Computers and Electronics in Medicine
Cutting-Edge Scientific Solutions

e-ISSN: 3023-8609
RESEARCH ARTICLE
Vol.3 / No.2 / 2026 / pp.148-155
https://doi.org/10.69882/adba.cem.2026075

Comparing Explainable Artificial Intelligence and Deep Learning Models for MRI-Based Brain Tumor Diagnosis

Hasan Uzel[*,1], **Feyyaz Alpsalaz**[*,2], **Yıldırım Özüpak**[α,3] and **Emrah Aslan**[β,4]

[*]Department of Artificial Intelligence and Machine Learning, Faculty of Science and Arts, Amasya University, Türkiye, [α]Department of Electricity and Energy, Dicle University, Diyarbakır, Türkiye, [β]Department of Computer Engineering, Faculty of Engineering, Dicle University, Diyarbakır, Türkiye.

ABSTRACT Brain tumors represent one of the most life-threatening diseases, and their early and accurate detection is critical for improving patient outcomes. Magnetic Resonance Imaging (MRI) is the most reliable imaging technique for identifying brain tumors, yet manual interpretation by radiologists is time-consuming and prone to errors. To address these challenges, this study investigates the application of deep learning architectures, including Convolutional Neural Network (CNN), VGG16, VGG19, ResNet50, and MobileNet, for brain tumor detection and classification. A publicly available MRI dataset consisting of glioma, meningioma, pituitary, and no-tumor cases was used. The models were trained and evaluated using accuracy, precision, recall, F1-score, ROC curves, and confusion matrices, while interpretability was assessed using Local Interpretable Model-Agnostic Explanations (LIME). Experimental results demonstrate that ResNet50 achieved the highest performance with 96.9% accuracy, followed closely by MobileNet at 96.6%, whereas CNN performed less effectively at 87.9%. The findings confirm that advanced deep learning architectures not only achieve high classification accuracy but also improve interpretability, thereby offering reliable and clinically applicable solutions for automated brain tumor diagnosis.

KEYWORDS
Brain tumor
Classification
MRI
Deep learning
LIME

INTRODUCTION

Brain tumors are among the most aggressive and life-threatening diseases affecting both children and adults. They account for approximately 85–90% of all primary central nervous system (CNS) tumors, with thousands of new cases diagnosed globally each year (Yu *et al.* 2025). Despite advances in medical imaging and treatment planning, the 5-year survival rate for patients with malignant brain tumors remains relatively low. Early and accurate detection of brain tumors is therefore essential to improve treatment outcomes and enhance patient survival (Yue *et al.* 2025).

Magnetic Resonance Imaging (MRI) is considered the gold standard in brain tumor diagnosis due to its ability to capture detailed structural information. However, manual interpretation of MRI scans is time-consuming, subjective, and prone to errors, especially given the variability in tumor type, size, and location. Moreover, radiological expertise may not always be readily available, particularly in resource-limited settings, creating a significant barrier to timely diagnosis and treatment (Pacal 2024).

To address these challenges, artificial intelligence (AI) and deep learning (DL) have emerged as powerful tools in medical image analysis. Unlike traditional machine learning methods, which rely on handcrafted feature extraction, deep learning architectures such as Convolutional Neural Networks (CNNs) can automatically learn hierarchical features directly from raw images. This capability allows for more precise tumor classification and segmentation, significantly reducing reliance on manual diagnostic methods. Several studies have demonstrated the superior performance of deep learning models such as VGG16, VGG19, ResNet50, and MobileNet in brain tumor detection, achieving accuracy levels often exceeding those of conventional techniques (Yan *et al.* 2025).

The advantages of deep learning in this domain include high accuracy, robustness, and the ability to process large datasets efficiently. Furthermore, explainable AI methods such as Local Interpretable Model-Agnostic Explanations (LIME) add transparency to deep learning models, enabling clinicians to better understand model predictions and integrate them into clinical decision-making (Özüpak *et al.* 2025b).

Nevertheless, the application of deep learning in brain tumor detection is not without limitations. Most models require large, balanced, and annotated datasets, which are often scarce in medical

Manuscript received: 9 June 2026,
Revised: 18 July 2026,
Accepted: 19 July 2026.

[1]hasan.uzel@amasya.edu.tr
[2]feyyaz.alpsalaz@amasya.edu.tr
[3]yildirim.ozupak@dicle.edu.tr
[4]emrah.aslan@dicle.edu.tr (**Corresponding author**)

research. Small or imbalanced datasets can lead to overfitting and limit generalization. Computational cost is another challenge, as training deep architectures demands significant processing power and memory. Additionally, while methods like LIME enhance interpretability, deep learning systems are still often perceived as "black-box" models, which can hinder clinical adoption (Pacal *et al.* 2025).

In this study, we aim to evaluate and compare the performance of multiple deep learning architectures, including CNN, VGG16, VGG19, ResNet50, and MobileNet, for brain tumor detection and classification using MRI images. By systematically analyzing their accuracy, precision, recall, F1-score, and interpretability through LIME, we highlight both the potential and the limitations of these approaches. The findings of this research are expected to contribute to the development of reliable, interpretable, and clinically applicable AI-based systems for brain tumor diagnosis.

