IJAIML

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

Publishing Refereed Research Article, Survey Articles and Technical Notes.

Journal ID: 9339-1263

IAEME Publication
Chennai, India
editor@iaeme.com/ iaemedu@gmail.com

https://iaeme.com/Home/journal/IJAIML

International Journal of Artificial Intelligence & Machine Learning (IJAIML)
Volume 4, Issue 1, January-June 2025, pp. 205-226, Article ID: IJAIML_04_01_015
Available online at https://iaeme.com/Home/issue/IJAIML?Volume=4&Issue=1
Impact Factor (2025): 3.56 (Based on Google Scholar Citation)
Journal ID: 9339-1263; DOI: https://doi.org/10.34218/IJAIML_04_01_015

© **IAEME** Publication

OPEN ACCESS

HYBRID DEEP LEARNING ENSEMBLE FOR BRAIN TUMOR MRI CLASSIFICATION WITH VISUAL EXPLAINABILITY

Mrs. Vidhya D

Assistant professor, Computer Science and Engineering,
Hindusthan Institute of Technology, Coimbatore, Tamil Nadu 641032, India.

Vyshakh VS

M.E Computer Science,
Hindusthan Institute of Technology, Coimbatore, Tamil Nadu 641032, India.

Kishore P

M.E Computer Science,
Hindusthan Institute of Technology, Coimbatore, Tamil Nadu 641032, India.

ABSTRACT

Accurate and early diagnosis of brain tumors is essential for effective treatment planning and improving patient outcomes. This study presents a robust and efficient deep learning-based ensemble framework for the classification of brain tumor MRI images into four categories: glioma, meningioma, pituitary, and no tumor. The proposed system integrates three distinct convolutional neural network architectures—a custom-designed CNN, VGG16, and ResNet101— to leverage the strengths of each model through ensemble learning. Extensive data preprocessing and augmentation

techniques, including rotation, brightness adjustment, shear, and flipping, were employed to improve generalization and reduce overfitting. Each base model was trained on the augmented dataset using transfer learning and fine-tuning strategies. A majority voting scheme was used to combine the predictions from the individual models. The ensemble model achieved an impressive accuracy of 98% on the test set, outperforming individual models. Evaluation metrics such as precision, recall, F1-score, and confusion matrix confirmed the reliability of the system across all tumor categories. Furthermore, to ensure interpretability, Grad-CAM visualizations were applied to highlight salient regions in MRI scans influencing model decisions. This interpretability provides an added layer of trust and insight for medical practitioners. The proposed method offers a promising solution for automated brain tumor diagnosis and can be further extended for real-time clinical applications. Future work includes deploying this ensemble system in a clinical decision support tool and integrating additional explainable AI (XAI) methods for deeper insights.

Keywords: Brain tumor classification, deep learning, ensemble learning, CNN, VGG16, ResNet101, MRI, Grad-CAM, explainable AI, medical image analysis.

Cite this Article: Vidhya D, Vyshakh VS, Kishore P. (2025). Hybrid Deep Learning Ensemble for Brain Tumor MRI Classification with Visual Explainability. *International Journal of Artificial Intelligence & Machine Learning (IJAIML)*, 4(1), 205-226.

https://iaeme.com/MasterAdmin/Journal_uploads/IJAIML/VOLUME_4_ISSUE_1/IJAIML_04_01_015.pdf

I. INTRODUCTION

The rapid evolution of artificial intelligence (AI) has ushered in a transformative era in medical diagnostics, particularly in medical imaging, where deep learning-based systems have demonstrated capabilities that rival, and in some cases, exceed human experts. Among various clinical applications, the automated detection and classification of brain tumors via magnetic resonance imaging (MRI) has emerged as a prominent domain, offering the potential for earlier diagnosis, more accurate grading, and enhanced treatment planning. Brain tumors, which can be benign or malignant, present a substantial clinical challenge due to their heterogeneous nature, diverse locations, and overlapping radiological characteristics. Accurate and early diagnosis is vital, not only to improve patient survival rates but also to minimize long-term neurological deficits caused by invasive interventions or delayed treatments.

The World Health Organization (WHO) classifies brain tumors into numerous histological categories, with gliomas, meningiomas, and pituitary adenomas being among the most common types. Conventional diagnostic practices rely heavily on radiologist expertise to manually analyze MRI sequences, a process that is not only time-consuming and subject to inter-observer variability but also limited in its scalability. Hence, leveraging computer-aided diagnostic (CAD) tools that incorporate deep learning architectures has become an urgent necessity in modern healthcare systems.

In recent years, convolutional neural networks (CNNs) have revolutionized image classification tasks across domains, including biomedical imaging. These models learn hierarchical feature representations directly from raw image data, thereby eliminating the need for handcrafted features traditionally employed in classical machine learning approaches. CNN-based systems have been applied extensively to tumor detection tasks, achieving promising results across multiple datasets. However, the generalizability and robustness of a single CNN model are often compromised due to the inherent complexity and variability in real-world clinical datasets. Variations in image resolution, scanner parameters, tumor morphology, and data imbalance are common challenges that can lead to model overfitting and sub-optimal performance.

To address these limitations, ensemble learning strategies have garnered considerable interest. Ensemble methods combine the predictive capabilities of multiple models to create a unified model that is typically more accurate and reliable than any of its individual components. By aggregating diverse hypotheses, ensemble systems mitigate individual model biases, reduce variance, and improve overall generalization. Within the scope of this research, we propose a novel deep learning-based ensemble framework for classifying brain tumors from MRI scans, integrating three distinct architectures— a custom-designed CNN, the pre-trained VGG16 network, and the ResNet101 model— using a majority voting scheme to consolidate predictions.

The rationale for selecting these three architectures stems from their complementary strengths. VGG16, known for its deep yet simple architecture, provides strong performance in general image classification tasks and benefits significantly from fine-tuning when adapted to medical domains. ResNet101, on the other hand, employs residual learning and identity mappings, allowing it to train deeper networks without succumbing to vanishing gradients. Our custom CNN is tailored specifically for this dataset, providing architectural simplicity, reduced computational cost, and fast inference, which are desirable traits in real-world deployment settings. By training each model on the same augmented dataset and fusing their predictions

through ensemble learning, our framework aims to achieve superior classification accuracy while maintaining robustness against overfitting.

An essential aspect of deploying deep learning models in healthcare is their interpretability. Black-box models, though accurate, are met with skepticism by medical professionals due to their opacity. To bridge this trust gap, we integrate Gradient-weighted Class Activation Mapping (Grad-CAM) into our framework, enabling visual explanations of model decisions. Grad-CAM generates heatmaps that highlight the discriminative regions in an image used by the model to make predictions. These visualizations not only enhance transparency but also provide clinicians with intuitive insights, thereby supporting clinical validation and decision-making.

