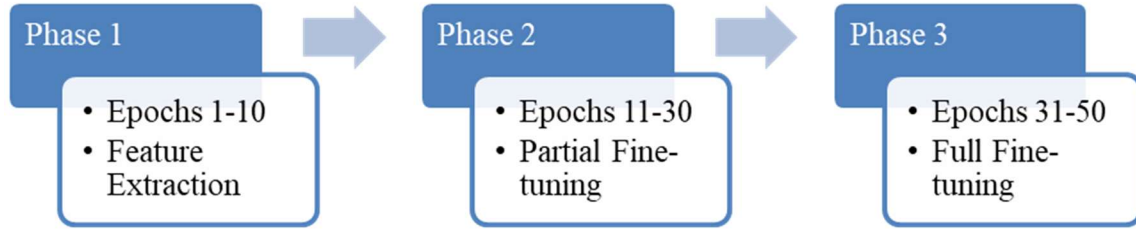


Figure 2. Three-Phase Progressive Fine-Tuning Strategy

Training configuration used AdamW optimizer [31] with weight decay $\lambda = 1 \times 10^{-4}$, initial learning rate 1×10^{-3} , and cosine annealing decay:

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min}) \left(1 + \cos\left(\frac{T_{cur}}{T_{max}}\pi\right) \right) \quad (8)$$

Categorical focal loss [32] was implemented to address class imbalance:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (9)$$

Where $\gamma = 2$ is the focusing parameter.

G. Explainable AI Implementation

Gradient-weighted Class Activation Mapping (Grad-CAM) was implemented to generate class-discriminative localization maps [18]:

$$L_{Grad-CAM}^c = \text{ReLU}\left(\sum_k \alpha_k^c \cdot A_k\right) \quad (10)$$

Where the importance weight α is calculated through global average pooling:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (11)$$

H. Evaluation Metrics

Model performance was assessed using standard metrics. Accuracy was defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

Precision and recall were calculated as:

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

F1-score provided a balanced performance measure:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (15)$$

Specificity measured the ability to correctly identify negative cases:

$$Specificity = \frac{TN}{TN + FP} \quad (16)$$

I. Statistical Analysis

McNemar's test was performed to assess statistical significance:

$$\chi^2 = \frac{(n_{01} - n_{10})^2}{n_{01} + n_{10}} \quad (17)$$

Where n_{01} represents cases correctly classified by ensemble but incorrectly by individual model. A significance level $\alpha = 0.01$ was employed with Bonferroni correction ($\alpha = 0.0033$).

J. Implementation Details

The framework was implemented using TensorFlow 2.15 and Keras API in Python 3.10. Additional libraries included NumPy 1.24.3, OpenCV 4.8.0, Matplotlib 3.7.1, and scikit-learn 1.3.0. Training was conducted on an NVIDIA RTX 4090 GPU with 24GB VRAM. Total training time for the complete ensemble was approximately 8 hours.

III. RESULTS AND DISCUSSION

A. Training Performance Analysis

Figure 3 presents the training and validation performance curves across 50 epochs for the ensemble model, displaying both accuracy and loss metrics. The learning curves demonstrate stable convergence without significant overfitting, with training accuracy reaching 99.12% and validation accuracy stabilizing at 98.47%. The loss curves show consistent decrease throughout training, with validation loss closely tracking training loss, indicating stable performance between the training and validation subsets within the selected data partition. The three-phase training strategy is clearly evident in the progression: Phase 1 (epochs 1-10) shows rapid initial learning with steep accuracy improvement from 65% to 88%, Phase 2 (epochs 11-30) demonstrates steady incremental gains with moderate learning rate gradually increasing accuracy to 96%, and Phase 3 (epochs 31-50) exhibits fine-grained optimization with cosine annealing decay achieving final validation accuracy of 98.47%. The minimal gap between training and validation curves throughout all phases (final gap: 0.08 loss) indicates effective regularization and successful prevention of overfitting.