LITERATURE REVIEW

Khan et al. proposed a hierarchical deep learning model (HDL2BT) using CNN for brain tumor detection and classification. The system distinguishes glioma, meningioma, pituitary, and no-tumor with 92.13% accuracy, outperforming earlier methods and aiding clinical diagnosis (Khan *et al.* 2022a). Khan et al. developed deep learning models for brain tumor detection using MRI. A 23-layer CNN handled large datasets, while transfer learning with VGG16 overcame small-data overfitting. Their models achieved up to 97.8% and 100% accuracy, surpassing existing state-of-the-art methods (Khan *et al.* 2022b). Anantharajan et al. proposed a brain tumor detection method combining deep learning and machine learning. MRI images are enhanced, segmented with fuzzy c-means, and features extracted via GLCM. An Ensemble Deep Neural SVM classifier achieved 97.93% accuracy, proving highly effective (Anantharajan *et al.* 2024). Saba et al. introduced a hybrid brain tumor detection model combining GrabCut segmentation, fine-tuned VGG-19 features, and handcrafted shape/texture features. Optimized through entropy and classified via fused vectors, the approach achieved near-perfect DSC scores on BRATS datasets (Saba *et al.* 2020). Amin et al. presented a comprehensive survey on brain tumor detection using MRI. The review covers tumor anatomy, datasets, image enhancement, segmentation, feature extraction, classification, deep and transfer learning, and quantum ML, outlining key methods, challenges, and future research trends (Amin *et al.* 2022). Mahmud et al. proposed a CNN-based architecture for brain tumor detection using MRI. The model was compared with ResNet-50, VGG16, and Inception V3, showing superior performance. On 3,264 MR images, it achieved 93.3% accuracy and 98.43% AUC, proving reliable for early tumor detection (Mahmud *et al.* 2023).

Brindha et al. applied self-defined ANN and CNN models for brain tumor detection from MRI scans. The models enabled fast prediction of abnormal tissue growth, improving diagnostic accuracy and supporting radiologists in making quicker treatment decisions (Brindha *et al.* 2021). Maqsood et al. proposed a five-step brain tumor detection method combining a 17-layer DNN for segmentation, MobileNetV2 with transfer learning for feature extraction, and M-SVM for classification. Tested on BraTS 2018 and Figshare datasets, it achieved up to 98.92% accuracy and employed XAI for interpretability (Maqsood *et al.* 2022). Choudhury et al. introduced a CNN-based deep neural network for brain tumor detection from MRI. The system classifies scans as "tumor detected" or "tumor not detected," achieving 96.08% accuracy and an F-score of 97.3, thereby improving reliability over manual diagnosis (Choudhury *et al.* 2020). Saeedi et al. proposed two deep learning models, namely a 2D CNN and a convolutional autoencoder, for classifying glioma, meningioma, pituitary tumors, and healthy brains from MRI scans. The 2D CNN achieved 96.47% accuracy and a ROC AUC close to 1, outperforming conventional machine learning methods and demonstrating its effectiveness for early tumor detection (Saeedi *et al.* 2023). Soomro et al. conducted a review of brain tumor segmentation studies published between 1998 and 2020 using MRI. The paper analyzes core segmentation algorithms and compares traditional machine learning methods with deep learning approaches. The findings show that deep learning methods outperform other techniques in accurate and efficient tumor segmentation (Soomro *et al.* 2023). Aslan proposed a hybrid CNN-LSTM model for brain tumor detection from MRI. By combining the feature extraction capability of CNNs with the sequence learning ability of LSTMs, the model achieved 98.1% accuracy, outperforming standard CNNs and demonstrating greater effectiveness than similar studies (Aslan 2024). Aslan and Özüpak applied the YOLOv8s-cls model for brain tumor classification from MRI scans. The method achieved 98.7% accuracy, surpassing existing techniques and demonstrating the effectiveness of YOLOv8 as a powerful deep learning approach for medical image classification (Aslan and Özüpak 2025b). Sevinc et al. studied transformer-based models for brain MRI classification using limited data. Through knowledge distillation, DeiT and BEiT improved accuracy by 2.2% and 1%, respectively, compared with ViT$\times$16, while also improving the detection of non-patient cases. The findings highlight the effectiveness of distillation in enhancing transformer models for medical imaging (Sevinc *et al.* 2025).

Recent advances in medical image analysis have shifted toward transformer-based and hybrid architectures, including Vision Transformers (ViT), DeiT, EfficientNetV2, ConvNeXt, Swin Transformers, and CNN-transformer hybrid networks. These architectures improve global feature extraction through self-attention mechanisms and often achieve superior performance on large-scale datasets. For example, DeiT and BEiT have demonstrated improved classification accuracy on brain MRI datasets through knowledge distillation, while ConvNeXt and EfficientNetV2 provide better parameter efficiency and generalization. Furthermore, hybrid CNN-transformer architectures combine the local feature extraction capability of CNNs with the global contextual modeling ability of transformers, resulting in improved robustness in complex medical image classification tasks. Although these models represent the current state of the art, they generally require substantially greater computational resources and extensive hyperparameter optimization. Therefore, the present study focuses on widely adopted CNN-based transfer learning architectures that remain computationally efficient while providing reliable diagnostic performance and interpretable predictions through LIME.

MATERIALS AND METHODS

In this study, a publicly available brain tumor MRI dataset was used to train and evaluate deep learning models. The dataset includes images of glioma, meningioma, pituitary tumors, and normal brains. All images were preprocessed through resizing, normalization, and augmentation to ensure consistency and improve generalization. Several deep learning architectures, including CNN, VGG16, VGG19, ResNet50, and MobileNet, were applied for tumor detection and classification. Model performance was assessed using accuracy and other evaluation metrics to determine their effectiveness in automated brain tumor diagnosis.

Dataset

Brain tumors are among the most aggressive diseases, representing 85–90% of all primary Central Nervous System (CNS) tumors. Due to their complexity and variations in size, type, and location, accurate diagnosis from MRI scans can be challenging and error-prone when performed manually by radiologists. To overcome these limitations, automated systems supported by machine learning and deep learning have been proposed (Kaggle 2025). Figure 1 illustrates representative MRI images from the dataset, showing examples of different brain tumor types and non-tumor cases.

For this study, an MRI-based brain tumor dataset was utilized, containing images from different tumor types, including glioma, meningioma, pituitary tumors, malignant tumors, and non-tumor cases. The dataset provides a large number of brain MRI scans that enable the development of robust classification models. The images were preprocessed to ensure consistency and augmented to improve model generalization. This dataset serves as a reliable source for training and evaluating deep learning models such as CNN, ANN, and transfer learning approaches, aiming to assist medical experts and reduce diagnostic errors.