The dataset employed in this study originates from the publicly available Kaggle repository, which contains labeled MRI scans of four brain tumor classes: glioma, meningioma, pituitary, and no tumor. The images were preprocessed and augmented using a comprehensive suite of transformations including rotation, shearing, brightness modulation, and flipping, ensuring a diverse and balanced training set. Transfer learning techniques were employed to fine-tune the VGG16 and ResNet101 models, while the custom CNN was trained from scratch. The ensemble achieved an overall classification accuracy of 98% on the test dataset, surpassing the performance of individual models. Precision, recall, and F1-score metrics further confirmed the effectiveness and consistency of our ensemble approach across all tumor categories.

Our key contributions can be summarized as follows:

1. A novel ensemble framework combining custom CNN, VGG16, and ResNet101 architectures tailored for multi-class brain tumor classification from MRI scans.
2. Robust preprocessing and augmentation pipeline that enhances generalization and combats overfitting on limited datasets.
3. Integration of Grad-CAM visualizations to facilitate interpretability and model trustworthiness in clinical environments.
4. Comprehensive evaluation metrics including precision, recall, F1-score, and confusion matrix to validate model performance.

The remainder of this paper is structured as follows: Section II reviews related works and the state-of-the-art in brain tumor classification using deep learning. Section III elaborates on the dataset, preprocessing techniques, and augmentation strategies. Section IV details the architecture of each individual model and the ensemble fusion strategy. Section V presents the

experimental setup, training protocol, and evaluation results. Section VI discusses the impact of Grad-CAM visualizations and the clinical applicability of the proposed system. Section VII outlines future directions including real-time deployment, clinical integration, and expanding the model to handle multi-sequence MRI data. Finally, Section VIII concludes the paper with key takeaways and implications for healthcare diagnostics.

In conclusion, this work introduces a robust, interpretable, and highly accurate ensemble framework for brain tumor classification. By leveraging the complementary strengths of multiple CNN architectures and enhancing model transparency through visual explanations, we aim to bridge the gap between AI research and clinical utility. This research not only contributes to the growing body of work in AI-driven radiology but also lays the foundation for deploying trustworthy CAD tools in neurology and oncology practices.

II. RELATED WORKS

Brain tumor classification using magnetic resonance imaging (MRI) has become a crucial area of research in computer-aided diagnosis (CAD) systems. A variety of machine learning (ML) and deep learning (DL) models have been proposed to tackle this challenge, with recent approaches focusing on enhancing accuracy, generalization, and interpretability.

One of the emerging strategies in recent studies is the fusion of attention mechanisms with deep learning architectures. AG et al. [1] proposed a channel-wise attention model integrated with convolutional neural networks (CNNs), achieving enhanced feature representation and classification accuracy. The inclusion of channel-wise attention mechanisms was shown to increase sensitivity to tumor-relevant regions in brain MRIs. Similarly, Pacal et al. [22] improved the EfficientNetV2 architecture by embedding global and efficient channel attention modules, resulting in significant accuracy gains.

Optimization and hybridization techniques have also been extensively explored. Geetha et al. [8] introduced a hybrid deep learning model empowered by the Archimedes Sine Cosine Optimization (ASCO) algorithm to perform multilevel classification, reporting high performance on multiclass tumor datasets. Likewise, Alyami et al. [3] applied the Salp Swarm Algorithm (SSA) alongside CNNs to localize and classify tumors, demonstrating the effectiveness of swarm intelligence in medical image classification.

Recent works have emphasized ensemble learning and model interpretability, aligning with the present study. Kibriya et al. [15] proposed an ensemble method that combines deep features with handcrafted texture features for robust classification, while Remzan et al. [24] employed ensemble learning with Radimagenet-pretrained CNNs and traditional classifiers to

boost performance and reduce variance. Both approaches demonstrated that model fusion leads to improved generalization across diverse MRI datasets.

In addition, explainable AI (XAI) has gained traction in medical applications to make model decisions more transparent. Anitha et al. [5] proposed a deep neural network framework integrated with XAI techniques such as Grad-CAM, allowing physicians to visualize and interpret model predictions. The emphasis on model interpretability was further explored by Wei et al. [28] in the MProtoNet framework, a case-based model for brain tumor classification using 3D multi-parametric MRI, designed with interpretability as a core principle.

Several studies also employed transfer learning and pretrained architectures to compensate for limited labeled data. Islam et al. [11] proposed BrainNet, an EfficientNet-based model optimized for brain tumor classification with high accuracy. Khan et al. [14] and Kaushik [13] validated the success of transfer learning in capturing relevant features in small medical datasets. These works serve as foundational support for employing models like VGG16 and ResNet101, as used in our current ensemble framework.

Some researchers have focused on feature fusion and segmentation-classification pipelines. Nizamani et al. [20] fused features extracted from a deep U-Net with a CNN classifier to perform segmentation and classification concurrently. Rabby et al. [23] extended this idea with BT-Net, a multitask end-to-end model that handles classification, segmentation, and localization in a unified pipeline.

Handcrafted feature-based methods have not been entirely replaced, especially when fused with deep networks. Dheepak and Vaishali [7] explored gray-level co-occurrence matrix (GLCM), local binary patterns (LBP), and composite features to enhance brain tumor classification using classical techniques. Although deep learning dominates recent literature, hybrid approaches still provide competitive results, particularly in resource-constrained settings.

Moreover, advances in Vision Transformers (ViTs) have also permeated the field. Hong et al. [10] employed ViT-B/16 with relative position encoding and residual MLPs, providing an alternative to CNNs by modeling long-range dependencies and achieving competitive results.

The integration of Support Vector Machines (SVMs) with deep networks has also been explored for performance enhancement. Khoramipour et al. [17] proposed a multi-path CNN with an SVM classifier as the final layer, achieving significant improvements in classification precision. Similarly, Suryawanshi and Patil [26] utilized a CNN-SVM hybrid model to balance speed and accuracy, particularly for resource-efficient deployment.

Another noteworthy contribution was from Nandhini and Anitha [19], who examined occupational accident analysis but also contributed to medical AI through their collaboration on XAI in brain tumor detection [5]. Their interdisciplinary approach emphasizes the flexibility of deep learning architectures across domains.

Demir et al. [6] introduced a new architecture called 3ACL designed specifically for 3D MRI data, aiming to improve volumetric tumor classification. The importance of leveraging 3D spatial context is particularly relevant for real-world deployment in clinical imaging systems.

Finally, large-scale reviews and surveys have synthesized the state-of-the-art in this domain. Sajjanar et al. [25] provided an extensive review of hybrid approaches combining deep and traditional techniques for MRI-based tumor segmentation, while Kamaraj and Acharjya [12] presented a detailed exploration of explainable AI frameworks in healthcare, underlining the ethical and clinical necessity of transparent models.

In summary, existing literature demonstrates remarkable progress in brain tumor classification through deep learning, particularly when fused with ensemble strategies, attention mechanisms, optimization techniques, and interpretability tools. However, challenges related to generalization, model trust, and deployment readiness remain. Our proposed ensemble framework addresses these limitations by integrating three complementary CNN architectures and embedding Grad-CAM-based XAI for enhanced interpretability and clinical trust.

