ISSN : 3105-0794

الهيئة الليبية للبحث العلمي

Libyan Authority For Scientific Research

Libyan International Journal of Natural Sciences (LIJNS)

VOLUME 2 ISSUE 1 MAY 2026

Bi-Annual, Peer- Reviewed, and Open Accessed e-Journal

DOLJ Directory of Online Libyan Journals

turnitin

Google Scholar

Puplished by

MILJS@aonsrt.ly

Libyan International Journal of Natural Sciences (LIJNS)
Vol. 2 Issue No 1 May 2026

Comparative Evaluation of ResNet50 and MobileNetV2 Enhanced with Explainable AI for Brain Tumor MRI Classification

* Abdelkader Alrabai

Physics Department, Education Faculty, Wadi Alshatti University, Wadi Alshatti, Libya

ARTICLE INFORMATION

Article history:
Received 6 *January 25*
Accepted 2 March 26
Published 10 *March 2026*

Keywords:
Brain Tumor
Classification
CNN
Grad-CAM
MRI

ABSTRACT

Accurate identification of brain malignancies via MRI is a cornerstone of effective neuro-oncological intervention. While deep learning offers high-speed automated analysis, the unclear reasoning mechanisms of these algorithms often limit their adoption in clinical environments. This study presented a comparative evaluation of two distinct convolutional neural network (CNN) architectures—ResNet50 and MobileNetV2—to determine the optimal balance between classification depth and computational efficiency. Both models were evaluated on an MRI dataset to distinguish between healthy and tumor-containing scans, with class weighting applied to mitigate the effects of class imbalance. Experimental results demonstrate high performance across both architectures. ResNet50 achieved an accuracy of 95.47%, supported by a precision of 95.16%, a recall of 95.44%, and an F1-score of 95.29%. In comparison, the lightweight MobileNetV2 reached an accuracy of 94.80%, with precision, recall, and F1-score recorded at 94.33%, 95.00%, and 94.63%, respectively. To address the requirement for medical transparency, Grad-CAM was integrated to generate visual evidence for each prediction. These heatmaps successfully localized tumor regions, aligning internal model logic with radiological findings. By benchmarking a high-capacity residual network against an efficient, mobile-ready architecture, this study confirms that both frameworks, when paired with explainability tools, provide a reliable secondary screening layer. This integration of high-performance classification with visual interpretability offers a clear path toward deploying trustworthy diagnostic tools in clinical workflows.

©Author(s) 2026. This article is distributed under the terms of the CC BY-NC 4.0.

1. Introduction

Brain tumors are a major contributor to global illness and death, accounting for approximately 90% of all adult cancers affecting the central nervous system [1]. A brain tumor refers to an abnormal mass of tissue that can develop in any structure within the skull, such as the brain itself, cranial nerves, meninges, skull, pituitary gland, or pineal gland. These tumors are typically classified based on the type of cells from which they originate. They can be primary, meaning they begin in the brain, or secondary, meaning they have metastasized to the brain from other parts of the body [2]. Several brain imaging techniques are employed in the diagnosis of brain tumors, helping to determine their size, location, shape, and other characteristics. Commonly used scanning methods include MRI, CT, and PET. Among these, MRI is the most widely used due to its non-invasive nature and its ability to provide excellent contrast in soft tissues. As a result, MRI is commonly found in diagnostic clinics and plays a key role in identifying brain tumors [3]. Brain tumor classification is a fundamental component of radiological assessment. Categorization is traditionally performed through clinical inspection or via automated diagnostic frameworks. While human-led interpretation is labor-intensive and prone to inter-observer variability, it remains the established clinical benchmark. Conversely, computational imaging systems provide substantial utility by refining the precision and throughput of tumor identification [4]. Computer-Aided Diagnosis (CAD) systems significantly assist radiologists in rapidly diagnosing brain tumors, contributing to a reduction in brain cancer-related mortality. The primary objective of CAD systems is to automate the detection of brain tumors in medical images with high