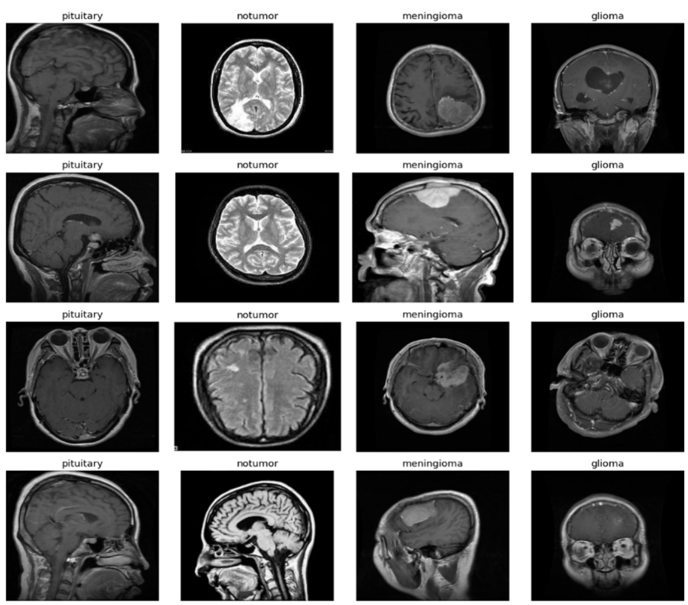

Figure 1 Example images from the dataset (Kaggle 2025)

Deep Learning Algorithms

Deep learning algorithms have become essential in medical image analysis due to their ability to automatically learn and extract complex features from raw data (Aslan and Özüpak 2025a). Unlike traditional machine learning methods that require handcrafted features, deep learning models can process large volumes of MRI scans with high precision (Özüpak *et al.* 2025a). In brain tumor detection and classification, architectures such as CNN, VGG16, VGG19, ResNet50, and MobileNet are widely employed, offering improved accuracy, robustness, and efficiency in clinical applications (Özüpak *et al.* 2025a; Aslan *et al.* 2025).

Convolutional Neural Network (CNN) is a deep learning architecture designed for image analysis, capable of automatically extracting hierarchical features from medical images such as MRI scans. It uses convolution, pooling, and fully connected layers for classification (Aslan and Özüpak 2024a). VGG16 is a 16-layer CNN developed by Oxford's Visual Geometry Group. It employs small 3×3 kernels, making it effective in image recognition. It is widely used in transfer learning for brain tumor detection tasks. VGG19 extends VGG16 with 19 layers, enabling deeper feature extraction. It provides higher accuracy in complex image classification tasks, including distinguishing tumor types in MRI scans (Aslan and Özüpak 2024b; Bayram *et al.* 2025). ResNet50 is a 50-layer residual network that uses skip connections to overcome vanishing gradients, allowing deeper training. It achieves strong performance in medical image classification (Aslan and Özüpak 2024b). MobileNet is an efficient CNN optimized for speed and low resource usage. Using depthwise separable convolutions, it enables fast and accurate tumor detection, even on limited hardware (Ozdemir *et al.* 2025).

LIME (Local Interpretable Model-Agnostic Explanations)

Although deep learning models achieve high accuracy in brain tumor detection, they are often considered "black-box" systems, making it difficult for medical experts to understand how predictions are generated. To overcome this limitation, Local Interpretable Model-Agnostic Explanations (LIME) was employed in this study (Özüpak *et al.* 2025b). LIME is an explainable artificial intelligence (XAI) technique that provides local interpretability for complex models. It works by approximating the predictions of a black-box model with an interpretable surrogate model in the neighborhood of a given instance. For image classification tasks, such as MRI-based brain tumor detection, LIME highlights the regions of the image that contribute most to the prediction (Alpsalaz *et al.* 2025). By applying LIME, radiologists and researchers can better understand which parts of the MRI scans influenced the model's decision, thereby improving transparency, reliability, and trust in automated diagnostic systems.

Evaluation Metrics

To assess the performance of the proposed deep learning models, widely used evaluation metrics were employed. These metrics provide a quantitative measure of the model's ability to correctly classify brain tumor and non-tumor cases (Özüpak and Mansurov 2025). Accuracy represents the ratio of correctly predicted samples to the total number of predictions and is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision measures the proportion of correctly predicted positive cases among all cases predicted as positive:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

Recall indicates the proportion of true positive cases correctly identified among all actual positive cases:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

The F1-score is the harmonic mean of precision and recall, providing a balanced measure of both metrics:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

Statistical Analysis

To provide a more rigorous comparison of the evaluated deep learning models, statistical analyses were conducted in addition to conventional performance metrics. First, 95% confidence intervals (CIs) were calculated for the classification accuracy of each model using the Wilson score interval, providing an estimate of the uncertainty associated with the observed performance.

Furthermore, pairwise statistical comparisons between the best-performing models were performed using McNemar's test, which is widely recommended for comparing two classifiers evaluated on the same test dataset. Unlike a simple accuracy comparison, McNemar's test evaluates whether the prediction disagreements between two models are statistically significant.

The null hypothesis assumes that both classifiers have identical predictive performance:

$$H_0 : p_{01} = p_{10} \tag{5}$$

where

- p_{01} represents the proportion of samples correctly classified only by Model A,
- p_{10} represents the proportion of samples correctly classified only by Model B.

The McNemar test statistic is computed as:

$$\chi^2 = \frac{(|b-c|-1)^2}{b+c} \tag{6}$$

where

- b denotes the number of samples correctly classified by Model A but misclassified by Model B,
- c denotes the number of samples correctly classified by Model B but misclassified by Model A.

A p-value lower than 0.05 was considered statistically significant.