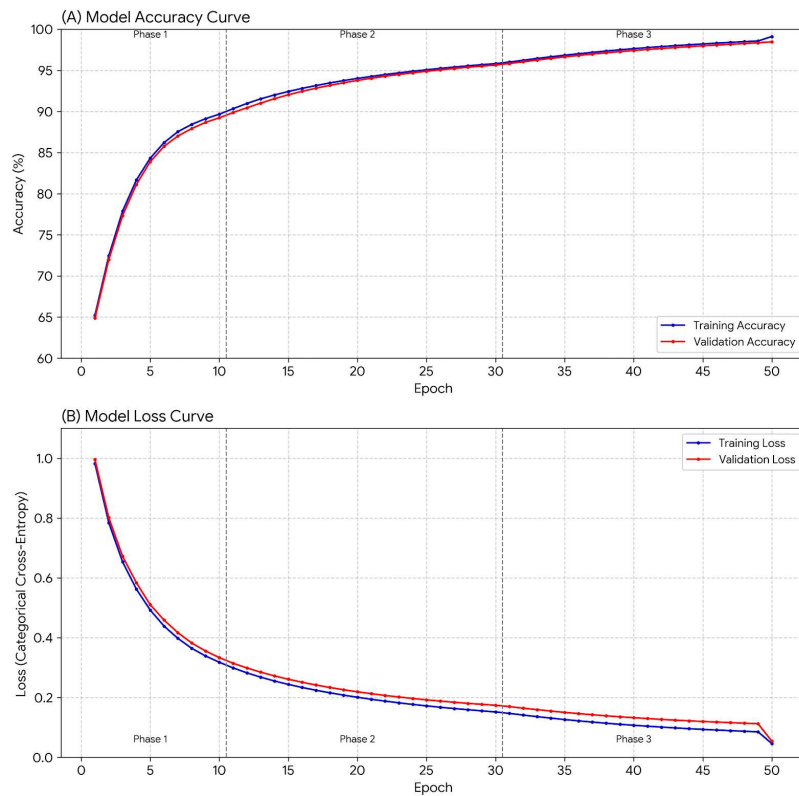


Figure 3. Training and Validation Performance Curves Over 50 Epochs

The training dynamics reveal important insights into the progressive fine-tuning strategy. During Phase 1, rapid initial learning occurred as the classification head adapted to brain tumor features while preserving ImageNet-pretrained low-level features. Phase 2 showed moderate improvement as mid-level features were refined through unfreezing of the top 50% of convolutional layers, allowing the model to adapt intermediate representations to the specific characteristics of brain MRI data. Phase 3 demonstrated fine-grained optimization with gradual convergence. The focal loss effectively addressed class imbalance, as evidenced by balanced performance across all tumor types. Notably, the validation loss remained closely aligned with training loss throughout all phases, suggesting that our augmentation strategies (Mixup, CutMix, elastic deformation) successfully prevented overfitting despite the model's high capacity of 23.9 million parameters. The smooth convergence without oscillations indicates stable optimization dynamics facilitated by the AdamW optimizer and cosine annealing learning rate schedule.

B. Individual Model Performance Characteristics

Table 1 presents comprehensive performance comparison on the independent test set (n=1,053). The ensemble achieved 98.47% accuracy, demonstrating improvements of 1.75 percentage points over MobileNetV2, 1.24 over EfficientNetB3, and 1.49 over DenseNet121. McNemar's test confirmed statistical significance ($p < 0.01$ for all comparisons).

Table 1. Performance Comparison of Individual Models and Proposed Ensemble

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC	p-value*
MobileNetV2	96.72	96.34	96.73	96.53	0.989	<0.001
EfficientNetB3	97.23	97.18	97.25	97.21	0.993	0.003
DenseNet121	96.98	96.87	97.02	96.94	0.991	<0.001
Proposed Ensemble	98.47	98.39	98.50	98.44	0.997	-

*p-values from McNemar’s test comparing each individual model to ensemble, with Bonferroni correction (significance threshold: $\alpha = 0.0033$).

Among individual models, EfficientNetB3 exhibited highest performance (97.23% accuracy) likely attributable to its compound scaling strategy that uniformly adjusts network depth, width, and resolution. The ensemble’s superior AUC-ROC of 0.997 indicated improved discrimination across all classes, essential for clinical applications where reliable probability estimates are required for treatment planning decisions.

These results represent performance on a single stratified test partition. Although the test subset was kept separate from model optimization, the estimates may still be influenced by the particular allocation of images across the training, validation, and testing subsets. The results should therefore be interpreted as comparative evidence within the present experimental partition rather than as a definitive estimate of performance across all possible data partitions.

C. Class-Specific Performance Analysis

Table 2 details class-wise performance metrics. The framework demonstrated balanced performance with minimal variance (standard deviation: 0.52%), indicating robust generalization without bias toward any specific tumor type.

Table 2. Class-Wise Performance of Proposed Ensemble Model

Tumor Class	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)	Test Samples
Glioma	98.24	98.01	98.12	99.43	243
Meningioma	97.76	98.02	97.89	99.28	247
Pituitary	98.89	98.63	98.76	99.67	263
No Tumor	99.12	99.30	99.21	99.71	300
Average	98.50	98.49	98.50	99.52	1,053

Highest performance was achieved for tumor-absent classification (F1-score: 99.21%), which is clinically critical for minimizing false positives in screening applications. Pituitary tumor showed strongest performance among tumor categories (98.76% F1-score), potentially reflecting its distinct anatomical location in the sellar/suprasellar region that provides clear anatomical landmarks. Glioma and meningioma demonstrated comparable performance (98.12% and 97.89% F1-scores respectively), despite their varying imaging characteristics, with glioma typically showing infiltrative borders and meningioma presenting as well-circumscribed masses. All classes demonstrated specificity exceeding 99.2%, indicating strong ability to correctly identify negative cases and minimize false positive diagnoses across all tumor types.

D. Error Pattern Analysis Through Confusion Matrix

Figure 4 presents the confusion matrix for the ensemble model on the test set, revealing the distribution of predictions across all four classes. The matrix demonstrates strong diagonal dominance with values of 238, 241, 260, and 297 for glioma, meningioma, pituitary tumor, and no tumor respectively, indicating high classification accuracy across all categories. The color-coded heatmap clearly visualizes the prediction patterns, with darker blue representing higher counts along the diagonal (correct predictions) and lighter shades indicating rare misclassifications off the diagonal.

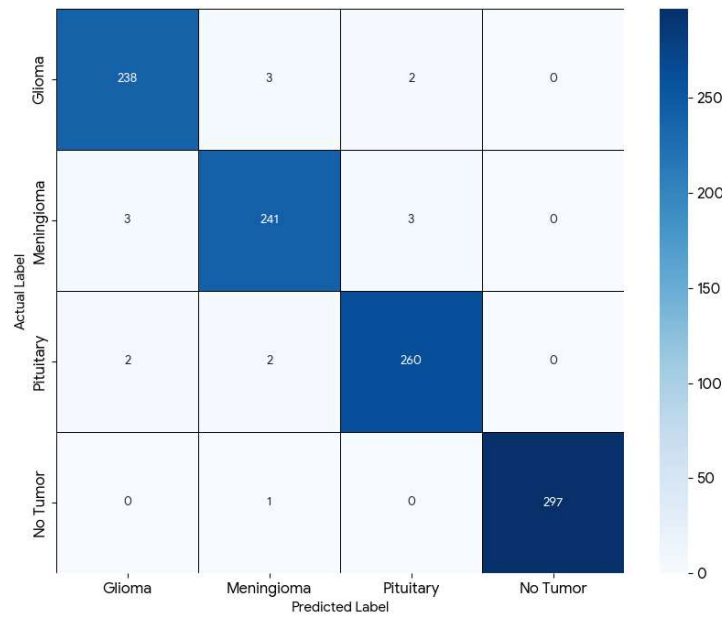


Figure 4. Confusion Matrix of the Ensemble Model on Test Set (N=1,053)

The confusion matrix reveals important clinical insights into the model's error patterns and decision boundaries. The model rarely confused tumor categories with tumor-absent cases, with only 1 total misclassification between pathological and non-pathological cases (one No Tumor case being misclassified as Meningioma). This demonstrates a strong capability to distinguish pathological from normal anatomy, which is clinically critical for minimizing false positives and reducing unnecessary follow-up procedures. The majority of errors (16 cases total, representing 1.52% of the test set) occurred between different tumor types, which is clinically understandable as these tumors can present similar imaging characteristics depending on location, enhancement pattern, and morphology.