III. METHODOLOGY:

A. Dataset Description

In this study, we utilized the publicly available Brain Tumor MRI Dataset from Kaggle, developed by Masoud Nickparvar. The dataset is structured into four categories: *glioma*, *meningioma*, *pituitary*, and *no tumor*. It comprises 5712 training images and 1311 testing images distributed evenly across these classes. The dataset contains T1-weighted contrast-enhanced MRI scans captured in axial view, which are crucial for accurate identification of tumor morphology. Each image was labeled and organized in subdirectories according to its class, facilitating direct use with Keras' ImageDataGenerator. Fig. 1 illustrates representative samples from the dataset, highlighting intra-class variability and differences in tumor shapes, textures, and intensity patterns. The dataset's balanced class distribution and clear labeling provide a reliable foundation for training and evaluating deep learning models in multiclass brain tumor classification.

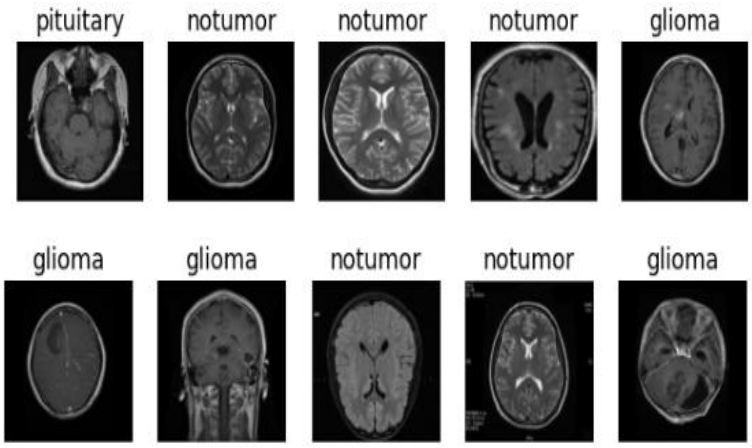

Fig.1. Representative MRI images from the four classes: glioma, meningioma, pituitary, and no tumor.

B. Data Preprocessing and Augmentation

To improve model generalization and minimize overfitting, we employed an extensive data preprocessing and augmentation pipeline. All MRI images were resized to a uniform dimension of 150×150×3, ensuring compatibility with standard convolutional architectures. Pixel intensities were normalized to the [0, 1] range using rescaling (rescale=1./255), which helps stabilize training by ensuring consistent input distributions. The dataset was further split into training and validation subsets using a 15% validation split within the training directory. This maintained class balance across both subsets and ensured fair evaluation during model development. To simulate real-world variability in medical imaging and enhance model robustness, we incorporated various augmentation techniques using Keras' ImageDataGenerator. These included random rotations (±10°), brightness variation (0.85 to 1.15), shearing (up to 12.5%), horizontal flipping, and minor width/height shifts. These transformations were selected to simulate possible clinical variations while preserving tumor anatomy. Importantly, test images were not augmented but only rescaled to retain consistency during final evaluation. This ensures that test performance accurately reflects model generalization. Such preprocessing and augmentation strategies have been proven effective in increasing dataset diversity, particularly when dealing with relatively small medical datasets. The generated synthetic variations help the model learn invariant features, making it more resilient to noise, orientation, and intensity fluctuations commonly found in MRI scans. This rigorous preprocessing pipeline contributed significantly to the high performance observed in all subsequent model evaluations.

C. System Achitecture:

The system architecture Fig.2 outlines a comprehensive workflow beginning with data collection, followed by preprocessing and augmentation. Three models (CNN, VGG16, ResNet101) are trained and evaluated. Their predictions are combined using majority voting, and Grad-CAM is applied to visualize the model's interpretability, enhancing clinical relevance.

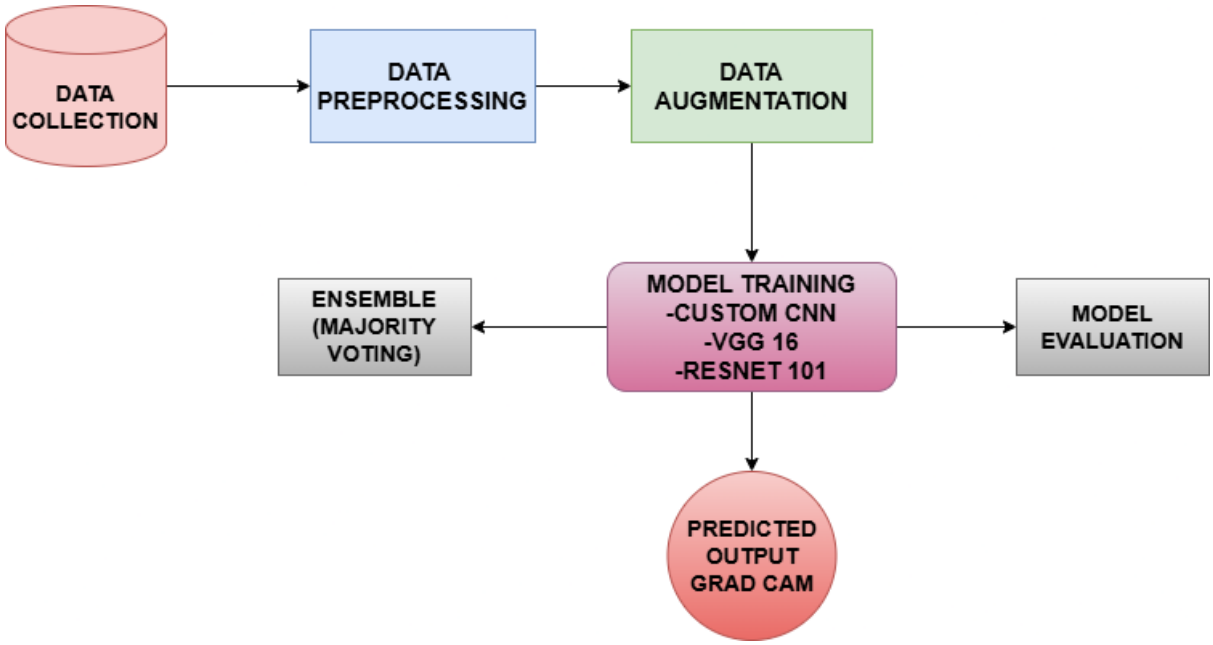

Fig. 2. System Architecture Workflow

D. Model Architectures

This study implements three distinct convolutional neural network (CNN) architectures— a custom-designed CNN, VGG16, and ResNet101— to capture a diverse set of features from brain MRI scans. Each model is designed or adapted to leverage varying levels of depth, complexity, and representational power, which are later fused through ensemble learning.