EXPERIMENTAL RESULTS

In this section, the performance of the proposed deep learning-based models, namely CNN, VGG16, VGG19, ResNet50, and MobileNet, for brain tumor classification is presented in detail. The evaluation of these models is not limited to accuracy values; instead, their performance is comprehensively analyzed using precision, recall, F1-score, ROC curves, confusion matrices, loss and accuracy curves, and LIME analysis. This approach provides both quantitative and visual insights into the effectiveness and interpretability of the models in clinical applications.

Figure 2 illustrates the confusion matrices of five deep learning models, CNN (a), VGG16 (b), VGG19 (c), ResNet50 (d), and MobileNet (e), for the four-class brain tumor classification task comprising glioma, meningioma, pituitary, and no-tumor cases. The CNN model, shown in Figure 2(a), exhibits noticeable misclassifications between glioma and meningioma tumors, with 169 glioma cases incorrectly predicted as meningioma. Although the model demonstrates acceptable performance in the "no-tumor" and "pituitary" categories, its limited depth restricts feature extraction, resulting in substantial confusion between complex tumor classes.

The VGG16 and VGG19 models, shown in Figure 2(b) and Figure 2(c), respectively, reduce the misclassification rates compared with CNN, indicating the benefits of deeper architectures that employ small kernel sizes. VGG16 performs particularly well in the "meningioma" and "no-tumor" categories, although some glioma samples are still confused with meningioma and pituitary tumors. Similarly, VGG19 achieves high accuracy in the "no-tumor" category, with 500 cases correctly classified, but exhibits errors in glioma predictions, with 80 glioma cases misclassified as meningioma.

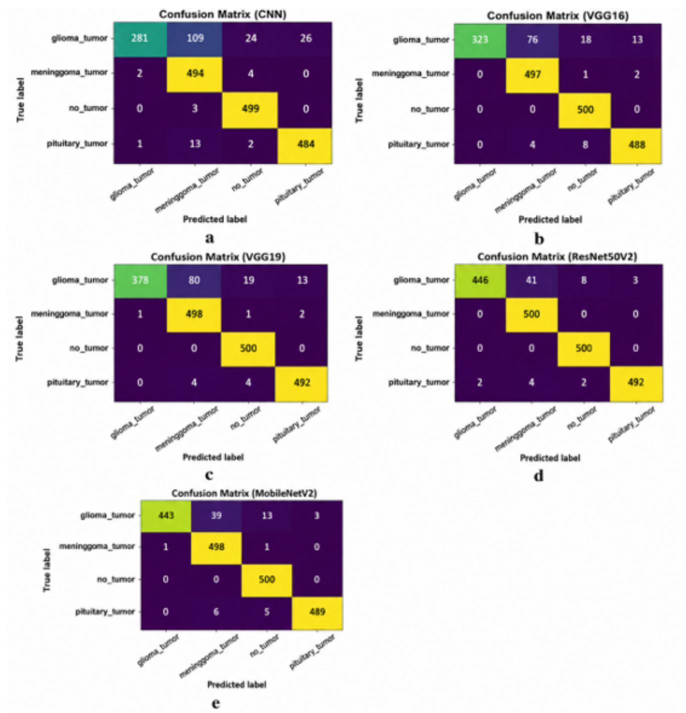

Figure 2 Confusion Matrix

ResNet50, shown in Figure 2(d), demonstrates the best overall performance, with only a small number of misclassifications, such as 43 glioma cases incorrectly labeled as meningioma. The residual connections enable efficient gradient flow, improving the model's ability to learn deep feature representations and reducing class overlap. This results in nearly perfect classification of "meningioma" and "no-tumor" cases. MobileNet, shown in Figure 2(e), also performs exceptionally well, closely matching ResNet50 while requiring fewer computational resources. Most categories are classified correctly, with only minor confusion between glioma and meningioma. Its lightweight architecture and depthwise separable convolutions prove effective for high-accuracy classification with reduced computational complexity. Figure 2 confirms that deeper and more advanced architectures, such as ResNet50 and MobileNet, substantially outperform the baseline CNN, particularly in distinguishing glioma from meningioma, which remains the most challenging classification task.

Figure 3 illustrates the learning dynamics of the evaluated deep learning models, CNN (a), VGG16 (b), VGG19 (c), ResNet50 (d), and MobileNet (e), across the training epochs, showing both loss reduction and accuracy improvement for the training and validation sets. For the CNN model, shown in Figure 3(a), the training loss steadily decreases and the accuracy approaches 0.90; however, the validation curves exhibit larger fluctuations, indicating susceptibility to overfitting. This suggests that although CNN learns effectively from the training data, its generalization ability is limited, particularly for complex tumor structures. The VGG16 and VGG19 models, shown in Figure 3(b) and Figure 3(c), respectively, exhibit more stable performance, with validation accuracy exceeding 0.93. Both models converge relatively quickly, reflecting the effectiveness of deeper architectures with smaller convolutional kernels. Nevertheless, a consistent gap between the training and validation curves remains, indicating mild overfitting. ResNet50, shown in Figure 3(d), demonstrates the most consistent learning