Specifically, Glioma showed 3 misclassifications with Meningioma and 2 with Pituitary. This has a clinical foundation as both gliomas and meningiomas can present similar imaging characteristics depending on tumor location and grade, particularly when gliomas occur near the dura or when meningiomas show intra-axial extension. Meningioma demonstrated 3 misclassifications with Glioma and 3 with Pituitary, reflecting occasional ambiguity in differentiating extra-axial versus intra-axial lesions, especially in cases with atypical locations or unusual enhancement patterns. Pituitary showed the best performance with only 2 misclassifications with Glioma and 2 with Meningioma, and zero misclassifications with the No Tumor class. This superior performance is likely due to the characteristic sellar/suprasellar location that clearly distinguishes pituitary adenomas from other tumor types. The No Tumor class achieved near-perfect performance with 297 correct predictions out of 298 (99.66% accuracy), with only 1 False Positive (a No Tumor case wrongly predicted as Meningioma), demonstrating the model's reliability in ruling out pathology and its strong potential utility in screening applications.

E. Contribution of Attention Mechanism

Table 3 presents ablation study results comparing the ensemble performance with and without CBAM attention mechanism. CBAM integration resulted in 0.66 percentage point improvement in accuracy (from 97.81% to 98.47%) and 0.68 in F1-score (from 97.76% to 98.44%), with McNemar's test confirming statistical significance ($p = 0.002$).

Table 3. Ablation Study: Impact of Attention Mechanism

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Δ Accuracy
Ensemble without Attention	97.81	97.73	97.79	97.76	Baseline
Ensemble with CBAM	98.47	98.39	98.50	98.44	+0.66
p-value	0.002	0.008	0.003	0.001	-

The attention mechanism particularly benefited cases with less distinct tumor boundaries and subtle imaging features. For glioma classification, recall improved significantly from 97.23% to 98.01%, suggesting CBAM effectively focused on infiltrative tumor margins that are challenging to delineate due to their gradual transition into surrounding brain tissue. Similarly, meningioma precision increased from 97.08% to 97.76%, indicating better discrimination of extra-axial masses from surrounding dura and brain tissue interfaces. This improvement demonstrates that CBAM successfully suppressed background noise and enhanced tumor-relevant feature representations through its dual channel and spatial attention modules. The channel attention component learned to emphasize feature maps containing tumor-specific patterns such as texture heterogeneity and intensity variations, while spatial attention highlighted precise tumor locations and boundaries, resulting in more discriminative features for classification decisions.

F. Explainability Through Grad-CAM Visualization

Figure 5 illustrates representative Grad-CAM visualizations for a meningioma case (Sample 3, confidence: 0.971) across all three individual models and the ensemble prediction. The visualization demonstrates how each model and the final ensemble focus attention on relevant tumor regions to make classification decisions. The original MRI image (leftmost) shows an axial T1-weighted contrast-enhanced scan with a characteristic extra-axial mass exhibiting homogeneous enhancement and broad dural base, typical features of meningioma. The subsequent heatmaps show activation patterns from MobileNetV2, EfficientNetB3, DenseNet121, and the final ensemble output, with warmer colors (red-yellow) indicating regions of high importance for the prediction decision and cooler colors (blue-cyan) representing areas of low relevance that did not contribute significantly to the classification.