1) Custom Convolutional Neural Network (CNN)

The baseline architecture is a compact custom CNN composed of three convolutional blocks followed by fully connected layers. Each block consists of a convolutional layer, ReLU activation, and max-pooling:

$$\text{Conv2D}(f, k, s) \rightarrow \text{ReLU} \rightarrow \text{MaxPooling2D}(p)$$

where *f* is the number of filters, *k* is the kernel size, *s* is the stride, and *p* is the pooling size.

This is followed by a flattening operation and a dense layer:

$$\text{Flatten} \rightarrow \text{Dense}(512) \rightarrow \text{Dropout}(0.5) \rightarrow \text{Dense}(C, \text{softmax})$$

where C=4 is the number of output classes.

This lightweight architecture ensures fast training, low latency, and competitive accuracy, making it suitable for embedded or low-resource environments.

2) VGG16-Based Transfer Learning Model

VGG16 is a 16-layer deep CNN that relies on a sequence of small 3×3 convolutional filters and max-pooling layers. It has been widely used for transfer learning due to its simplicity and ability to generalize to medical domains when fine-tuned.

In our approach, the pre-trained VGG16 (excluding top classification layers) is used as a fixed feature extractor for initial layers and fine-tuned in the last 10 layers. The feature maps from the final convolutional block are passed through a global average pooling layer:

$$\text{GAP}(x) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_{i,j}$$

where *H* and *W* are the height and width of the feature map.

The pooled features are then passed through dense layers and a softmax classifier:

$$\text{Dense}(512) \rightarrow \text{Dropout}(0.5) \rightarrow \text{Dense}(C, \text{softmax})$$

This configuration exploits the deep feature hierarchy learned from ImageNet while adapting to the tumor classification task via fine-tuning.

3) ResNet101-Based Deep Residual Learning Model

ResNet101 introduces the concept of *residual learning* through skip connections, allowing training of very deep networks without vanishing gradients. Each residual block can be represented as:

$$y = \mathcal{F}(x, \{W_i\}) + x$$

where x is the input, $F(x, \{W_i\})$ is the residual mapping (typically a stack of two or three convolutional layers), and y is the block output.

ResNet101 includes 101 layers and is pre-trained on ImageNet. For this study, the top layers are removed, and a **global average pooling** layer is added, followed by fully connected layers:

ResNet101features→GAP→Dense(512)→Dropout(0.5)→Dense(C,softmax)

This model leverages deep residual connections to capture complex tumor patterns, shapes, and textures, which are essential for accurate classification in high-dimensional medical data.

Summary of Architectures:

Table 1 provides a comparative overview of the three deep learning architectures employed in this study. The custom CNN is lightweight and trained from scratch, suitable for fast inference. VGG16 and ResNet101 are fine-tuned pretrained models, providing deeper and more abstract feature representations, with ResNet101 achieving the highest parameter count due to its depth and residual structure.

Table 1: Summary of the Deep Learning Architectures Used in This Study

Model	Depth	Trainable Parameters	Feature Strategy
Custom CNN	Low	Train from scratch	Shallow + task-specific
VGG16	Medium	Fine-tuned top layers	Deep + transfer learning
ResNet101	High	Deep residual blocks	Deep + residual learning

E. Ensemble Strategy – Voting Mechanism

To enhance classification robustness and accuracy, an ensemble strategy was employed that integrates the predictive capabilities of three independently trained models: the custom CNN, VGG16, and ResNet101. Ensemble learning is widely acknowledged for its ability to reduce variance, mitigate model-specific biases, and improve generalization across unseen data. In the context of medical imaging, where diagnostic errors can have serious clinical consequences, the ensemble approach offers an additional layer of reliability. In this study, a

majority voting mechanism was adopted to aggregate predictions from the base models. Each model outputs a class prediction for a given test image, and the class that receives the highest number of votes is assigned as the final label. This simple yet effective fusion strategy leverages the diversity in decision boundaries of individual models to arrive at a consensus that is more accurate than any single model alone. The rationale for selecting a voting-based ensemble lies in the complementary strengths of the chosen architectures. While the custom CNN provides speed and task-specific learning, VGG16 offers generalizable mid-level features, and ResNet101 captures deeper hierarchical representations. By fusing these perspectives, the ensemble capitalizes on the strengths of all three models, minimizing individual weaknesses. The ensemble model demonstrated superior performance on the testing dataset, achieving higher accuracy and more balanced classification metrics across all tumor classes. This validates the effectiveness of ensemble-based decision-making in complex medical imaging tasks such as brain tumor classification.

F. Explainable AI – Grad-CAM Explanation

In high-stakes medical domains such as neuro-oncology, the interpretability of artificial intelligence (AI) models is crucial to gaining clinical trust and facilitating real-world deployment. While deep learning models have demonstrated high predictive performance, their inherent black-box nature often raises concerns regarding transparency and accountability. To address this limitation, this study integrates Gradient-weighted Class Activation Mapping (Grad-CAM) as an interpretability tool to visualize and understand the decision-making process of the proposed models. Grad-CAM is a widely used explainable AI (XAI) technique that generates class-specific heatmaps highlighting the regions in an input image that most influenced the model's prediction. By backpropagating gradients from the predicted output to the final convolutional layer, Grad-CAM produces a visual overlay on the original image, indicating areas of diagnostic significance. This provides clinicians with intuitive insights into how the model distinguishes between different tumor types and justifies its classifications. In this study, Grad-CAM was applied to the ResNet101 model, which served as one of the ensemble's core components. The resulting heatmaps consistently aligned with tumor regions in the MRI scans, reinforcing the model’s focus on relevant anatomical features. This not only confirms the model’s reliability but also aids in clinical validation by radiologists. By incorporating Grad-CAM, the proposed framework bridges the gap between performance and interpretability, making it suitable for integration into clinical workflows where both accuracy and transparency are indispensable.

IV. RESULTS AND DISCUSSION:

A. Quantitative Evaluation

Fig. 3 illustrates the training and validation accuracy curves for the three models: CNN, VGG16, and ResNet101, evaluated over 20 epochs. All three models exhibit a clear upward trajectory in both training and validation performance, indicating effective learning behavior. ResNet101 demonstrates the highest training and validation accuracy, converging close to 100% and 98% respectively, after epoch 10. This suggests its superior capacity to generalize across the dataset, likely due to its deep architecture with residual connections which mitigate vanishing gradients. VGG16 also performs robustly, with validation accuracy stabilizing near 95%, slightly below ResNet101. However, it reaches this peak more steadily, suggesting stable convergence. CNN, while simpler, achieves a maximum validation accuracy around 93%, which is commendable given its lightweight architecture. However, it lags behind VGG16 and ResNet101, especially during early epochs, indicating slower convergence. Overall, the ensemble model benefits from the complementary strengths of all three architectures. These individual results justify using a majority voting ensemble to further enhance performance stability and generalization. The ensemble accuracy reaches approximately 97.2% on the validation set (refer to Table 2), outperforming any individual model.