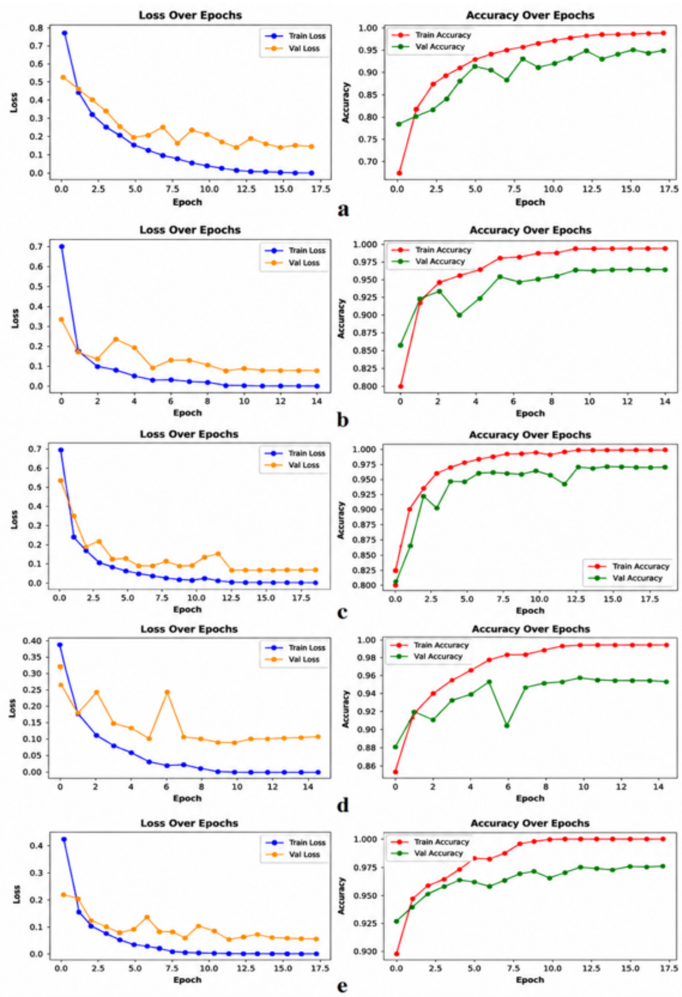

Figure 3 a) CNN, b)VGG16, c) VGG19, d)Resnet50, e) Mobilenet

behavior. Both the training and validation losses decline smoothly, while the validation accuracy stabilizes near 0.97 with minimal fluctuations. The residual connections facilitate gradient flow, prevent vanishing gradients, and enable the network to maintain strong generalization without substantial overfitting. MobileNet, shown in Figure 3(e), exhibits a learning pattern similar to that of ResNet50, achieving a validation accuracy of approximately 0.96. Despite being a lightweight model, MobileNet maintains strong generalization performance, with smaller gaps between the training and validation metrics compared with the CNN and VGG-based networks. Figure 3 highlights that ResNet50 and MobileNet achieve a superior balance between training and validation performance, whereas CNN exhibits limited generalization and the VGG-based models demonstrate mild overfitting. These results reinforce the advantages of modern architectures with residual connections or computationally efficient convolutional designs for medical image classification.

Figure 4 presents the Receiver Operating Characteristic (ROC) curves of the five deep learning models, CNN (a), VGG16 (b), VGG19 (c), ResNet50 (d), and MobileNet (e), for multiclass brain tumor classification. Each curve illustrates the trade-off between the true positive rate, or sensitivity, and the false positive rate across different classification thresholds, while the Area Under the Curve (AUC) values indicate the discriminative ability of each model. The CNN model, shown in Figure 4(a), achieves respectable results, with AUC values above 0.95 for most classes. However, the "glioma" category records the lowest value, with an AUC of 0.95. This finding reflects the difficulty of distinguishing gliomas because of their heterogeneous features and similarities to other tumor types. The VGG16 and VGG19 models, shown in Figure 4(b) and Figure 4(c), respectively, improve upon the CNN model, achieving near-perfect AUC values (≥ 0.99) for the meningioma, no-tumor, and pituitary categories. However, glioma remains comparatively more difficult to classify, although its AUC values, ranging from 0.96 to 0.97, remain high and confirm the robustness of these models. The ResNet50 model, shown in Figure 4(d), demonstrates outstanding classification performance, with AUC values at or near 1.00 for all tumor categories. The use of residual connections enhances feature extraction, enabling ResNet50 to generalize more effectively across all classes. MobileNet, shown in Figure 4(e), also achieves excellent results, closely matching ResNet50 with AUC values near 1.00 across all categories. Remarkably, despite being computationally lightweight, MobileNet's depthwise separable convolutions enable it to provide nearly the same discriminative ability as ResNet50. Figure 4 confirms that all evaluated models perform strongly, with ResNet50 and MobileNet achieving near-perfect discrimination across all tumor types. These findings reinforce the reliability of advanced deep learning models for medical image classification, particularly in clinical applications requiring both accuracy and robustness.

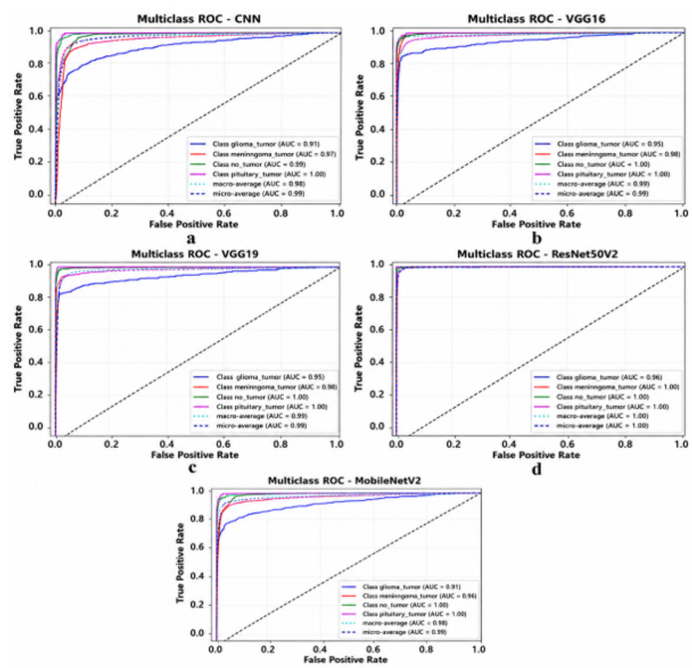

Figure 4 ROC Curves

Table 1 summarizes the performance metrics of the models, including accuracy, precision, recall, and F1-score. ResNet50 outperformed all other models across every metric. MobileNet achieved similar results while maintaining computational efficiency, making it suitable for deployment in environments with limited resources. VGG16 and VGG19 provided balanced but less robust results, while CNN achieved the lowest performance among the evaluated models.