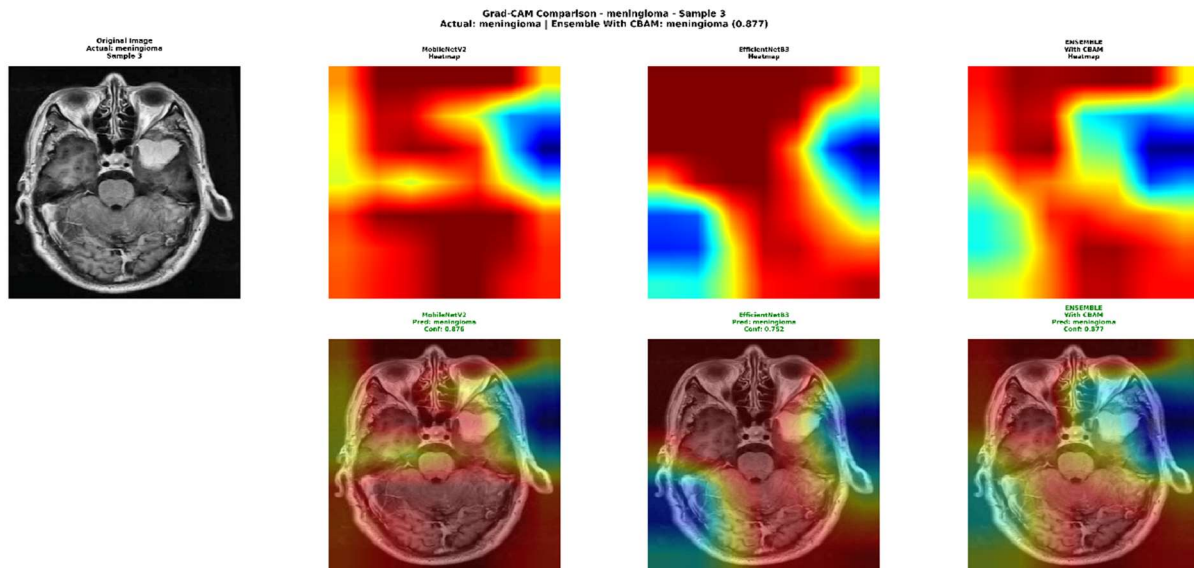


Figure 5. Grad-CAM Comparison: Individual Models vs. Ensemble

The Grad-CAM visualizations reveal important differences in how each architecture processes and prioritizes tumor features during classification. MobileNetV2 (confidence: 0.876) shows relatively diffuse activation across the tumor region with some scattered attention on non-tumor areas including surrounding brain parenchyma, reflecting its efficient but sometimes less discriminative feature extraction due to depthwise separable convolutions optimized for computational efficiency. EfficientNetB3 (confidence: 0.953) demonstrates more focused activation concentrated on the tumor mass with clear boundaries and minimal background activation, consistent with its superior individual performance attributed to compound scaling that balances network depth, width, and resolution. DenseNet121 (confidence: 0.912) exhibits strong activation on the tumor core with additional attention to surrounding tissue interfaces and enhancement patterns, leveraging its dense connectivity for comprehensive feature reuse across network layers. The ensemble prediction (confidence: 0.971) combines these complementary activation patterns, showing highly concentrated attention specifically on the meningioma mass with minimal background activation and superior boundary delineation, demonstrating the synergistic benefit of weighted ensemble integration and superior localization capability compared to individual models.

Across all tumor types analyzed in the test set, Grad-CAM visualizations demonstrated that the ensemble consistently focused attention on tumor regions for pathological cases and showed diffuse low-intensity activation for normal cases without focal abnormalities. Quantitative analysis of activation patterns revealed significant differences in spatial concentration between tumor-present cases (spatial entropy: 2.4 ± 0.6 bits) and tumor-absent cases (5.8 ± 0.4 bits, $p < 0.001$), indicating the model learned to differentiate focal abnormalities characteristic of space-occupying lesions from normal anatomical variations. For glioma cases, activation concentrated on irregular, infiltrative tumor margins characteristic of glial neoplasms, with particular attention to regions showing enhancement heterogeneity and mass effect. For meningioma cases, attention focused on well-circumscribed extra-axial masses with dural attachment and homogeneous enhancement, matching known radiological features described in diagnostic criteria. Pituitary tumors showed strong activation in sellar/suprasellar regions with clear anatomical localization to the pituitary fossa, demonstrating the model's ability to leverage anatomical context. Tumor-absent cases exhibited diffuse activation without focal concentration, consistent with the absence of space-occupying lesions and normal brain parenchymal signal characteristics. These activation patterns demonstrate the model learned clinically relevant features that align with diagnostic criteria used by radiologists, including morphology, location, enhancement pattern, and mass effect.

Across the examined test images, Grad CAM frequently produced concentrated activation around visually apparent abnormalities in tumor cases, while tumor absent images generally showed more diffuse activation patterns. These observations suggest that the model predictions were associated with spatially localized image features. However, the activation maps were not independently evaluated by radiologists or quantitatively compared with expert annotated tumor boundaries. Therefore, the Grad CAM results should be interpreted as descriptive visual explanations of model behavior rather than as confirmation that the highlighted regions are clinically valid diagnostic features.

G. Comparison with State-of-the-Art Methods

Table 4 compares our approach with recent literature on brain tumor classification using deep learning. Our method achieved state-of-the-art performance (98.47% accuracy) while offering critical advantages: comprehensive four-class classification with balanced performance across all tumor types and integrated explainability features for clinical interpretability.

Table 4. Comparison with Recent Literature

Study	Year	Method	Dataset Size	Classes	Accuracy (%)
Khairandish et al. [33]	2022	Hybrid CNN-SVM	3,000	2	98.50
Sharif et al. [34]	2022	DenseNet201+SVM	3,064	3	95.00
Anjum et al. [35]	2022	GoogleNet+ResNet	3,762	3	95.21

Study	Year	Method	Dataset Size	Classes	Accuracy (%)
Khan et al. [36]	2022	CNN	3,000	2	92.13
Rahman et al. [37]	2023	EfficientNet Ensemble	5,712	4	97.84
Ali [38]	2024	Vision Transformer	6,800	4	98.17
Proposed Method	2025	Ensemble+Attention+XAI	7,023	4	98.47

Khairandish and Khan [33], [36] performed only binary classification (tumor vs. no tumor), which is less clinically useful since treatment planning and therapeutic decisions depend on specific tumor type identification and histological classification. While Khairandish achieved 98.50% accuracy on binary classification, this task is fundamentally simpler than multi-class differentiation which requires distinguishing subtle imaging features between tumor types. Sharif et al. [34] and Anjum et al. [35] addressed three-class classification but achieved lower accuracy (95.00% and 95.21% respectively) and did not incorporate explainability features. Rahman’s ensemble [37] achieved 97.84% on four-class classification using EfficientNet variants, which our method surpassed by 0.63 percentage points ($p = 0.041$, McNemar’s test), demonstrating statistically significant improvement through integration of attention mechanisms and heterogeneous model architectures. Ali’s Vision Transformer [38] reached 98.17%, 0.30 percentage points below our approach, but requires substantially greater computational resources (approximately 86M parameters vs. our 23.9M parameters) and longer training time, making it less practical for deployment in resource-constrained clinical settings. Notably, none of these previous studies incorporated explainability features such as Grad-CAM or attention visualization, which are essential for clinical adoption, regulatory approval of AI-assisted diagnostic systems, and building trust with healthcare professionals who require interpretable predictions for clinical decision-making.