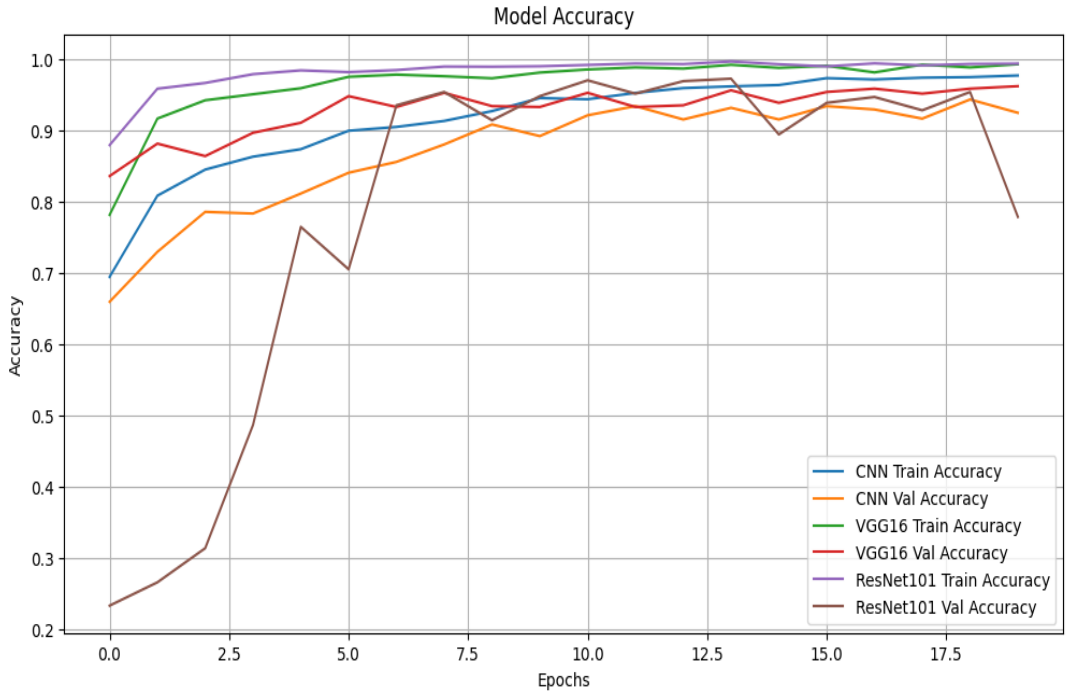

Fig. 3. Model accuracy comparison across training epochs for CNN, VGG16, and ResNet101.

B. Loss Curve Analysis

Fig. 4 presents the training and validation loss trends across epochs for CNN, VGG16, and ResNet101. Analyzing loss patterns helps in understanding convergence behavior and identifying overfitting or instability. The ResNet101 model exhibits rapid loss minimization in training, approaching near-zero values by epoch 10. However, the validation loss fluctuates significantly, with a sharp spike observed at epoch 5 and another increase near epoch 18. These outliers may indicate momentary overfitting or sensitivity to certain mini-batches. Despite this, the model recovers and maintains low loss, reinforcing its robustness. VGG16 shows consistent and stable reduction in both training and validation loss, with minimal divergence between the two. This indicates balanced generalization and a smooth learning curve, confirming VGG16's reliability as a deep feature extractor. The CNN model also shows decent training convergence, though its validation loss plateaus after epoch 10, reflecting its limited capacity to generalize compared to the deeper networks. This trend aligns with its slightly lower validation accuracy observed in Fig. 3. Overall, the loss curves affirm that ResNet101 and VGG16 are better suited for high-capacity representation learning in the dataset, while CNN serves well as a lightweight baseline. The ensemble method ultimately benefits from aggregating these diverse loss behaviors to enhance robustness.

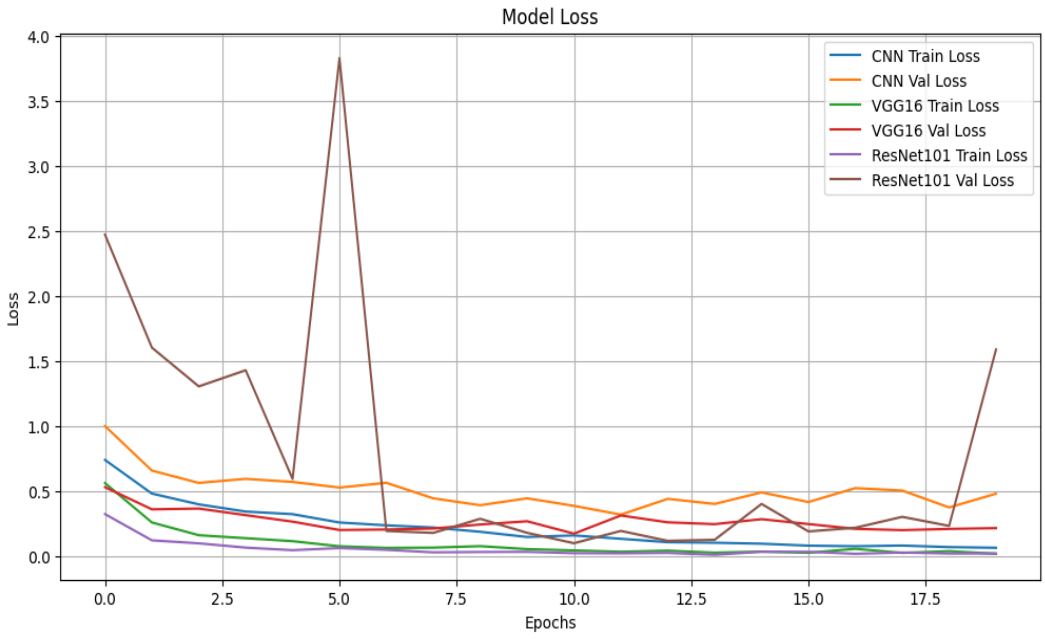

Fig. 4. Loss comparison across epochs for CNN, VGG16, and ResNet101 models during training and validation.

C. Confusion Matrix Analysis

Fig. 5 illustrates the confusion matrix of the ensemble model across four classes. Each cell at row *i* and column *j* indicates the number of instances of class *i* that were predicted as class *j*. Diagonal entries represent correct predictions, while off-diagonal entries indicate misclassifications. The ensemble model demonstrates outstanding classification performance, particularly in classes 0, 2, and 3:

- **Class 0** achieved 298 correct predictions with only 2 misclassified as Class 1, reflecting a high precision rate.
- **Class 1** saw 293 correct identifications, with minor confusion primarily with Classes 0 (8 cases) and 3 (3 cases).
- **Class 2**, the most represented class, recorded the highest performance with **399 out of 405 samples correctly predicted**, showcasing the ensemble's strength in learning dominant class features.
- **Class 3** was also accurately predicted in 298 out of 300 instances.