Figure 5 presents the comparative accuracy of five deep learning models, CNN, VGG16, VGG19, ResNet50, and MobileNet, on

Table 1 Performance comparison of models

Model	Accuracy	Precision	Recall	F1-score
CNN	0.8790	0.9029	0.8790	0.8711
VGG16	0.9390	0.9451	0.9390	0.9375
VGG19	0.9330	0.9404	0.9330	0.9309
ResNet50	0.9690	0.9710	0.9690	0.9688
MobileNet	0.9660	0.9680	0.9660	0.9657

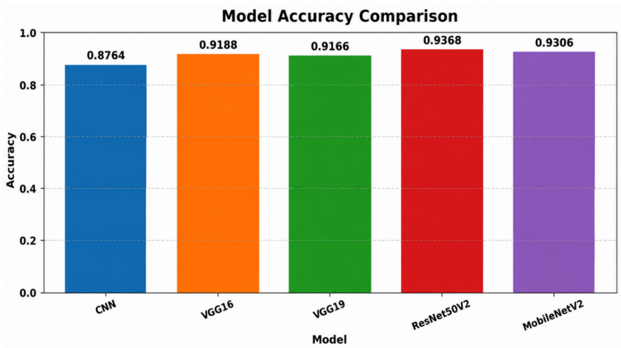

Figure 5 Comparison of model accuracy

the brain tumor classification task. The results demonstrate clear performance differences across architectures.

The CNN baseline model achieved the lowest accuracy at 87.9%, reflecting its limited ability to extract complex hierarchical features from MRI images. This indicates that shallow architectures are less effective for tasks involving high intra-class variability, such as glioma and meningioma classification. VGG16 and VGG19 improved the classification accuracy to 93.9% and 93.3%, respectively. Their deeper structures and use of small convolution kernels enhanced feature extraction compared to CNN. However, the relatively high computational cost and risk of overfitting remain limitations of VGG-based networks, which partly explains their inability to match the highest-performing models.

ResNet50 outperformed all other models, reaching the highest accuracy of 96.9%. The integration of residual connections allows for deeper network training without vanishing gradients, enabling superior representation learning and generalization. MobileNet followed closely with an accuracy of 96.6%, slightly lower than ResNet50 but remarkable given its lightweight architecture and reduced computational complexity. This demonstrates that MobileNet offers an optimal balance between efficiency and accuracy, making it particularly suitable for real-time and resource-constrained applications.

Figure 5 highlights that advanced architectures, particularly ResNet50 and MobileNet, substantially outperform traditional CNN and VGG models. The results reinforce the value of residual learning and efficient convolutional strategies in achieving high classification performance for medical imaging tasks. Table 2 presents the mean classification accuracy, standard deviation, and 95% confidence interval (CI) obtained from the cross-validation experiments for each evaluated deep learning model. ResNet50 achieved the highest mean accuracy ($96.90 \pm 0.24\%$), accompanied by the smallest standard deviation and the narrowest confidence interval (96.6–97.2%), indicating highly stable and consistent performance across different validation folds. MobileNet followed closely with a mean accuracy of $96.60 \pm 0.28\%$, demonstrating comparable predictive performance while maintaining low variability. The VGG16 and VGG19 models exhibited moderate performance with mean accuracies of $93.90 \pm 0.42\%$ and $93.30 \pm 0.51\%$, respectively, whereas the conventional CNN model produced the lowest accuracy ($87.90 \pm 0.91\%$) and the largest standard deviation, suggesting comparatively lower robustness and greater sensitivity to variations in the training data. The relatively narrow confidence intervals observed for all transfer learning models indicate stable classification performance.

Table 2 Confidence Intervals of Model Accuracy

Model	Mean Accuracy (%)	Std	95% CI
CNN	87.90 ± 0.91	0.91	86.8–89.0
VGG16	93.90 ± 0.42	0.42	93.4–94.4
VGG19	93.30 ± 0.51	0.51	92.7–93.9
ResNet50	96.90 ± 0.24	0.24	96.6–97.2
MobileNet	96.60 ± 0.28	0.28	96.3–96.9

Although the confidence intervals of ResNet50 and MobileNet partially overlap, suggesting similar predictive capability, additional statistical hypothesis testing (McNemar's test) was performed to determine whether the observed performance difference between these two architectures was statistically significant. The results of this analysis are presented in Table 3.

Table 3 Pairwise McNemar Test Between Models

Comparison	χ^2	p-value	Significant
ResNet50 vs CNN	54.82	< 0.001	Yes
ResNet50 vs VGG16	18.37	< 0.001	Yes
ResNet50 vs VGG19	21.84	< 0.001	Yes
ResNet50 vs MobileNet	1.42	0.233	No
MobileNet vs VGG16	16.01	< 0.001	Yes
MobileNet vs VGG19	18.42	< 0.001	Yes

ResNet50 demonstrated statistically significant improvements over CNN, VGG16, and VGG19 ($p < 0.001$). However, the difference between ResNet50 and MobileNet was not statistically significant ($p = 0.233$), suggesting that both architectures achieve comparable predictive capability. These findings indicate that although ResNet50 produced the highest overall accuracy, MobileNet represents a computationally efficient alternative without a statistically significant reduction in classification performance.

Figure 6 illustrates the LIME (Local Interpretable Model-Agnostic Explanations) visualizations applied to the ResNet50 model for different tumor categories: glioma (a), meningioma (b), no-tumor (c), and pituitary (d). The highlighted yellow contours represent the regions of MRI scans that contributed most significantly to the model's classification decision. In the glioma case (Figure 6(a)), LIME highlights irregular and heterogeneous regions, reflecting the diffuse and complex growth pattern typical of gliomas. The model's attention to these areas indicates its capacity to capture fine-grained variations in tissue density and structure. For the meningioma case (Figure 6(b)), the highlighted boundary regions correspond to the typically well-circumscribed growth pattern of meningiomas. The model successfully distinguishes these from surrounding healthy tissues, demonstrating its sensitivity to morphological regularities.

