An Advanced XAI Framework for MRI-Based Glioma Classification Using Vision Transformer and CNN-Based Approaches

Hana Charaabi
Laboratory of Advanced Technologies for Medicine and Signals, National Engineering School of Sfax, University of Sfax, Tunisia | Digital Research Center of Sfax, Technopole of Sfax, Tunisia
hana.charaabi@enis.tn (corresponding author)

Ridha El Hamdi
Laboratory of Advanced Technologies for Medicine and Signals, National Engineering School of Sfax, University of Sfax, Tunisia | Digital Research Center of Sfax, Technopole of Sfax, Tunisia
hamdi.ridha@univgb.tn

Mohamed Njah
Laboratory of Advanced Technologies for Medicine and Signals, National Engineering School of Sfax, University of Sfax, Tunisia | Digital Research Center of Sfax, Technopole of Sfax, Tunisia
mohamed.njah@enis.tn

Rachid Jennane
Denis Poisson Institute UMRS CNRS 7013, University of Orleans, Orleans, France
Rachid.Jennane@univ-orleans.fr

Fatma Kolsi
Department of Neurosurgery, Habib Bourguiba University Hospitals, Sfax, Tunisia
Kolsi_fatma@medecinesfax.org

Mohamed Zaher Boudawara
Department of Neurosurgery, Habib Bourguiba University Hospitals, Sfax, Tunisia
boudawarazaher@gmail.com

Received: 31 July 2025 | Revised: 9 September 2025 | Accepted: 19 September 2025

 DOI: https://doi.org/10.48084/etasr.13716

ABSTRACT

This paper presents an advanced eXplainable Artificial Intelligence (XAI) framework for Magnetic Resonance Imaging (MRI)-based glioma classification, designed to bridge the gap between high-performing Deep Learning (DL) models and interpretability in clinical diagnostics. The proposed framework leverages three Convolutional Neural Network (CNN) models (ResNet50V2, EfficientNet-B6, and Xception) and a Vision Transformer (ViT) model to differentiate between Low-Grade Gliomas (LGG) and High-Grade Gliomas (HGG). A Max-Voting strategy integrates the predictions of these models to enhance classification reliability. To promote transparency, the framework incorporates four XAI methods (LIME, SHAP, Integrated Gradients, and GradCAM), providing interpretive insights by highlighting salient features and regions within the MRI scans. Experiments using the BraTS2019 dataset demonstrate that the suggested framework achieves a high classification accuracy of 98.89%, while also offering visual attention maps and interpretative information flow visualizations that support radiologists in clinical decision-making. This work not only advances accurate glioma classification but also emphasizes the role of transparency in AI, highlighting how explainable AI has the potential to revolutionize medical imaging.

Keywords-explainable artificial intelligence (XAI); XAI evaluation metrics; deep transfer learning; vision transformer (ViT); brain tumor diagnosis; MRI; max-voting

I. INTRODUCTION

Gliomas are the most common primary brain tumors, with Glioblastoma Multiforme (GBM) representing the most aggressive form, accounting for more than half of all gliomas and the majority of malignant brain tumors[1]. Accurate and early classification into Low-Grade (LGG) and High-Grade (HGG) gliomas is vital for treatment planning, but remains challenging due to the complexity and variability of MRI features [2]. Manually analyzing MRI images depends mostly on the knowledge of neuroradiologists, which occasionally results in errors or omitted information. Recent advances in Artificial Intelligence (AI), particularly Deep Learning (DL), have proven to be highly effective in automating feature extraction from MRI scans, significantly enhancing diagnostic accuracy and consistency. Computer-Aided Diagnosis (CAD) systems based on DL have shown promise in supporting clinical decision-making. Compared to traditional classification algorithms in medical imaging, DL methods are gaining more attention today due to their ability to automatically extract relevant features [3]. Despite the efficacy of CNN-based transfer learning models, they still face limitations in capturing long-range dependencies and modeling complex tumor characteristics. To overcome this, Vision Transformers (ViT) have gained attention for their ability to process images as sequences and extract global contextual features more effectively than traditional CNNs.