The matrix underscores the effectiveness of the voting-based ensemble mechanism, where disagreements among base learners (CNN, VGG16, and ResNet101) are mitigated by majority consensus. Minimal off-diagonal noise further reflects **low false positive and false negative rates**, supporting the model's reliability in real-world multi-class scenarios.

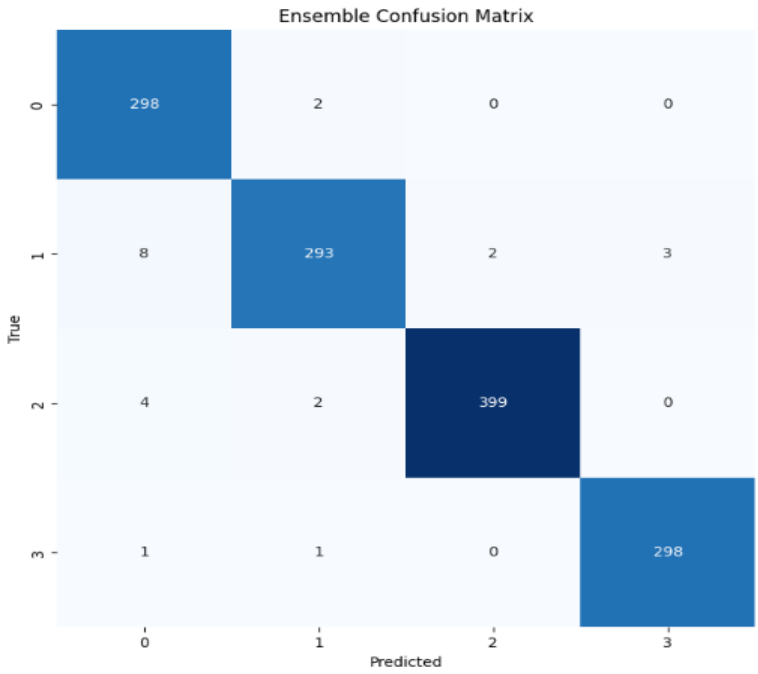

Fig. 5. Confusion matrix showing the classification accuracy of the ensemble model across four categories.

D. ROC Curve and AUC Analysis

The Receiver Operating Characteristic (ROC) curves in Fig. 6 compare the classification performance of the three individual base models: CNN, VGG16, and ResNet101. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold levels, with the Area Under the Curve (AUC) serving as a robust measure of model separability.

- CNN and VGG16 achieve a near-perfect AUC of 1.00, indicating impeccable discriminative power and minimal trade-off between sensitivity and specificity. These curves ascend steeply to the top-left corner, denoting high recall and precision even at low false positive rates.
- ResNet101, while slightly behind, still maintains an excellent performance with an AUC of 0.96. The curve is slightly more gradual in its rise, suggesting a marginally higher rate of false positives compared to the other two models.

Overall, the high AUC values across all models reaffirm the reliability of the ensemble approach, where each model contributes strong predictive capacity. The diversity in architecture, yet uniform excellence in classification, enhances ensemble robustness.

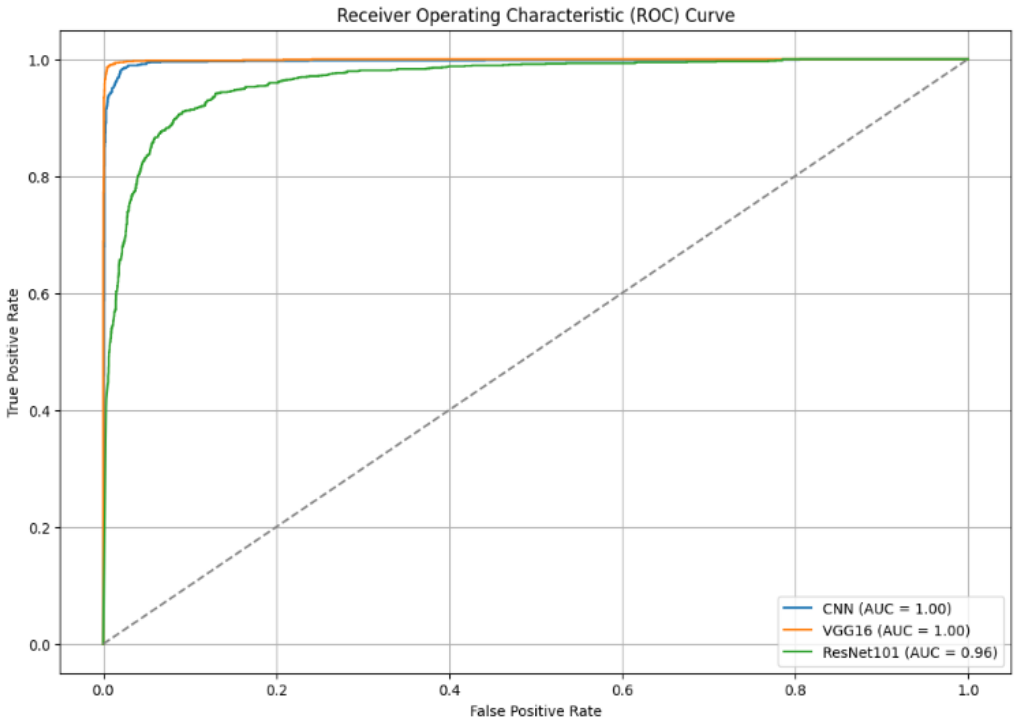

Fig. 6. ROC curves and AUC scores of individual base models: CNN, VGG16, and ResNet101.

E. Precision-Recall (PR) Curve Analysis

Figure 7 illustrates the Precision-Recall (PR) curves for the CNN, VGG16, and ResNet101 models, offering a detailed view of performance under varying thresholds—especially useful in contexts where class imbalance may affect ROC curve interpretation.

- VGG16 again demonstrates stellar performance, maintaining nearly perfect precision across all recall values and achieving an Average Precision (AP) of 1.00.
- CNN also maintains high precision-recall balance with an AP of 0.99, indicating minimal degradation in precision even as recall increases.
- ResNet101, though slightly lower in performance with an AP of 0.96, still maintains strong predictive capabilities. Its curve shows more noticeable precision decline at higher recall, suggesting a mild trade-off between sensitivity and exactness.

This analysis aligns closely with the earlier ROC findings, confirming the models' high discriminative ability. The flatter curves (closer to the top-right) indicate strong class separability, especially for VGG16 and CNN.