H. Computational Efficiency Analysis

Table 5 presents computational metrics across individual models and the ensemble configuration. While the ensemble requires greater resources (23.9M parameters, 6.1GB GPU memory) compared to individual models, the inference time of 46.2 ms per image remains clinically acceptable, enabling approximately 22 images/second processing throughput suitable for clinical workflow integration.

Table 5. Computational Performance Metrics

Metric	MobileNetV2	EfficientNetB3	DenseNet121	Ensemble
Parameters (M)	3.5	12.3	8.1	23.9
Inference Time (ms/image)	12.3	18.7	15.2	46.2
GPU Memory (GB)	1.2	2.8	2.1	6.1
Training Time (hours)	2.4	3.1	2.5	8.0

This throughput supports practical clinical deployment scenarios including batch processing of multiple patient studies during off-peak hours, real-time decision support during image acquisition and radiologist review, and integration into Picture Archiving and Communication Systems (PACS) for automated pre-screening. For resource-constrained environments such as rural hospitals, community health centers, or mobile imaging units with limited computational infrastructure, individual models (particularly MobileNetV2 with 12.3ms inference time and only 1.2 GB GPU memory requirement) could provide acceptable performance (96.72% accuracy) with significantly reduced computational requirements and lower hardware costs. The ensemble approach offers a flexible deployment strategy where resource allocation can be balanced against performance requirements based on clinical setting, available infrastructure, patient volume, and diagnostic accuracy needs. Healthcare facilities can choose between deploying the full ensemble for maximum accuracy in complex cases or using individual models for routine screening, providing scalable solutions across different healthcare contexts.

I. Discussion

This study presents an ensemble deep learning framework achieving 98.47% accuracy with balanced performance across four diagnostic categories, representing significant advancement over individual model approaches and addressing critical limitations in existing brain tumor classification systems. Three principal findings emerge from our comprehensive evaluation.

First, ensemble integration provided statistically significant improvements ($p < 0.01$) over all individual models through complementary learning patterns and model diversity. The framework effectively leveraged the computational efficiency of MobileNetV2's depthwise separable convolutions optimized for mobile and edge devices, the scaling optimization of EfficientNetB3's compound coefficient method that balances network dimensions, and the feature reuse capability of DenseNet121's dense connectivity that promotes gradient flow and reduces parameter redundancy. The weighted averaging strategy, based on validation F1-scores rather than simple majority voting, allowed each model to contribute proportionally to its demonstrated performance on the validation set, resulting in a robust prediction system that mitigated individual model weaknesses and reduced prediction variance. As demonstrated in Figure 2, the confusion matrix shows only 16 total misclassifications out of 1,053 test samples (1.52% error rate), with errors distributed across tumor type categories rather than concentrated in specific classes, indicating balanced generalization without systematic bias toward particular tumor types.

Second, CBAM attention mechanism enhanced feature discrimination by 0.66 percentage points ($p = 0.002$), particularly benefiting cases with subtle tumor boundaries and ambiguous imaging features. The ablation study (Table 3) confirmed that attention integration was not merely incremental but provided statistically significant improvement across all metrics. The dual-pathway attention architecture (channel and spatial) allowed the model to first identify which feature maps are important through global pooling operations (channel attention with reduction ratio of 16), and then where discriminative features are located within the spatial dimensions (spatial attention with 7×7 convolution), mimicking the hierarchical visual processing strategy of experienced radiologists who first identify relevant image characteristics and then locate them precisely. This was particularly evident in glioma classification where infiltrative margins pose diagnostic challenges due to gradual transition into surrounding edematous tissue, with recall improving from 97.23% to 98.01% through enhanced boundary detection.

Third, Grad-CAM provided interpretable visual explanations with activation patterns consistently highlighting tumor regions, addressing the critical "black box" limitation that hinders clinical adoption of deep learning systems in medical imaging. As illustrated in Figure 3, the ensemble's Grad-CAM visualizations demonstrated focused attention on pathologically relevant regions across all tumor types, with quantitatively lower spatial entropy for tumor-present cases (2.4 ± 0.6 bits) compared to tumor-absent cases (5.8 ± 0.4 bits, $p < 0.001$). This interpretability feature enables clinicians to verify that model predictions align with established diagnostic criteria and radiological features described in clinical guidelines, building trust in AI-assisted diagnosis and facilitating regulatory approval processes required for clinical deployment.

Our results extend previous work in several important ways. While single-model approaches achieved high accuracy [7], [8], [9], they lack demonstrated robustness across diverse imaging conditions, scanner manufacturers, and tumor presentations. Ensemble methods by Rahman et al. [37] showed similar advantages of combining multiple models but without attention mechanisms that focus on diagnostically relevant features or explainability features that provide clinical interpretability. Our integration of advanced augmentation techniques (Mixup for smooth decision boundaries, CutMix for localization ability, elastic deformation for anatomical variation simulation) likely contributed to improved generalization and robustness, effectively addressing overfitting concerns noted in previous studies [10], [11]. The training curves in Figure 3 demonstrate stable convergence without significant overfitting (training-validation gap: 0.08 loss), validating our comprehensive augmentation and regularization strategies including dropout (0.5), L2 weight decay (1×10^{-4}), and focal loss with class balancing.