* Corresponding author: *E*-mail addresses: a.alrabai@wau.edu.ly

accuracy and dependability [5]. Recent advances in smart healthcare, particularly the use of AI in radiology, have made significant strides in automating image-recognition tasks to detect various objects within imaging data. AI techniques excel at automatically identifying complex patterns in medical images. A wide array of AI-driven applications has rapidly advanced the field. Among these, deep learning-based classification and segmentation are regarded as the most effective approaches for detecting and extracting features from MRI images [6]. Deep learning is frequently described as a black box due to its complex and opaque decision-making processes. There is growing concern across many fields that these models might contain hidden biases that remain undetected. This issue is particularly critical in medical applications, where unnoticed bias can lead to serious consequences. As a result, there is an increasing demand for methods that provide greater insight into how these models operate. These methods are generally known as interpretable deep learning or explainable AI [7]. Providing explanations for a model's output is essential, especially in precision medicine, where specialists need more detailed information than just a basic yes-or-no prediction to effectively support their diagnoses [8].

Accurate and early identification of brain tumors plays a vital role in guiding effective treatment and improving patient prognosis. While visual assessment of MRI scans by specialists remains dependable, it is often time-consuming and may vary between observers. In this work, deep learning techniques are investigated for automated binary classification of brain MRI images into normal (healthy) and abnormal categories, using ResNet50 and MobileNetV2 architectures. To better understand the basis of the model predictions, Grad-CAM was applied to reveal image regions that most strongly influenced the classification outcomes. The study presents a comparative evaluation of the two employed models for binary brain MRI classification and demonstrates how interpretability methods can be integrated into diagnostic workflows. This integration enhances transparency and fosters trust, supporting the development of reliable and clinically applicable diagnostic systems.

2. Related Works

Numerous studies have employed CNNs for brain MRI classification tasks, showcasing their effectiveness in medical image analysis. Some of these studies have tackled multi-class classification problems, aiming to differentiate among various tumor types. Others have concentrated on binary classification, where the objective is to determine the presence or absence of abnormal brain tissue. Since the current study is designed as a binary classification task, the related works reviewed will primarily emphasize research that has adopted a similar binary classification approach.

Khan et al. [9] developed deep learning models for both binary and multiclass brain tumor classification using MRI data. A 23-layer CNN was applied to the larger dataset, while overfitting on the smaller dataset was addressed by integrating transfer learning with the VGG16 architecture. Mijwil et al. [10] investigated the use of a deep learning approach for brain disease detection based on MRI data. In their study, the MobileNetV1 architecture was applied to a dataset consisting of 1,265 images. The experimental analysis demonstrated that the model achieved a classification accuracy exceeding 97%, demonstrating its effectiveness for automated brain abnormality detection. Badjie and Ülker [11] employed a transfer learning strategy based on the AlexNet CNN for automated brain tumor analysis. Their approach demonstrated high reliability, achieving an overall classification accuracy of 99.62%. The study also reported the capability of the proposed framework to identify and categorize tumors across different stages and sizes, highlighting its potential contribution to improving efficiency and consistency in clinical diagnostic workflows. Younis et al. [12] examined the utilization of VGG16-based CNNs for brain tumor detection from MRI data. Their study evaluated a CNN framework trained on 253 brain MRI images and reported strong performance, with VGG16 achieving an accuracy of 98.5%, outperforming conventional detection methods. Almadhoun and Abu-Naser [13] investigated deep learning approaches for brain tumor detection using MRI images. Their study evaluated multiple architectures, including VGG16, ResNet50, MobileNet, and InceptionV3, on a dataset of 10,000 images. The reported results showed strong performance across models, with InceptionV3 and VGG16 achieving the highest F-score accuracies. Anand et al. [14] proposed a weighted ensemble deep learning framework for brain tumor classification using MRI data. Their approach combined feature representations from VGG19 and two CNN models, evaluated on a TCGA lower-grade glioma dataset, and demonstrated improved performance over individual models by effectively reducing overfitting. Gurunathan and Krishnan [15] introduced an automated framework for brain tumor analysis from MRI images that integrates preprocessing, classification, and segmentation stages. Their method employed CNN-based classification with data augmentation and morphological segmentation to identify tumor regions and assess severity, with performance evaluated on public datasets and validated by expert radiologists.