Despite their high performance, DL models are still opaque, making it difficult to understand how they make decisions [4]. This is particularly problematic in the medical field, where clinical trust is crucial. The lack of transparency in how these models arrive at predictions has earned them the label of "black-boxes," limiting their acceptance in critical applications such as glioma diagnosis. To bridge this gap, Explainable Artificial Intelligence (XAI) has gained increasing attention. XAI techniques aim to make model predictions more transparent and interpretable, allowing clinicians to understand the reasoning behind automated decisions [5]. This not only helps identify potential errors but also fosters trust and accountability. Therefore, integrating XAI into medical AI systems is vital for their safe and effective use in clinical practice.

Several recent studies have integrated DL and XAI for glioma classification using MRI. In [6], a comprehensive study investigated the explainability of deep neural networks, employing XAI techniques such as Grad-CAM, Layer-wise Relevance Propagation (LRP), and Occlusion Sensitivity to visualize critical regions influencing CNN predictions. ResNet50-based models achieved an accuracy of 98.62% on the BraTS dataset, underscoring the potential of interpretable AI in clinical settings. In [7], a hybrid Caps-VGGNet model reached 98.9% accuracy, with Capsule Networks offering inherently interpretable part-whole feature representations. In [8], a deep spatio-spatial model achieved 96.98% accuracy by employing class activation maps to visually localize tumor regions, enhancing radiological understanding. In [9], 98.87% accuracy was achieved with a multi-sequence MRI model for glioma grading, incorporating a post-hoc decision tree to approximate CNN behavior, enabling rule-based interpretability. In [10], a multimodal classification system combined CNNs and SVMs, reaching 98.5% accuracy, and using Grad-CAM to provide saliency maps that justified the model's tumor classifications. In [11], a lightweight CNN achieved 92.85% accuracy in binary glioma classification. Although XAI was not central to this study, the future integration of methods such as SmoothGrad and Integrated Gradients was emphasized to enhance transparency. In [12], a brain tumor classification model used deep transfer learning and ensemble stacking, achieving very high accuracy (~98.6%) using seven pre-trained CNN architectures fine-tuned on the BraTS19 dataset.

This study introduces a practical and interpretable DL framework for glioma classification using MRI scans. The proposed framework strategically combines ViT with CNNs and integrates XAI methods. Evaluated on the BraTS2019 dataset [13], the proposed approach not only delivers strong classification performance but also enhances radiologists' trust and understanding of AI predictions. The key contributions include:

- The introduction of an innovative framework that successfully integrates XAI techniques, enhancing the interpretability of ViT and CNN models.
- The achievement of high classification accuracy while fostering transparency in model predictions through visual attention maps and information flow visualizations.
- A comprehensive comparative analysis of multiple XAI methods, emphasizing their strengths and limitations in the context of glioma diagnosis. This includes an objective and multi-dimensional evaluation using quantitative metrics.

II. PROPOSED FRAMEWORK FOR MRI GLIOMA TUMOR CLASSIFICATION

This study introduces a highly precise and fully automated MRI classification framework, specifically designed to differentiate between two primary categories of gliomas: HGG and LGG. The framework leverages advanced DL models and XAI techniques to facilitate interpretable decision-making processes. Figure 1 shows a diagram that outlines the proposed framework, which basically involves the following tasks:

- Preprocessing involves the projection of input images, normalization, resizing, and data augmentation.
- Classification: Pre-trained models were used on the large-scale ImageNet Dataset. Then, a meticulous fine-tuning process was undertaken to determine the optimal hyperparameters, ensuring the highest level of performance. The Max-Voting technique in ensemble learning combines predictions from the CNN and the ViT models.
- Explanation: Integrates four post-hoc methods, two based on gradient (GradCAM and Integrated Gradients), and two based on perturbation (LIME and SHAP).