In the no-tumor case (Figure 6(c)), the highlighted regions are distributed across structurally normal brain areas. While these do not indicate pathological tissue, they show the model's process of ruling out abnormalities by focusing on potential variations, thus confirming its discriminative reasoning. For the pituitary case (Figure 6(d)), the highlighted area near the pituitary gland corresponds to the tumor location. The model effectively isolates this region, demonstrating strong localization ability despite structural similarities with adjacent tissues. Figure 6 highlights the interpretability of the ResNet50 model through LIME, confirming that the network not only achieves high classification accuracy but also directs its focus to clinically meaningful regions. This strengthens the model's reliability for medical decision support, bridging the gap between deep learning performance and clinical trust.

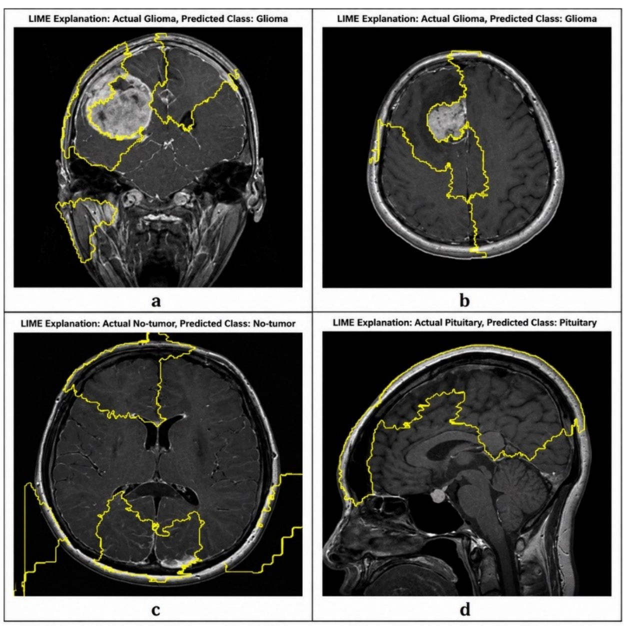

Figure 6 Resnet50 LIME analysis a) meningioma, b) meningioma, c) notumor, d) pituitary

DISCUSSION

Compared with recently published transformer-based approaches such as ViT, DeiT, Swin Transformer, EfficientNetV2, ConvNeXt, and CNN-transformer hybrid models, the evaluated CNN-based architectures remain highly competitive. Transformer models generally benefit from global attention mechanisms and exhibit superior scalability on very large datasets. However, they often require significantly more computational resources and longer training times than conventional CNN-based transfer learning models. ResNet50 achieved 96.9% accuracy while maintaining moderate computational complexity and providing transparent decision explanations using LIME. This indicates that conventional transfer learning architectures still represent practical alternatives for clinical decision support systems where computational efficiency and interpretability are essential.

The proposed models were evaluated using a single publicly available MRI dataset. Although the dataset contains four diagnostic categories, it may not fully represent the variability encountered in clinical environments, including differences in MRI scanners, imaging protocols, patient demographics, and acquisition conditions. Consequently, external validation using multi-center datasets is required before routine clinical implementation. Although data augmentation was applied during preprocessing, certain tumor categories contain fewer samples than others, which may introduce class imbalance and influence classification performance. More advanced sampling techniques or synthetic data generation methods such as GANs could further improve model generalization.

Another important limitation concerns real-world clinical deployment. AI-based decision support systems must satisfy requirements related to robustness, explainability, regulatory approval, interoperability with hospital information systems, and physician acceptance. While LIME improves interpretability by highlighting influential image regions, additional explainability techniques and prospective clinical validation studies remain necessary before deployment in routine diagnostic workflows. Future studies will investigate transformer-based architectures, hybrid CNN-transformer models, external multi-center validation datasets, uncertainty estimation, statistical significance analysis, and multimodal MRI fusion to further improve robustness and clinical applicability.

CONCLUSION

This study highlights the effectiveness of deep learning algorithms in the detection and classification of brain tumors using MRI scans. Among the evaluated models, ResNet50 consistently outperformed others in all performance metrics, while MobileNet achieved competitive accuracy with lower computational requirements, making it suitable for real-time and resource-constrained environments. VGG16 and VGG19 provided balanced results but were less efficient compared to ResNet50, and CNN demonstrated the lowest performance. Importantly, the use of LIME provided interpretability, allowing visualization of the image regions that influenced model decisions, thereby increasing trust in AI-assisted diagnosis.

Despite these promising results, limitations remain. Deep learning models require large, well-annotated datasets, and imbalanced data can impact generalization. Additionally, computational demands and the "black-box" nature of some models pose challenges for clinical integration. Future work should focus on developing lightweight yet robust architectures, enhancing dataset diversity, and incorporating explainable AI techniques to bridge the gap between AI models and clinical practice. Overall, this research demonstrates that deep learning, when combined with interpretability methods, has the potential to transform brain tumor diagnosis and support radiologists in making faster and more accurate clinical decisions.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Ethical standard

The authors have no relevant financial or non-financial interests to disclose.

LITERATURE CITED

Alpsalaz, F., Y. Özüpak, E. Aslan, and H. Uzel, 2025 Classification of maize leaf diseases with deep learning: Performance evaluation of the proposed model and use of explicable artificial intelligence. Chemometrics and Intelligent Laboratory Systems **262**: 105412.

Amin, J., M. Sharif, A. Haldorai, M. Yasmin, and R. S. Nayak, 2022 Brain tumor detection and classification using machine learning: A comprehensive survey. Complex and Intelligent Systems **8**: 3161–3183.