Previous studies demonstrate substantial progress in the use of deep learning techniques for brain tumor detection and classification from MRI data. Transfer learning approaches have been widely adopted, particularly to mitigate challenges related to limited datasets and model overfitting. Several investigations report that, when trained on sufficiently large and well-labeled datasets, deep learning models can achieve performance comparable to, or exceeding, that of experienced clinicians. Despite these advances, limited model transparency remains a key concern, as many approaches offer little insight into the basis of their predictions, which can hinder clinical adoption. To address this limitation, visualization methods such as Grad-CAM have been increasingly applied to identify image regions that contribute most to classification outcomes. Motivated by these findings, the present study examines binary brain tumor classification using ResNet50 and MobileNetV2

architectures, incorporating Grad-CAM to support visual interpretation of model decisions. By jointly considering predictive performance and explainability, this work responds to the growing emphasis on interpretable AI solutions in medical imaging and strengthens their potential for clinical integration.

3. Methodology

This work introduces a deep learning–driven framework for binary classification of brain MRI scans, distinguishing normal cases from those exhibiting tumor-related abnormalities. The proposed approach is organized into a sequence of well-defined stages, with the complete workflow illustrated in Fig. 1 to clarify the structure of the pipeline and the interaction between its components.

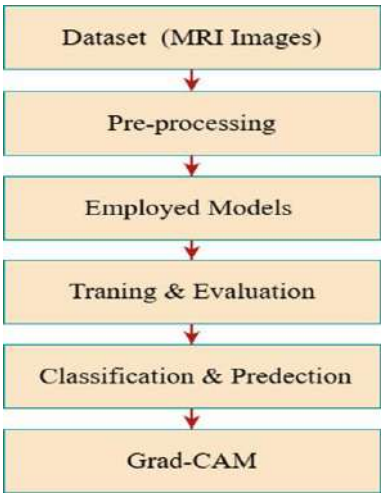

Fig. 1: Method summary

3.1 Dataset and Preprocessing

The dataset used in this study comprises brain MRI images divided into two categories: healthy (normal) images, which include scans without any visible signs of brain tumors, and tumor (abnormal) images, containing scans that clearly exhibit brain tumors. This dataset, sourced from Kaggle [16], consists of a total of 5,000 images, with 2,000 belonging to the healthy class and 3,000 to the tumor class. Fig. 2 illustrates representative samples from both categories, providing a visual overview of the dataset used.

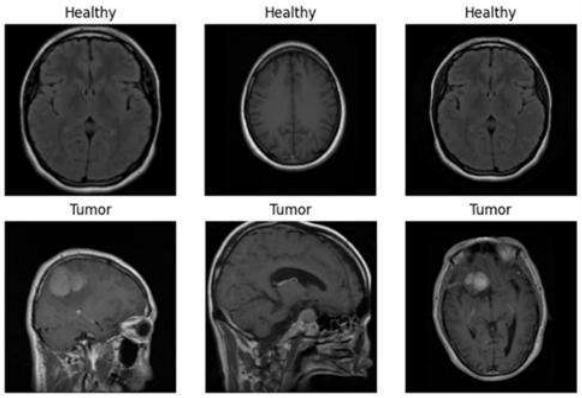

Fig. 2: Representative sample.

The dataset was divided into training, validation, and testing sets using a 70%, 15%, and 15% distribution, respectively. This configuration was chosen to balance effective model training along with reliable performance evaluation. Assigning 70% of the data to training allows the models to learn representative features adequately, while reserving 15% for validation supports hyperparameter adjustment and early stopping. The remaining 15% was held out for testing to provide an independent assessment of generalization capability.

All images were uniformly resized to 224×224 pixels to ensure consistency in input dimensions. Additionally, pixel values were normalized to standardize the intensity range across the dataset.

3.2 Employed Models

Pretrained CNN models, specifically ResNet50 and MobileNetV2, were adopted in this work. These architectures, initially trained on the ImageNet dataset, were employed as feature extractors for the binary classification of brain MRI images. By building upon the representations learned during pretraining, the networks were subsequently fine-tuned to differentiate normal brain scans from those containing tumors. This adaptation highlights the suitability of transfer learning approaches for this task.