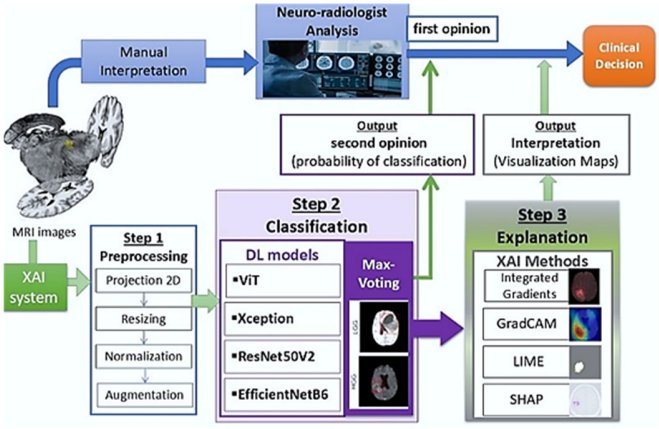

Fig. 1. Architecture of the proposed XAI framework.

A. *Preprocessing of the MRI Dataset*

Preprocessing is a critical step that involves image projection, normalization, resizing, and data augmentation. These steps collectively contribute to the creation of a well-structured and standardized dataset, ultimately enhancing the effectiveness and robustness of the model. This phase involves:

- 2D projection: Following [14], the original 3D MRI scans (NIfTI format) were converted into 2D PNG images. The 155 slices from each scan were reduced to representative 2D images using the Maximum Intensity Projection (MIP) technique. Each projection is composed of five (5) consecutive slices, preserving almost all relevant tumor information. This approach enables the use of powerful 2D CNN architectures pre-trained on large-scale datasets.
- Intensity normalization: The obtained data may be regarded as highly diverse since it was gathered using several scanning methods and acquisition protocols. A straightforward Min-Max normalization was used for this.
- Data augmentation: Data augmentation strategies were explored to alleviate the overfitting issue and expand the dataset size to enhance the performance of pre-trained models. The most compatible data augmentation technique for brain images is horizontal flipping.

B. *CNNs and ViT-Based Classification with Max-Voting*

After conducting a comprehensive comparative analysis, which involved evaluating various pre-trained models, considering the latest advances in the field, three models were selected: ResNet50V2, EfficientNet-B6, and Xception. Deep Transfer Learning (DTL) enables applying pre-learned weights and parameters from models trained on rich datasets such as ImageNet. Xception [15] is a DL model designed for efficiency and high performance. It replaces standard convolutions with depthwise separable convolutions, reducing the number of parameters while improving feature extraction. With 36 convolutional layers and residual connections to support better gradient flow, it is both lightweight and powerful. Its use of global average pooling further streamlines the architecture. EfficientNet-B6 [16] is a high-performing yet efficient DL model, designed to balance accuracy and computational cost. Built with compound scaling, it adjusts depth, width, and resolution in harmony, allowing it to capture detailed features across different levels. Its architecture includes depthwise separable convolutions, inverted residual blocks, and the Swish activation function, making it powerful and lightweight. ResNet50V2 [17] is an improved version of the original ResNet architecture, specifically designed to address issues such as vanishing gradients when training deep networks. It introduces a refined residual block structure with three convolutional layers (a bottleneck block), allowing for more stable and efficient training, even in very deep models. Using shortcut connections and residual learning, the model maintains a strong gradient flow, which is crucial for deep feature extraction.