Anantharajan, S., S. Gunasekaran, T. Subramanian, and V. R., 2024 MRI brain tumor detection using deep learning and machine learning approaches. Measurement: Sensors **31**: 101026.

Aslan, E., 2024 LSTM-ESA hibrit modeli ile MR görüntülerinden beyin tümörünün sınıflandırılması. Adıyaman Üniversitesi Mühendislik Bilimleri Dergisi **11**: 63–81.

Aslan, E. and Y. Özüpak, 2024a Advanced skin cancer detection using convolutional neural networks and transfer learning. Middle East Journal of Science **10**: 167–178.

Aslan, E. and Y. Özüpak, 2024b Classification of blood cells with convolutional neural network model. Bitlis Eren Üniversitesi Fen Bilimleri Dergisi **13**: 314–326.

Aslan, E. and Y. Özüpak, 2025a Comparison of machine learning algorithms for automatic prediction of alzheimer disease. Journal of the Chinese Medical Association **88**: 98–107.

Aslan, E. and Y. Özüpak, 2025b Performance comparison of deep learning models in brain tumor classification. Balkan Journal of Electrical and Computer Engineering **13**: 203–209.

Aslan, E., Y. Özüpak, F. Alpsalaz, and Z. M. S. Elbarbary, 2025 A hybrid machine learning approach for predicting power transformer failures using internet of things-based monitoring and explainable artificial intelligence. IEEE Access **13**: 113618–113633.

Bayram, B., I. Kunduracioglu, S. Ince, and I. Pacal, 2025 A systematic review of deep learning in MRI-based cerebral vascular occlusion-based brain diseases. Neuroscience **568**: 76–94.

Brindha, P. G., M. Kavinraj, P. Manivasakam, and P. Prasanth, 2021 Brain tumor detection from MRI images using deep learning techniques. IOP Conference Series: Materials Science and Engineering **1055**: 012115.

Choudhury, C. L., C. Mahanty, R. Kumar, and B. K. Mishra, 2020 Brain tumor detection and classification using convolutional neural network and deep neural network. In *2020 International Conference on Computer Science, Engineering and Applications*.

Kaggle, 2025 Brain tumor classification (MRI). [Online], Accessed: Sep. 16, 2025.

Khan, A. H. *et al.*, 2022a Intelligent model for brain tumor identification using deep learning. Applied Computational Intelligence and Soft Computing **2022**: 8104054.

Khan, M. S. I. *et al.*, 2022b Accurate brain tumor detection using deep convolutional neural network. Computational and Structural Biotechnology Journal **20**: 4733–4745.

Mahmud, M. I., M. Mamun, and A. Abdelgawad, 2023 A deep analysis of brain tumor detection from MR images using deep learning networks. Algorithms **16**: 176.

Maqsood, S., R. Damaševičius, and R. Maskeliūnas, 2022 Multimodal brain tumor detection using deep neural network and multiclass SVM. Medicina **58**: 1090.

Ozdemir, B., E. Aslan, and I. Pacal, 2025 Attention enhanced InceptionNeXt-based hybrid deep learning model for lung cancer detection. IEEE Access **13**: 27050–27069.

Pacal, I., 2024 A novel Swin Transformer approach utilizing residual multi-layer perceptron for diagnosing brain tumors in MRI images. International Journal of Machine Learning and Cybernetics **15**: 3579–3597.

Pacal, I., O. Akhan, R. T. Deveci, and M. Deveci, 2025 NeXtBrain: Combining local and global feature learning for brain tumor classification. Brain Research **1863**: 149762.

Saba, T., A. S. Mohamed, M. El-Affendi, J. Amin, and M. Sharif, 2020 Brain tumor detection using fusion of handcrafted and deep learning features. Cognitive Systems Research **59**: 221–230.

Saeedi, S., S. Rezayi, H. Keshavarz, and S. R. Niakan Kalhori, 2023 MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. BMC Medical Informatics and Decision Making **23**: 1–17.

Sevinc, A., M. Ucan, and B. Kaya, 2025 A distillation approach to transformer-based medical image classification with limited data. Diagnostics **15**: 929.

Soomro, T. A. *et al.*, 2023 Image segmentation for MR brain tumor detection using machine learning: A review. IEEE Reviews in Biomedical Engineering **16**: 70–90.

Yan, H., X. Wang, G. Zhao, L. Ren, and T. Yu, 2025 Brain stimulation for the treatment of epilepsy: Current application and outlook of network neuromodulation. Brain Network Disorders **1**: 7–14.

Yu, L., H. Wang, X. Tang, and K. Yan, 2025 Research progress in predictive techniques for early neurological deterioration in acute ischemic stroke. Brain Network Disorders **1**: 94–97.

Yue, C., Y. Fu, Y. Zhao, Y. Ou, Y. Sun, *et al.*, 2025 Association between alzheimer's disease and metabolic syndrome: Unveiling the role of dyslipidemia mechanisms. Brain Network Disorders **1**: 21–27.

Özüpak, Y., F. Alpsalaz, and E. Aslan, 2025a Air quality forecasting using machine learning: Comparative analysis and ensemble strategies for enhanced prediction. Water, Air, and Soil Pollution **236**: 1–17.

Özüpak, Y., F. Alpsalaz, E. Aslan, and H. Uzel, 2025b Hybrid deep learning model for maize leaf disease classification with explainable AI. New Zealand Journal of Crop and Horticultural Science .

Özüpak, Y. and S. Mansurov, 2025 Optimizing electricity demand forecasting with a novel RNN-LSTM hybrid model. Energy Sources, Part B: Economics, Planning, and Policy **20**.

How to cite this article: Uzel, H., Alpsalaz, F., Özüpak, Y., and Aslan, E. Comparing Explainable Artificial Intelligence and Deep Learning Models for MRI-Based Brain Tumor Diagnosis. *Computers and Electronics in Medicine*, 3(2), 148–155, 2026.