ResNet50 [17] is a deep CNN designed around the concept of residual learning, which helps alleviate the vanishing gradient issue commonly observed in very deep architectures. The network introduces shortcut connections that enable information to pass directly across layers, facilitating stable and efficient training as depth increases. Comprising 50 layers, ResNet50 is capable of capturing high-level and intricate visual patterns and has been successfully employed across a wide range of image analysis applications, including medical imaging. Owing to its reliability and strong representational capacity, it is frequently selected for challenging tasks such as brain tumor detection and classification. MobileNetV2 [18] is a lightweight CNN specifically designed for efficient deployment on mobile and embedded devices. Its architecture introduces the concept of inverted residual blocks with linear bottlenecks, which allows the network to maintain accuracy while reducing computational cost. By connecting thin layers directly to each other and expanding only in intermediate layers, MobileNetV2 captures essential features without the redundancy often seen in conventional deep networks. This balance of efficiency and representational power makes it particularly suitable for tasks like medical image classification, where both performance and computational resources are critical considerations.

3.3 Training and Evaluation

Both models were trained using a carefully selected set of hyperparameters to ensure optimal performance. An initial learning rate of 0.0001 was applied, with a batch size of 32, and the Adam optimizer was employed to efficiently update the network weights. Training was carried out for 50 epochs, giving the models sufficient opportunity to learn from the

dataset. Class weighting was also applied to address any imbalance between classes. The same hyperparameter settings were used for both models to allow an unbiased comparison. After training, their performance was evaluated on the test set using common evaluation metrics.

3.4 Explainable AI

To improve interpretability and support clinical confidence, Grad-CAM was employed to provide visual explanations for the MRI images. The resulting heatmaps highlight the spatial regions that most influenced the model's predictions. Grad-CAM [19] is a technique that produces heatmaps to highlight the regions of an image most influential in the prediction. This approach improves interpretability and reliability, which is especially important in sensitive applications such as medical diagnostics.

4. Results and Discussions

This section summarizes the experimental findings from applying ResNet50 and MobileNetV2 to the binary classification of brain MRI scans, differentiating between normal and tumor cases. Fig. 3 depicts the training progression of both models across 50 epochs, showing trends in training and validation accuracy as well as loss. These plots offer a clear view of how each model learned over time and highlight the consistency of their performance throughout the training process.

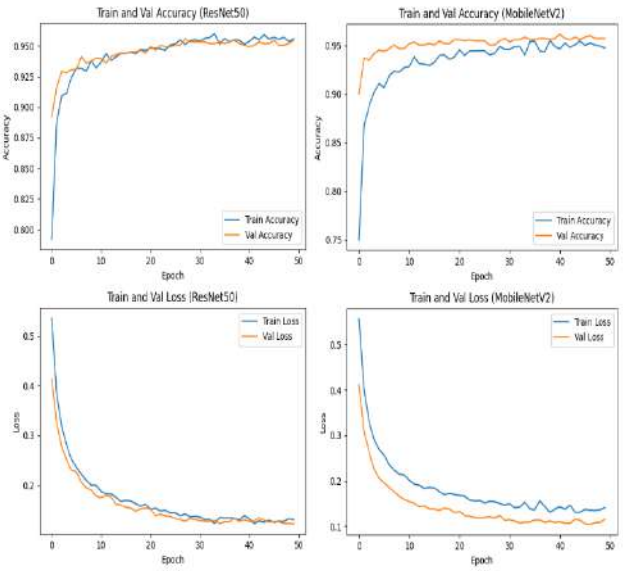

Fig. 3: Accuracy and loss curves for both models

The evaluation results across all common metrics for both models are summarized in Table 1.

Table 1: Evaluation metrics results.

Metrics	ResNet50	MobileNetV2
Accuracy	0.9547	0.9480
Precision	0.9516	0.9433
Recall	0.9544	0.9500
F1 score	0.9529	0.9463

Table 1 summarizes the performance of ResNet50 and MobileNetV2 in classifying brain MRI images into healthy and tumor categories. ResNet50 achieved slightly higher results across all evaluation metrics, with an accuracy of 95.47%, a precision of 95.16%, a recall of 95.44%, and an F1 score of 95.29%. MobileNetV2 also demonstrated strong performance, with an accuracy of 94.80%, a precision of 94.33%, a recall of 95.00%, and an F1 score of 94.63%. While the differences between the two models are modest, ResNet50 shows a small advantage in overall predictive accuracy, whereas MobileNetV2 offers competitive results with the added benefit of being more computationally efficient. These results indicate that both models are effective for brain tumor detection, with the choice depending on the trade-off between performance and resource constraints.