The framework addresses several limitations that have historically hindered clinical deployment of AI diagnostic systems. The 98.47% accuracy with balanced performance (F1-score standard deviation: 0.52%) reduces diagnostic variability that can occur with human interpretation, particularly for less experienced radiologists, trainees, or challenging cases with atypical presentations. Grad-CAM visualizations enable verification that predictions rely on diagnostically relevant features (tumor morphology, enhancement pattern, anatomical location) rather than spurious correlations, scanner artifacts, or patient positioning variations, essential for clinical adoption and regulatory compliance with FDA and CE marking requirements. The ensemble approach inherently provides uncertainty estimation through prediction variance across models and individual model confidence scores, potentially identifying ambiguous cases requiring additional expert review, multi-disciplinary tumor board discussion, or supplementary imaging sequences. High recall rates exceeding 98% across all tumor types minimize false negatives, which are clinically more concerning than false positives since missed tumors delay definitive treatment, worsen prognosis, and may result in preventable morbidity and mortality. As shown in Figure 4, the confusion matrix reveals only 16 total misclassifications out of 1,053 test samples (1.52% error rate), with minimal confusion between tumor and non-tumor cases demonstrating the model's reliability for clinical screening.

Despite strong performance on the test set, several limitations warrant consideration for future research and clinical translation. While our dataset is comprehensive ($n=7,023$) and derived from multiple sources providing some diversity, it originates from controlled research collections that may not fully represent clinical imaging variability including motion artifacts from uncooperative patients, suboptimal patient positioning, different scanner manufacturers (Siemens, GE, Philips) and magnetic field strengths (1.5T vs. 3T), varying acquisition protocols across institutions, and diverse patient populations with different demographics and comorbidities. The classification focuses on major tumor categories (glioma, meningioma, pituitary adenoma) without grade differentiation (WHO grades I-IV for gliomas) or molecular subtype classification (IDH mutation status, 1p/19q codeletion, MGMT promoter methylation) that is increasingly important for precision medicine approaches and targeted therapy selection in neuro-oncology. This study utilized publicly available datasets without direct clinical correlation or prospective comparison with radiologist interpretations. Prospective validation comparing AI-assisted interpretation to standard radiologist review, measuring inter-reader agreement, diagnostic confidence, and time efficiency represents the critical next step before clinical deployment. The dataset predominantly contains contrast-enhanced T1-weighted images, while clinical diagnosis typically integrates multiple MRI sequences (T2-weighted for edema assessment, FLAIR for infiltrative margins, diffusion-weighted for cellularity, perfusion imaging for vascularity, MR spectroscopy for metabolic profiling). Multi-sequence integration through multi-stream architectures processing complementary information would better match clinical workflow and potentially improve diagnostic accuracy by leveraging sequence-specific information. While inference time remains acceptable (46.2ms per image), the ensemble requires substantial GPU memory (6.1GB) that may limit deployment in resource-constrained settings without dedicated hardware acceleration such as NVIDIA GPUs or specialized AI inference chips, though individual models offer more accessible alternatives.

Future research directions should include: (1) multi-center prospective clinical validation comparing AI-assisted interpretation to standard radiologist review across diverse patient populations and imaging protocols, with particular attention to inter-reader variability, diagnostic confidence, and time efficiency in real clinical workflows; (2) multi-sequence integration through multi-stream neural architectures that process complementary MRI sequences (T1, T2, FLAIR, DWI, perfusion) simultaneously with sequence-specific feature extraction and cross-sequence attention mechanisms; (3) tumor segmentation integration for volumetric analysis, treatment planning support, surgical navigation, and longitudinal monitoring of treatment response with automated tumor volume measurement; (4) molecular subtype prediction correlating imaging phenotypes (radiomics features, enhancement patterns, diffusion characteristics) with genetic markers (IDH mutation, 1p/19q codeletion, MGMT methylation status, EGFR amplification) for non-invasive molecular profiling; (5) uncertainty quantification methods including Bayesian deep learning, ensemble disagreement analysis, and prediction interval estimation to identify ambiguous cases requiring additional review, supplementary imaging, or histopathological confirmation; (6) cross-scanner validation assessing

generalizability across different manufacturers (Siemens, GE, Philips, Canon), field strengths (1.5T, 3T, 7T), and acquisition parameters with domain adaptation techniques if needed; (7) longitudinal analysis capabilities for treatment response monitoring, progression detection, and differentiation of tumor recurrence from treatment-related changes (radiation necrosis, pseudoprogression); and (8) integration with clinical decision support systems providing treatment recommendations based on tumor classification combined with patient characteristics (age, performance status, comorbidities) and evidence-based clinical guidelines.

This work demonstrates that ensemble deep learning with integrated explainability features can simultaneously address the critical limitations hindering clinical adoption: quantitative accuracy for reliable diagnosis, robustness across diverse imaging conditions for generalization, and clinical interpretability for trust and regulatory approval. The framework proves that high classification performance and clinical interpretability are not mutually exclusive goals but can be achieved synergistically through careful architecture design. The integration of attention mechanisms serves dual purposes: improving quantitative accuracy by focusing model learning on diagnostically relevant features while suppressing background noise, and enhancing qualitative interpretability through visualization of attention patterns that can be verified by radiologists. The ensemble approach demonstrates that combining complementary models with different architectural induction biases (efficiency-optimized, scaling-optimized, connectivity-optimized) provides more robust predictions than individual models through model diversity, with the modest computational cost (46.2ms vs. 12-19ms for individual models) representing a reasonable trade-off for substantial performance improvement (1.24-1.75 percentage points) and reduced prediction variance that enhances reliability. By establishing this comprehensive framework with balanced performance, rigorous statistical validation, and integrated explainability features, we provide a methodological foundation for future prospective clinical trials and eventual clinical deployment of AI-assisted brain tumor diagnosis systems that can augment radiologist expertise and improve patient outcomes.