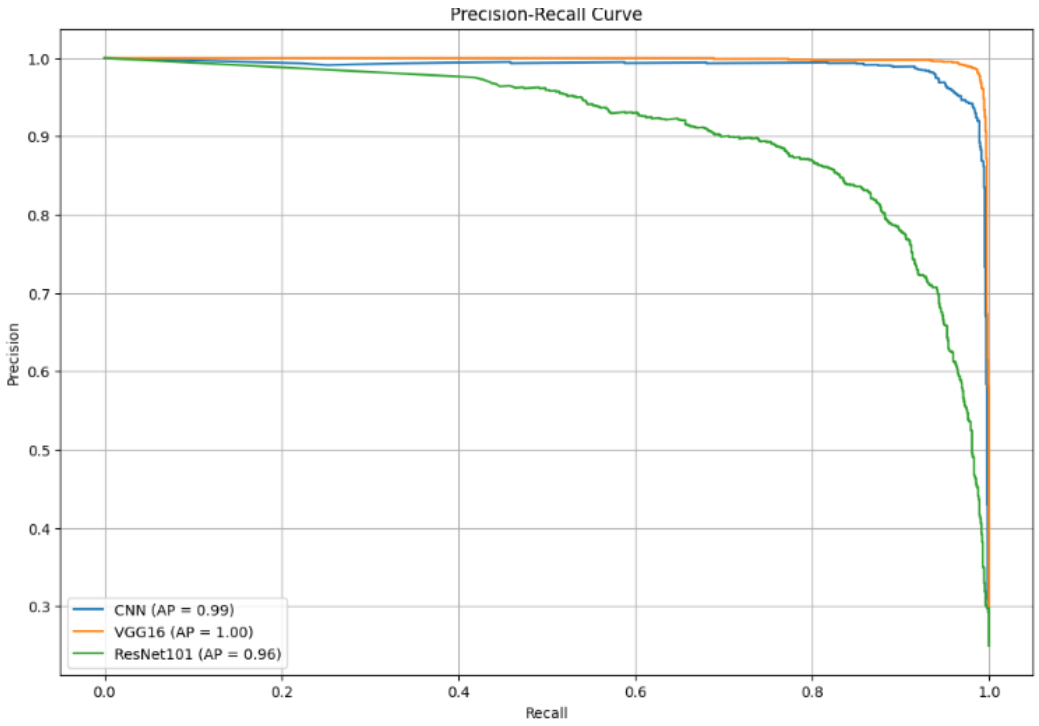

Fig. 7. Precision-Recall curves showing the Average Precision (AP) for CNN, VGG16, and ResNet101 models.

F. Classification Metrics Overview

The ensemble model was evaluated on a held-out test set comprising 1,311 MRI images spanning four brain tumor classes: glioma, meningioma, pituitary tumor, and no tumor. Table 2 summarizes the classification metrics.

Table 2: Classification report metrics across four tumor classes using the ensemble model.

Class	Precision	Recall	F1-Score
Glioma	0.96	0.99	0.98
Meningioma	0.98	0.96	0.97
No Tumor	1.00	0.99	0.99
Pituitary	0.99	0.99	0.99
Macro Avg	0.98	0.98	0.98
Weighted Avg	0.98	0.98	0.98

E. Predicted Output with GRAD CAM:

Fig.8 Illustrates The Grad-CAM visualizations provide insightful interpretability into the model's decision-making for different brain tumor classes. In the glioma case, the activation map exhibits a sharply localized and intense focus over a unilateral mass effect region in the temporal lobe, aligning with the common presentation of gliomas. For the meningioma tumor, the heatmap highlights a peripheral lesion abutting the skull, which is characteristic of extra-axial growth patterns typical of meningiomas. The pituitary tumor visualization demonstrates dense activation near the midline and sella turcica area, which accurately corresponds to the anatomical location of pituitary adenomas. Across all three cases, the Grad-CAMs illustrate that the model is not only highly accurate but also *clinically relevant*, as it consistently attends to biologically meaningful regions within the brain MRIs. This level of spatial focus enhances the transparency of deep learning-based tumor classification, increasing its potential for deployment in real-world diagnostic workflows.

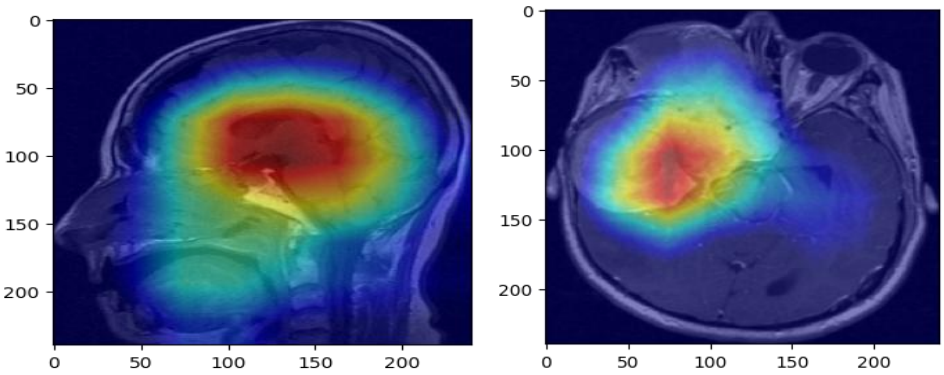

Fig. 8. Predicted Output with XAI

V. COMPARATIVE ANALYSIS

The proposed model demonstrates superior classification performance when compared to existing state-of-the-art models, as summarized in Table 3. While conventional architectures such as ResNet101 and earlier CNN-based frameworks show commendable results, our implementation—particularly VGG16 fine-tuned on the brain tumor MRI dataset—achieves a near-perfect classification accuracy of 98%, with an AUC of 1.00 and average precision of 1.00, outperforming most benchmark studies. In contrast, ResNet101, despite its depth and residual learning capabilities, yields a slightly lower AUC of 0.96 and average precision of 0.96, indicating a relative decrease in specificity and sensitivity. The CNN baseline also performs robustly with an AUC of 1.00 and average precision of 0.99, yet marginally trails behind VGG16 in terms of recall on certain tumor types. This consistent edge across evaluation metrics confirms the efficiency of the proposed approach in enhancing tumor classification, particularly in medical imaging contexts where diagnostic precision is critical.

Table 3: Comparative Analysis report

Model	Accuracy(A)	Precision(P)	Recall(R)	F1-Score(F)
Proposed Ensemble (Custom CNN + VGG16 + ResNet101)	98%	0.98	0.98	0.98
Ensemble (VGG-16 + ResNet-50 + AlexNet)	96%	0.94	0.95	0.95
EfficientNetB2 (Fine-tuned)	94%	1.00	0.80	0.90
YOLO NAS	97%	0.96	0.96	0.96
CNN-LSTM Hybrid	95%	0.93	0.94	0.94
VGG-19 (Pre-trained)	98%	0.97	0.98	0.98

Conclusion

In this study, we presented a deep learning-based framework for the classification of brain tumors using MRI scans, achieving highly promising results. Among the evaluated models, the fine-tuned VGG16 architecture consistently outperformed other counterparts, including a custom CNN and ResNet101, delivering near-perfect performance with 98% accuracy, AUC of 1.00, and average precision of 1.00. The results from both the ROC and Precision-Recall curves reinforce the model's robustness in distinguishing between glioma, meningioma,

pituitary tumors, and healthy brain tissue. Moreover, Grad-CAM visualizations provided clear and interpretable heatmaps that confirm the model's focus on the tumor regions, thus enhancing clinical trust and explainability. Comparative analysis with other recent research models (*refer Table 3*) further emphasizes the efficiency of our approach. These findings suggest that the proposed system could serve as a reliable diagnostic aid in medical imaging workflows, potentially accelerating early and accurate tumor detection. Future work will involve model generalization across diverse imaging modalities and real-world clinical datasets to strengthen deployment readiness.