To complement CNN models, the proposed framework strategically integrates a ViT, which significantly enhances both feature extraction and classification performance. While CNNs primarily rely on local receptive fields and hierarchical pooling to capture features, ViT adopts a fundamentally different approach. It partitions each MRI slice into fixed-size, non-overlapping patches, which are then linearly embedded and processed as a sequence of tokens. These tokens are subsequently fed into a series of transformer encoder layers, where Multi-Head Self-Attention (MHSA) mechanisms are central. Unlike the localized operations of CNNs, MHSA allows ViT to model long-range dependencies across the entire image. This global attention mechanism inherently preserves spatial structure without the information loss typically associated with successive convolutional and downsampling operations in CNNs. Consequently, ViT is uniquely positioned to capture both global and fine-grained local tumor characteristics more effectively, offering a richer contextual understanding of MRI data [18]. In this study, each MRI slice was divided into 16×16 patches, linearly embedded in a 768-dimensional space, and processed through 12 transformer encoder layers with MHSA. The model was fine-tuned using the Adam optimizer with an initial learning rate of 3e-5 and a dropout rate of 0.1 to prevent overfitting.

In the proposed framework, combining the ViT with three advanced CNN models boosts the system's ability to detect complex patterns in MRI scans. The Max-Voting technique was applied to further improve reliability, where each model votes independently, and the most frequent prediction determines the final output [19].

C. *XAI Module of the Proposed Framework*

The XAI module in the proposed framework generates visual explanations that highlight the key regions of MRI scans influencing the classification model's decisions [20]. Four state-of-the-art XAI methods were employed, as follows:

- Integrated Gradients is a gradient-based attribution method that addresses the saturation problem in saliency maps. It computes the average of the gradients as the input varies along a linear path from a baseline (usually a black image) to the actual input [21].

$$IG(x_i) = (x_i - x_i') \times \int_0^1 \partial F(x' + \alpha(x - x')) / \partial x_i \, d\alpha \quad (1)$$

where x is the input image, x' is the baseline input, F is the model's output function, $\alpha \in [0,1]$ is the interpolation scalar, and $\partial F/\partial x_i$ denotes the gradient of the output.

- Grad-CAM (Gradient-weighted Class Activation Mapping) generates coarse localization heatmaps by propagating gradients of the target class back to the final convolutional layer and weighting its activation maps accordingly [22]. This method effectively highlights tumor regions (in warm colors) and suppresses background noise.

$$G^c = ReLU(\Sigma_k \delta_k{}^c \times A_k) \quad (2)$$

where $\delta_k{}^c$ is the weight assigned to the feature map A_k.

- LIME (Local Interpretable Model-agnostic Explanations) provides local interpretability by approximating the complex model with a simpler, interpretable one near the input sample [23]. It perturbs the image by masking superpixels and learns a linear model to mimic the classifier's behavior locally.

$$\varphi(x) = argmin_{(}g \in G_{)} [\, L(f, g, \pi_x) + \Omega(g) \,] \quad (3)$$

where G is the set of interpretable models, L is a loss function ensuring fidelity between the black-box model f and the interpretable model g, π_x defines the locality, and $\Omega(g)$ penalizes model complexity.

- SHAP (SHapley Additive exPlanations) assigns feature importance scores based on Shapley values from cooperative game theory [24]. It provides fine-grained pixel-wise explanations, offering deeper insights into model decisions and enhancing transparency.

$$g(z') = \varphi_0 + \Sigma_1{}^M \varphi_i \times z'_i \quad (4)$$

where φ_i are the SHAP values, z' is a binary feature, and φ_0 is the model output for a baseline input.

III. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the results, assessing the effectiveness of the proposed pre-trained model and integrated XAI techniques. In addition, a comprehensive discussion is provided with a comparative analysis of recent advances.

A. Dataset Structure

This study used the BraTS 2019 MRI dataset, which comprises multi-modal brain scans (T1, T2, T1-contrast, and FLAIR) at a resolution of 240×240×155 voxels. It includes 76 LGG and 259 HGG cases, each annotated with classification labels [13]. This imbalance between HGG and LGG could bias the classification process. To alleviate this issue, extensive data augmentation was applied to increase the effective number of LGG samples, and class-balanced sampling was used during training to reduce bias toward the majority class, ensuring fairer model performance across both glioma grades.

B. Classification and XAI Evaluation Metrics

The performance of the classification models was evaluated using various metrics, including Accuracy, Precision, Recall, and F1-score, offering a thorough assessment of the model's effectiveness.