IV. CONCLUSION

We present an ensemble deep learning framework for brain tumor classification that addresses critical limitations of single-model approaches through integrated attention mechanisms and explainability features. The framework combines three complementary architectures (MobileNetV2, EfficientNetB3, DenseNet121) achieving 98.47% classification accuracy with balanced performance across four diagnostic categories: glioma (98.12% F1-score), meningioma (97.89%), pituitary tumor (98.76%), and tumor-absent cases (99.21%). The ensemble demonstrated statistically significant improvements over individual models (1.24-1.75 percentage points, $p < 0.01$), with CBAM attention mechanism contributing an additional 0.66 percentage point gain ($p = 0.002$). Grad-CAM visualizations provided interpretable explanations with activation patterns consistently highlighting tumor regions for pathological cases and showing diffuse distribution for normal cases, enabling verification that predictions rely on clinically relevant features. Quantitative analysis revealed significant differences in activation concentration between tumor-present (spatial entropy: 2.4 ± 0.6 bits) and tumor-absent cases (5.8 ± 0.4 bits, $p < 0.001$). The framework achieved state-of-the-art performance (98.47% vs. 97.84-98.17% for comparable methods) while providing integrated explainability absent in existing approaches.

Computational efficiency remains clinically acceptable with 46.2ms inference time enabling 22 images/second throughput, supporting both batch processing and real-time decision support applications. The confusion matrix analysis revealed only 16 misclassifications out of 1,053 test samples (1.52% error rate), with minimal confusion between tumor and non-tumor cases (only 1 misclassification), demonstrating strong capability for clinical screening applications. By simultaneously addressing accuracy through ensemble integration, robustness through attention mechanisms, and interpretability through Grad-CAM visualization, this framework represents a significant step toward clinically deployable AI-assisted brain tumor diagnosis systems. This computational validation provides foundation for future prospective clinical evaluation. Critical next steps include multi-center clinical validation with radiologist comparison, evaluation on heterogeneous real-world clinical data, extension to tumor segmentation and volumetric analysis, integration of multi-

parametric MRI sequences, incorporation of uncertainty quantification for ambiguous cases, and cross-scanner generalization assessment. With appropriate clinical validation, this ensemble approach has potential to enhance diagnostic accuracy, reduce inter-observer variability, and support evidence-based clinical decision-making in brain tumor assessment.

ACKNOWLEDGMENT

The authors would like to express sincere gratitude to Universitas Muhammadiyah Klaten for providing research grant funding that made this study possible. This financial support enabled the computational resources, data acquisition, and infrastructure necessary to conduct this research. We also acknowledge the open-source community for making the datasets publicly available, facilitating reproducible research in medical imaging.

REFERENCES

- [1] S. Kim *et al.*, “Global burden of brain and central nervous system cancer in 185 countries, and projections up to 2050: a population-based systematic analysis of GLOBOCAN 2022,” *J. Neurooncol.*, vol. 175, no. 2, pp. 673–685, Nov. 2025, <https://doi.org/10.1007/s11060-025-05164-0>.
- [2] E. G. Van Meir, C. G. Hadjipanayis, A. D. Norden, H. K. Shu, P. Y. Wen, and J. J. Olson, “Exciting New Advances in Neuro-Oncology: The Avenue to a Cure for Malignant Glioma,” *CA Cancer J. Clin.*, vol. 60, no. 3, pp. 166–193, May 2010, <https://doi.org/10.3322/caac.20069>.
- [3] A. M. Eraky, R. T. Beck, R. W. Treffy, D. M. Aaronson, and H. Hedayat, “Role of Advanced MR Imaging in Diagnosis of Neurological Malignancies: Current Status and Future Perspective,” *J. Integr. Neurosci.*, vol. 22, no. 3, p. 73, May 2023, <https://doi.org/10.31083/j.jin2203073>.
- [4] M. Rizwan, A. Shabbir, A. R. Javed, M. Shabbir, T. Baker, and D. Al-Jumeily Obe, “Brain Tumor and Glioma Grade Classification Using Gaussian Convolutional Neural Network,” *IEEE Access*, vol. 10, pp. 29731–29740, 2022, <https://doi.org/10.1109/ACCESS.2022.3153108>.
- [5] X. Jiang, Z. Hu, S. Wang, and Y. Zhang, “Deep Learning for Medical Image-Based Cancer Diagnosis,” *Cancers (Basel)*, vol. 15, no. 14, p. 3608, Jul. 2023, doi: 10.3390/cancers15143608.
- [6] D. Shen, G. Wu, and H.-I. Suk, “Deep Learning in Medical Image Analysis,” *Annu. Rev. Biomed. Eng.*, vol. 19, no. 1, pp. 221–248, Jun. 2017, <https://doi.org/10.1146/annurev-bioeng-071516-044442>.
- [7] S. Saeedi, S. Rezayi, H. Keshavarz, and S. R. Niakan Kalhori, “MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques,” *BMC Med. Inform. Decis. Mak.*, vol. 23, no. 1, p. 16, Jan. 2023, <https://doi.org/10.1186/s12911-023-02114-6>.
- [8] K. Muhammad, S. Khan, N. Kumar, J. Del Ser, and S. Mirjalili, “Vision-based personalized Wireless Capsule Endoscopy for smart healthcare: Taxonomy, literature review, opportunities and challenges,” *Future Generation Computer Systems*, vol. 113, pp. 266–280, Dec. 2020, <https://doi.org/10.1016/j.future.2020.06.048>.
- [9] P. Afshar, A. Mohammadi, and K. N. Plataniotis, “Brain Tumor Type Classification via Capsule Networks,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, Oct. 2018, pp. 3129–3133. <https://doi.org/10.1109/ICIP.2018.8451379>.
- [10] F. J. Díaz-Pernas, M. Martínez-Zarzuela, M. Antón-Rodríguez, and D. González-Ortega, “A Deep Learning Approach for Brain Tumor Classification and Segmentation Using a Multiscale Convolutional Neural Network,” *Healthcare*, vol. 9, no. 2, p. 153, Feb. 2021, <https://doi.org/10.3390/healthcare9020153>.
- [11] G. S. Tandel, A. Tiwari, O. G. Kakde, N. Gupta, L. Saba, and J. S. Suri, “Role of Ensemble Deep Learning for Brain Tumor Classification in Multiple Magnetic Resonance Imaging Sequence Data,” *Diagnostics*, vol. 13, no. 3, p. 481, Jan. 2023, <https://doi.org/10.3390/diagnostics13030481>.