Fig. 4 displays the confusion matrices for the applied models, providing a detailed overview of their classification results by showing both correct and incorrect predictions. This representation facilitates a deeper understanding of each model's performance and highlights specific areas where misclassifications occurred.

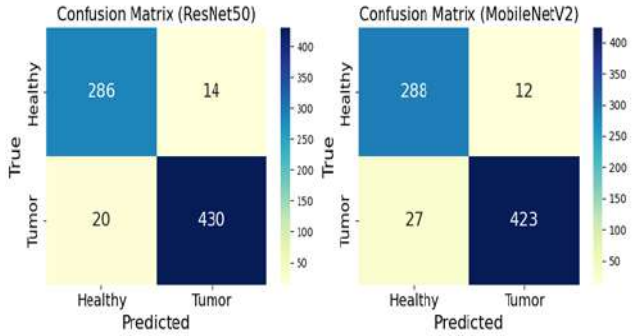

Fig. 4: Confusion matrices

The confusion matrices in Fig. 4 provide a detailed insight into the classification performance of ResNet50 and MobileNetV2 on the brain MRI dataset. For ResNet50, the model correctly identified 286 healthy cases and 430 tumor cases, while misclassifying 14 healthy scans as tumors and 20 tumor scans as healthy. In comparison, MobileNetV2 correctly classified 288 healthy and 423 tumor images, with 12 healthy scans incorrectly predicted as tumors and 27 tumor scans misclassified as healthy.

Overall, both models demonstrate strong performance, with high numbers of correct predictions in both classes. ResNet50 shows a slightly better balance in minimizing false negatives (20 tumor cases misclassified) compared to MobileNetV2 (27 tumor cases misclassified), which is important in medical diagnosis to reduce missed detections. On the other hand, MobileNetV2 slightly outperforms in correctly identifying healthy cases, with fewer false positives than ResNet50. These results suggest that while both models are effective, ResNet50 may be preferable when prioritizing tumor detection accuracy, whereas MobileNetV2 offers a slightly better performance in avoiding mislabeling healthy scans.

Figure 5 shows Grad-CAM visualizations of brain MRI predictions for ResNet50 and MobileNetV2. ResNet50 focuses precisely on relevant anatomical regions, especially around tumors, reflecting its higher accuracy and F1 score. MobileNetV2 also highlights important regions but with slightly more diffuse attention in some cases, aligning with its slightly lower recall. Both models rely on meaningful anatomical cues, with ResNet50 providing more precise localization, while MobileNetV2 offers a good balance of interpretability and efficiency.