The evaluation metrics of XAI methods assess how well the explanations align with the model's decision-making process. They assess whether the explanations are sufficiently detailed for understanding and evaluate their consistency across different decision instances. The key XAI evaluation metrics are:

- The Perturbation Stability Index (PSI) measures the stability of explanations under slight input variations.
- Fidelity measures how closely the provided explanation aligns with the model's actual decision-making process, with a higher fidelity indicating a better match.
- Sufficiency assesses whether the explanation contains enough information for users to understand why the model made a specific decision.
- Consistency evaluates if the model consistently provides similar explanations for similar inputs, promoting trust in the model's behavior.

C. Evaluation of the Proposed Models

Figure 2 illustrates the evolution of accuracy as a function of the number of epochs during the training and validation phases. All models quickly reach a high accuracy rate, close to 99%, from the first iterations, which demonstrates good learning capacity. The ViT model is distinguished by a more spread-out learning curve over time, requiring more epochs to stabilize its performance, but it reaches a level of accuracy comparable to the other models at the end of training.

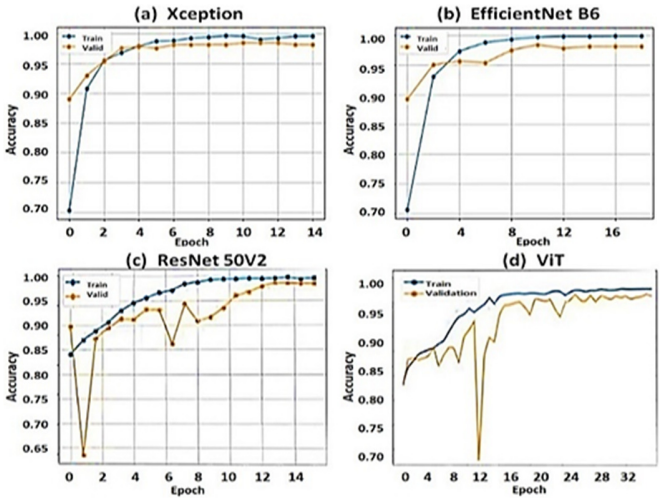

Fig. 2. Evolution of accuracy over the number of epochs, (a): Xception, (b): EfficientNetB6, (c): ResNet50V2, and (d): ViT.

TABLE I. EVALUATION OF THE PROPOSED FRAMEWORK

Models	Accuracy	Precision	Recall	F1-score
Xception	98.32%	98.83%	99.06%	98.94%
EfficientNetB6	96.64%	98.57%	97.18%	97.87%
ResNet50V2	97.37%	98.57%	98.10%	98.34%
ViT	96.27%	98.33%	96.95%	97.64%
MaxVoting	98.89%	99.07%	99.53%	99.30%

Table I presents a summary of the classification performance achieved by various approaches on the BraTS2019 dataset. The Max-Voting ensemble achieved the highest accuracy, reaching 98.89%. Among the individual models, Xception demonstrated the strongest performance with an accuracy of 98.32%, while the Vision Transformer (ViT) model showed a slightly lower, though still respectable, accuracy of 96.27%. The superior performance of the Max-Voting approach underscores the synergistic advantages of combining diverse DL architectures.

D. Comparative Study

Table II illustrates that many DTL models achieve high accuracy. The proposed framework, which leverages a Max-Voting ensemble on the BraTS2019 dataset, achieves an outstanding accuracy of 98.89%. This performance not only demonstrates the effectiveness of the proposed approach but also positions it favorably among state-of-the-art results reported in the literature, surpassing the accuracies of most individual models.