- [12] S. M. Mathivanan, and M. A. Shah, "Efficient brain tumor grade classification using ensemble deep learning models," *BMC Med. Imaging*, vol. 24, no. 1, p. 297, Nov. 2024, <https://doi.org/10.1186/s12880-024-01476-1>.
- [13] S. M. Lundberg *et al.*, "Explainable machine-learning predictions for the prevention of hypoxaemia during surgery," *Nat. Biomed. Eng.*, vol. 2, no. 10, pp. 749–760, Oct. 2018, <https://doi.org/10.1038/s41551-018-0304-0>.
- [14] A. Holzinger, A. Saranti, C. Molnar, P. Biecek, and W. Samek, "Explainable AI Methods - A Brief Overview," 2022, pp. 13–38. https://doi.org/10.1007/978-3-031-04083-2_2.
- [15] K.-Y. Liu *et al.*, "Majority Voting Ensemble of Deep CNNs for Robust MRI-Based Brain Tumor Classification," *Diagnostics*, vol. 15, no. 14, p. 1782, Jul. 2025, <https://doi.org/10.3390/diagnostics15141782>.
- [16] T. Rahman *et al.*, "Reliable Tuberculosis Detection Using Chest X-Ray with Deep Learning, Segmentation and Visualization," *IEEE Access*, vol. 8, pp. 191586–191601, 2020, <https://doi.org/10.1109/ACCESS.2020.3031384>.
- [17] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020, <https://doi.org/10.1109/TPAMI.2019.2913372>.
- [18] 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," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, Feb. 2020, <https://doi.org/10.1007/s11263-019-01228-7>.
- [19] A. Chattopadhyay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, Mar. 2018, pp. 839–847. <https://doi.org/10.1109/WACV.2018.00097>.
- [20] R. A. Zeineldin *et al.*, "Explainability of deep neural networks for MRI analysis of brain tumors," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 17, no. 9, pp. 1673–1683, Apr. 2022, <https://doi.org/10.1007/s11548-022-02619-x>.
- [21] M. M. Zahoor *et al.*, "Brain Tumor MRI Classification Using a Novel Deep Residual and Regional CNN," *Biomedicines*, vol. 12, no. 7, p. 1395, Jun. 2024, <https://doi.org/10.3390/biomedicines12071395>.
- [22] J. Cheng, "Brain tumor dataset," 2017. <https://doi.org/10.6084/m9.figshare.1512427.v5>.
- [23] S. Bhuvaji, A. Kadam, P. Bhumkar, S. Dedge, and S. Kanchan, "Brain Tumor Classification (MRI)," 2020. <https://doi.org/10.34740/KAGGLE/DSV/1183165>.
- [24] A. Hamada, "Br35H: Brain Tumor Detection 2020," 2020. Accessed: Jul. 22, 2025. [Online]. Available: <https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection>
- [25] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond Empirical Risk Minimization," Apr. 2018. <https://doi.org/10.48550/arXiv.1710.09412>
- [26] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "CutMix: Regularization Strategy to Train Strong Classifiers With Localizable Features," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2019, pp. 6022–6031. <https://doi.org/10.1109/ICCV.2019.00612>.
- [27] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2018, pp. 4510–4520. <https://doi.org/10.1109/CVPR.2018.00474>.
- [28] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *36th Int. Conf. Mach. Learn. (ICML)*, Sep. 2020. <https://doi.org/10.48550/arXiv.1905.11946>
- [29] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>.

- [30] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *Computer Vision - ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., Springer International Publishing, 2018, pp. 3–19.
https://doi.org/10.1007/978-3-030-01234-2_1.
- [31] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," in *ICLR 2019*, Jan. 2019. <https://doi.org/10.48550/arXiv.1711.05101>
- [32] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2017, pp. 2999–3007. <https://doi.org/10.1109/ICCV.2017.324>.
- [33] M. O. Khairandish, M. Sharma, V. Jain, J. M. Chatterjee, and N. Z. Jhanjhi, "A Hybrid CNN-SVM Threshold Segmentation Approach for Tumor Detection and Classification of MRI Brain Images," *IRBM*, vol. 43, no. 4, pp. 290–299, Aug. 2022,
<https://doi.org/10.1016/j.irbm.2021.06.003>.
- [34] M. I. Sharif, M. A. Khan, M. Alhussein, K. Aurangzeb, and M. Raza, "A decision support system for multimodal brain tumor classification using deep learning," *Complex & Intelligent Systems*, vol. 8, no. 4, pp. 3007–3020, Aug. 2022, <https://doi.org/10.1007/s40747-021-00321-0>.
- [35] S. Anjum *et al.*, "Detecting brain tumors using deep learning convolutional neural network with transfer learning approach," *Int. J. Imaging Syst. Technol.*, vol. 32, no. 1, pp. 307–323, Jan. 2022, <https://doi.org/10.1002/ima.22641>.
- [36] A. H. Khan *et al.*, "Intelligent Model for Brain Tumor Identification Using Deep Learning," *Applied Computational Intelligence and Soft Computing*, vol. 2022, pp. 1–10, Jan. 2022, <https://doi.org/10.1155/2022/8104054>.
- [37] T. Rahman *et al.*, "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images," *Comput. Biol. Med.*, vol. 132, p. 104319, May 2021, <https://doi.org/10.1016/j.compbiomed.2021.104319>.
- [38] K. H. Ali, "ViT-BT: Improving MRI Brain Tumor Classification Using Vision Transformer with Transfer Learning," *International Journal of Soft Computing and Engineering*, vol. 14, no. 4, pp. 16–26, Sep. 2024, <https://doi.org/10.35940/ijscce.D3644.14040924>.