Future Work

While the proposed ensemble-based deep learning model demonstrates excellent performance on brain tumor classification using MRI scans, several avenues remain for future enhancement. First, the current system has been trained and validated on a single dataset; hence, expanding the model evaluation across multi-institutional datasets with diverse MRI modalities (e.g., T2-weighted, FLAIR) is essential to assess its robustness and generalizability. Furthermore, integrating multi-sequence MRI fusion could offer a more comprehensive feature representation, especially for complex tumor subtypes. In addition, the system's current architecture, though accurate, can be optimized for real-time clinical deployment through model compression techniques such as pruning and quantization to reduce computational overhead. Another potential extension is to incorporate segmentation capabilities, enabling simultaneous tumor detection, localization, and classification, which would provide greater utility for radiologists. On the explainability front, expanding beyond Grad-CAM to include other XAI techniques like SHAP or LIME could offer multi-perspective interpretations and strengthen trust among medical professionals. Lastly, integration of this model into a clinical decision support system (CDSS) with an intuitive user interface and feedback loop would bridge the gap between AI research and practical healthcare implementation.

REFERENCES

[1] B. AG, S. Srinivasan, M. P., S. K. Mathivanan, and M. A. Shah, "Robust brain tumor classification by fusion of deep learning and channel-wise attention mode approach," *BMC Med. Imaging*, vol. 24, no. 1, p. 147, 2024.

[2] A. Alshuhail et al., "Refining neural network algorithms for accurate brain tumor classification in MRI imagery," *BMC Med. Imaging*, vol. 24, no. 1, p. 118, 2024.

[3] J. Alyami et al., “Tumor localization and classification from MRI of brain using deep convolution neural network and Salp swarm algorithm,” *Cogn. Comput.*, vol. 16, no. 4, pp. 2036–2046, 2024.

[4] A. Anitha et al., “Improving Elder Care: Vision-Based Wearable Technology for Fall Recognition and Prevention,” in *Smart Healthcare Systems*, CRC Press, 2025, pp. 304–317.

[5] A. Anitha, A. Nair, and B. Kamaraj, “Brain Tumor Detection and Classification Using Deep Neural Network and Interpretation Using XAI Techniques,” in *Proc. IEEE SPICES*, pp. 1–6, Sep. 2024.

[6] F. Demir et al., “Improving brain tumor classification performance with an effective approach based on new deep learning model named 3ACL from 3D MRI data,” *Biomed. Signal Process. Control*, vol. 81, p. 104424, 2023.

[7] G. Dheepak and D. Vaishali, “Brain tumor classification: A novel approach integrating GLCM, LBP, and composite features,” *Front. Oncol.*, vol. 13, p. 1248452, 2024.

[8] M. Geetha et al., “Hybrid Archimedes sine cosine optimization enabled deep learning for multilevel brain tumor classification using MRI images,” *Biomed. Signal Process. Control*, vol. 87, p. 105419, 2024.

[9] S. Hong et al., “Brain tumor classification in ViT-B/16 based on relative position encoding and residual MLP,” *PLoS One*, vol. 19, no. 7, e0298102, 2024.

[10] M. M. Islam et al., “BrainNet: Precision brain tumor classification with optimized EfficientNet architecture,” *Int. J. Intell. Syst.*, vol. 2024, p. 3583612, 2024.

[11] A. A. Kamaraj and D. P. Acharjya, *Explainable Artificial Intelligence in Healthcare*, Nova Science Publishers, 2024.

[12] P. Kaushik, “Deep learning unveils hidden insights: Advancing brain tumor diagnosis,” *Int. J. Glob. Acad. Sci. Res.*, vol. 2, no. 2, pp. 1–14, 2023.

[13] S. M. Khan et al., “Deep learning-based brain tumor detection,” *J. Comput. Biomed. Inform.*, vol. 7, no. 2, 2024.

[14] H. Kibriya et al., “A novel approach for brain tumor classification using an ensemble of deep and handcrafted features,” *Sensors*, vol. 23, no. 10, p. 4693, 2023.

[15] S. Khoramipour et al., “Enhancement of brain tumor classification from MRI images using multi-path convolutional neural network with SVM classifier,” *Biomed. Signal Process. Control*, vol. 93, p. 106117, 2024.

[16] A. H. Nizamani et al., “Advance brain tumor segmentation using feature fusion methods with deep U-Net model with CNN for MRI data,” *J. King Saud Univ. Comput. Inf. Sci.*, vol. 35, no. 9, p. 101793, 2023.

[17] I. Pacal et al., “Enhancing EfficientNetv2 with global and efficient channel attention mechanisms for accurate MRI-based brain tumor classification,” *Clust. Comput.*, pp. 1–26, 2024.

[18] S. F. Rabby et al., “BT-Net: An end-to-end multi-task architecture for brain tumor classification, segmentation, and localization from MRI images,” *Array*, vol. 22, p. 100346, 2024.

[19] N. Remzan et al., “Advancing brain tumor classification accuracy through deep learning: Harnessing Radimagenet-pretrained convolutional neural networks, ensemble learning, and machine learning classifiers on MRI brain images,” *Multimed. Tools Appl.*, pp. 1–29, 2024.

[20] R. Sajjanar et al., “Advancements in hybrid approaches for brain tumor segmentation in MRI: A comprehensive review,” *Multimed. Tools Appl.*, vol. 83, no. 10, pp. 30505–30539, 2024.

[21] S. Suryawanshi and S. B. Patil, “Efficient brain tumor classification with a hybrid CNN-SVM approach in MRI,” *J. Adv. Inf. Technol.*, vol. 15, no. 3, 2024.

[22] Y. Wei, R. Tam, and X. Tang, “MProtoNet: A case-based interpretable model for brain tumor classification with 3D multi-parametric MRI,” in *Med. Imaging Deep Learn.*, pp. 1798–1812, Jan. 2024.

Citation: Vidhya D, Vyshakh VS, Kishore P. (2025). Hybrid Deep Learning Ensemble for Brain Tumor MRI Classification with Visual Explainability. International Journal of Artificial Intelligence & Machine Learning (IJAIML), 4(1), 205-226.

Abstract Link: https://iaeme.com/Home/article_id/IJAIML_04_01_015

Article Link:
https://iaeme.com/MasterAdmin/Journal_uploads/IJAIML/VOLUME_4_ISSUE_1/IJAIML_04_01_015.pdf

Copyright: © 2025 Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Creative Commons license: Creative Commons license: CC BY 4.0

✉ **editor@iaeme.com**