TABLE II. COMPARISON TO STATE-OF-THE-ART WORKS

Study	Model	Dataset	Accuracy (%)
[6]	ResNet50	BraTS2019	98.62
[7]	CapsVGGNet	BraTS2019	98.9
[8]	ResNet	BraTS2018	96.98
[9]	EfficientNetB0	BraTS2019	98.87
[10]	Densenet201	BraTS2018	98.8
[11]	SVM	BraTS2018	84.2
[12]	VGG16	BraTS2019	95.16
	VGG19		93.68
	MobileNet		93.71
	InceptionV3		91.78
	Xception		90.39
	InceptionResV2		95.45
	DenseNet121		91.97
	Stacking Model		98.06
Proposed framework	Xception	BraTS2019	98.32
	EfficientNetB6		96.64
	ResNet50V2		97.37
	ViT model		96.27
	Max-voting		**98.89**

E. XAI Visualization Maps

Distinct colormaps were used for the different interpretability methods to help visualize model decisions, thereby enhancing trust in classifier predictions. Integrated Gradients uses a "hot" colormap (black-to-red-yellow-white) to highlight important pixels, where lighter colors indicate stronger feature attributions accumulated along a path from a baseline. Grad-CAM applies warm-to-cool color maps to localize critical areas influencing predictions. LIME produces binary black-and-white masks to highlight key superpixels, offering an intuitive region-based explanation. SHAP visualizes contributions using red and blue shades, reflecting positive and negative feature impacts. These diverse visual cues allow clinicians and researchers to verify whether the model focuses on medically relevant areas, thus increasing confidence in its diagnostic reliability.

Figure 3 compares the visual explanations generated by the XAI methods on MRI scans with HGG and LGG. Although all methods succeeded in identifying tumor regions in HGG cases, their performance varied in clarity and localization. Grad-CAM offers smooth, globally consistent heatmaps, particularly effective in HGG, whereas Integrated Gradients provides fine-grained, dense saliency maps across both tumor types. LIME yields sharp but sometimes overly simplistic explanations, and SHAP produces scattered attributions with limited spatial precision. In LGG cases, the differences become more evident due to the tumors' subtle appearance. Grad-CAM emerged as the most reliable technique for visual interpretability, balancing precision and clinical relevance. These findings underscore the importance of selecting appropriate XAI tools for effective deployment in medical imaging.

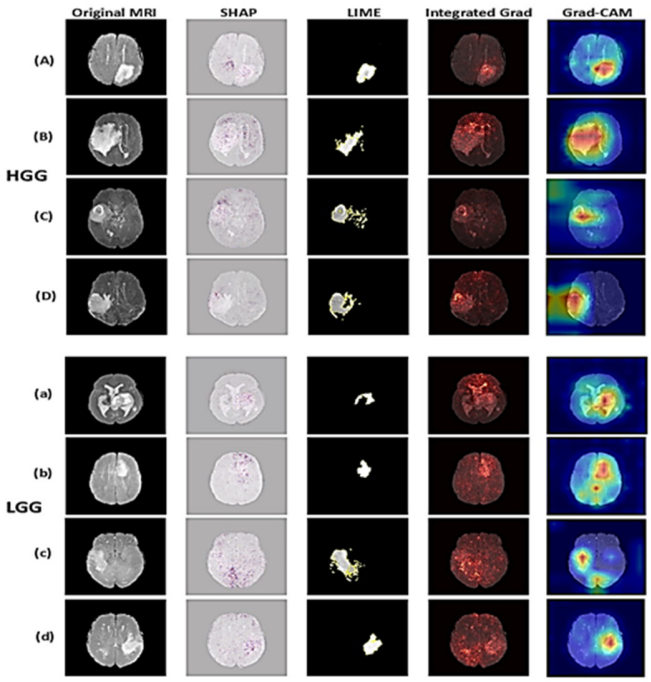

Fig. 3. Examples of XAI visualization images for the classification of gliomas in the brain using MRI. The first four rows provide sensitivity maps for HGG cases, whereas the final four rows display sensitivity maps for LGG cases. Left to right shows MRI images, SHAP, LIME, Integrated Gradients, and Grad-CAM.

Table III shows the performance of the evaluated XAI methods. Grad-CAM achieved the highest overall scores, demonstrating strong stability (92.15%), sufficiency (88.24%), fidelity (95.95%), and consistency (93.86%), but its visual explanations remain relatively coarse and may miss small or subtle tumor regions. Integrated Gradients generates more fine-grained saliency maps, offering improved localization, but its sensitivity to noise and limited robustness under input perturbations can reduce its reliability. LIME offers intuitive and localized explanations, but struggles with high-resolution MRI images due to its reliance on superpixel segmentation. SHAP produces detailed attributions with a clear distinction between positive and negative contributions.

TABLE III. EVALUATION OF XAI METHODS

XAI methods	PSI	Sufficiency	Fidelity	Consistency
Grad-CAM	92.15%	88.24%	95.95%	93.86%
Integrated Gradients	87.24%	82.90%	88.71%	85.33%
LIME	85.50%	85.06%	90.11%	88.58%
SHAP	84.81%	80.19%	86.35%	82.01%

IV. CONCLUSION

This study presented a powerful and interpretable DL framework for glioma classification from brain MRI scans, combining the strengths of multiple pretrained models (Xception, ResNet50V2, EfficientNet-B6, and ViT) with advanced XAI techniques, such as Grad-CAM, Integrated Gradients, LIME, and SHAP. By selectively applying LIME and SHAP and optimizing their computation load, the framework minimizes overhead. Meanwhile, combining Grad-CAM with Integrated Gradients improves visual clarity and reduces noise. By projecting 3D MRI onto optimized 2D representations and applying a max-voting strategy across models, the classification accuracy reached 98.89%. This ensemble approach not only improves diagnostic performance but also enhances reliability by leveraging the complementary strengths of diverse architectures. This integration of XAI bridges the gap between high-performing black-box models and the need for clinical transparency. The visual explanations generated by the proposed framework help radiologists understand, validate, and trust the model's decisions.

The clinical accuracy and relevance of the model's predictions, along with the interpretability of the XAI visualizations, were validated by Dr. Fatma Kolsi of the Department of Neurosurgery at Habib Bourguiba University Hospital, Sfax, Tunisia, using a local hospital database. Future work will include systematic validation on additional datasets and across multiple institutions. Unlike previous work limited to visual illustrations, this study conducts a rigorous metric-based evaluation of XAI methods, highlighting the strengths and weaknesses of each method. This comparison provides practical insights for selecting XAI tools in clinical diagnostics.

REFERENCES

[1] M. Weller *et al.*, "EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood," *Nature Reviews Clinical Oncology*, vol. 18, no. 3, pp. 170–186, Mar. 2021, https://doi.org/10.1038/s41571-020-00447-z.

[2] M. Zhou *et al.*, "Radiomics in Brain Tumor: Image Assessment, Quantitative Feature Descriptors, and Machine-Learning Approaches," *American Journal of Neuroradiology*, vol. 39, no. 2, pp. 208–216, Feb. 2018, https://doi.org/10.3174/ajnr.A5391.

[3] M. Rasool, A. Noorwali, H. Ghandorh, N. A. Ismail, and W. M. S. Yafooz, "Brain Tumor Classification using Deep Learning: A State-of-the-Art Review," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16586–16594, Oct. 2024, https://doi.org/10.48084/etasr.8298.

[4] Z. C. Lipton, "The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery.," *Queue*, vol. 16, no. 3, pp. 31–57, June 2018, https://doi.org/10.1145/3236386.3241340.

[5] Q. Xue and M. C. Chuah, "Explainable deep learning based medical diagnostic system," *Smart Health*, vol. 13, Aug. 2019, Art. no. 100068, https://doi.org/10.1016/j.smhl.2019.03.002.

[6] R. A. Zeineldin *et al.*, "Explainability of deep neural networks for MRI analysis of brain tumors," *International Journal of Computer Assisted Radiology and Surgery*, vol. 17, no. 9, pp. 1673–1683, Sept. 2022, https://doi.org/10.1007/s11548-022-02619-x.

[7] A. Jabbar, S. Naseem, T. Mahmood, T. Saba, F. S. Alamri, and A. Rehman, "Brain Tumor Detection and Multi-Grade Segmentation Through Hybrid Caps-VGGNet Model," *IEEE Access*, vol. 11, pp. 72518–72536, 2023, https://doi.org/10.1109/ACCESS.2023.3289224.

[8] S. Chatterjee, F. A. Nizamani, A. Nürnberger, and O. Speck, "Classification of brain tumours in MR images using deep spatiospatial models," *Scientific Reports*, vol. 12, no. 1, Jan. 2022, Art. no. 1505, https://doi.org/10.1038/s41598-022-05572-6.

[9] Z. Khazaee, M. Langarizadeh, and M. E. S. Ahmadabadi, "Developing an Artificial Intelligence Model for Tumor Grading and Classification, Based on MRI Sequences of Human Brain Gliomas," *International Journal of Cancer Management, Research*, Oct. 2021, Art. no. 15.

[10] 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.

[11] Q. U. Ain, I. Duaa, K. Haroon, F. Amin, and M. Z. ur Rehman, "MRI Based Glioma Detection and Classification into Low-grade and High-Grade Gliomas," in *15th International Conference on Open Source Systems and Technologies (ICOSST)*, Sept. 2021, pp. 1–5, https://doi.org/10.1109/ICOSST53930.2021.9683838.

[12] H. E. Hamdaoui *et al.*, "High precision brain tumor classification model based on deep transfer learning and stacking concepts," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 1, Oct. 2021, Art. no. 167, https://doi.org/10.11591/ijeecs.v24.i1.pp167-177.

[13] S. Bakas *et al.*, "Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge." arXiv, Apr. 23, 2019, https://doi.org/10.48550/arXiv.1811.02629.

[14] K. H. Kim and H. W. Park, "A fast progressive method of maximum intensity projection," *Computerized Medical Imaging and Graphics*, vol. 25, no. 5, pp. 433–441, Sept. 2001, https://doi.org/10.1016/S0895-6111(01)00003-9.

[15] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, July 2017, pp. 1800–1807, https://doi.org/10.1109/CVPR.2017.195.

[16] A. U. I. Rafid, S. Sanjana, M. B. Munir, and N. Sharmin, "An early-stage diagnosis of diabetic retinopathy based on ensemble framework," *Signal, Image and Video Processing*, vol. 18, no. 1, pp. 735–749, Feb. 2024, https://doi.org/10.1007/s11760-023-02796-5.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks," in *Computer Vision – ECCV 2016*, 2016, vol. 4, pp. 630–645, https://doi.org/10.1007/978-3-319-46493-0_38.

[18] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," presented at the International Conference on Learning Representations, 2021.

[19] G. S. Tandel, A. Tiwari, and O. G. Kakde, "Performance optimisation of deep learning models using majority voting algorithm for brain tumour classification," *Computers in Biology and Medicine*, vol. 135, Aug. 2021, Art. no. 104564, https://doi.org/10.1016/j.compbiomed.2021.104564.

[20] H. Charaabi, H. Mzoughi, R. Hamdi, and M. Njah, "EXplainable Artificial Intelligence (XAI) for MRI brain tumor diagnosis: A survey," in *2023 International Conference on Cyberworlds (CW)*, July 2023, Sousse, Tunisia, pp. 171–178, https://doi.org/10.1109/CW58918.2023.00033.

[21] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic Attribution for Deep Networks," in *Proceedings of the 34th International Conference on Machine Learning*, July 2017, pp. 3319–3328.

[22] 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," in *2017 IEEE International Conference*

on Computer Vision (ICCV), Venice, Italy, Oct. 2017, pp. 618–626, https://doi.org/10.1109/ICCV.2017.74.

[23] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, May 2016, pp. 1135–1144, https://doi.org/10.1145/2939672.2939778.

[24] S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, 2017, vol. 30, pp. 4765–4774.