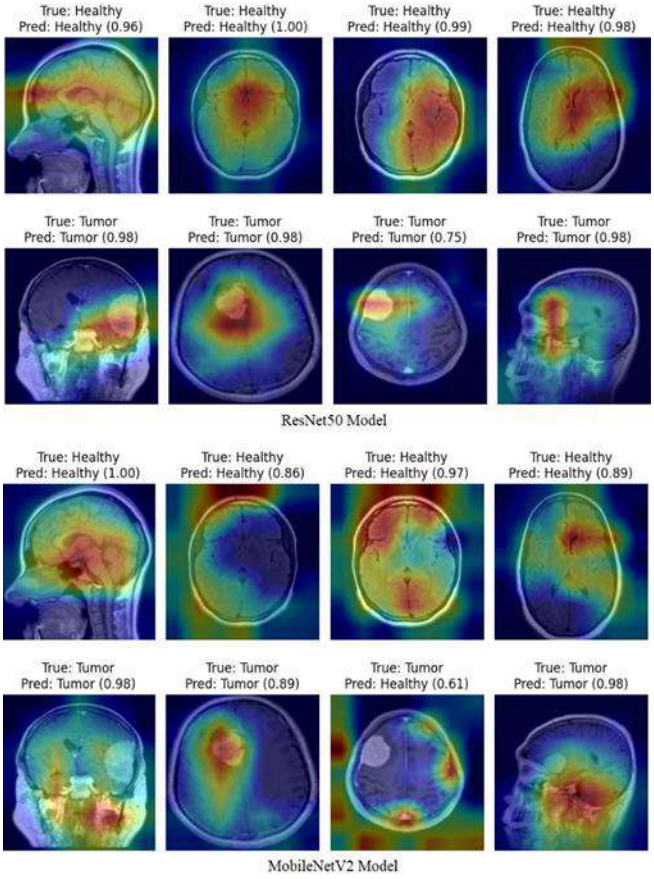

Fig. 5: Grad-CAM results.

The ResNet50 visualizations suggest that the model focuses strongly on tumor regions when making predictions. For healthy brains, the attention maps are spread fairly uniformly but show more emphasis around central brain regions, which may correspond to normal anatomical structures that the model relies on to identify (healthy) tissue. For tumor cases, the highlighted areas closely match the tumor locations in the MRI scans, indicating that the model is correctly identifying and localizing abnormal regions. The predicted confidence is consistently high (0.75–0.98), reflecting strong certainty in both classes. Even in the case with a slightly lower tumor prediction (0.75), the attention still emphasizes the tumor area, showing that ResNet50 can localize subtle anomalies effectively.

MobileNetV2 also shows correct predictions most of the time, but with slightly lower confidence in some cases. For healthy brains, the attention maps are somewhat more diffuse and occasionally highlight areas that do not clearly correspond to anatomical regions, suggesting less precise localization than ResNet50. For tumor cases, the model identifies tumors but sometimes misplaces or underrepresents the key tumor region, as in the example where the model predicts (healthy) with 0.61 confidence. This shows MobileNetV2 is less consistent in both localization and classification for tumors compared to ResNet50. Nevertheless, in most cases, the model still focuses on relevant areas.

ResNet50 appears more robust and precise than MobileNetV2 in both classifying and localizing tumors. Its attention maps are concentrated over relevant regions, even for subtle tumors, and its predictions are consistently confident. MobileNetV2, being a lighter model, trades off some accuracy and attention precision; its attention maps are often broader or misplaced, leading to occasional misclassification. For healthy cases, both models perform well, but ResNet50 demonstrates a more anatomically grounded attention, suggesting better feature extraction capabilities. For accurate tumor detection and precise localization in brain MRI scans, ResNet50 clearly outperforms MobileNetV2. MobileNetV2 could be preferred when computational efficiency is critical, but it may compromise reliability in borderline or challenging cases.

5. Conclusion

The study utilized ResNet50 and MobileNetV2 deep learning architectures to classify brain MRI scans into healthy and tumor categories. Both models achieved strong results across all evaluation metrics, with ResNet50 demonstrating slightly superior performance compared to MobileNetV2. The approach proved effective in identifying brain tumors, highlighting its potential as a diagnostic tool. Explainable AI techniques were employed to generate heatmaps, which consistently highlighted tumor-specific regions in abnormal scans. This not only improved the reliability of the models but also provided valuable visual insights for clinical interpretation. By offering clear explanations of the model's predictions, these methods enhance confidence in automated diagnostic outputs. Overall, the findings indicate that these deep learning models can play a significant role in supporting radiologists with early tumor detection, potentially reducing diagnostic times and improving patient care. Incorporating explainable AI into the workflow represents an important advancement toward making automated systems more transparent, trustworthy, and suitable for practical clinical application.

Despite the strong performance of both models, several limitations remain. The dataset imbalance continues to pose a challenge; even with class weighting applied, predictions for underrepresented tumor types may still be affected. Moreover, while Grad-CAM offers valuable interpretability, it cannot fully capture the complex interactions between features and may not account for all factors influencing the models' decisions. Lastly, although some previous studies have reported higher performance using alternative

architectures, the models presented here provide a well-rounded trade-off between accuracy, interpretability, and computational efficiency.

Future work could focus on addressing dataset limitations by incorporating larger and more balanced MRI datasets, including diverse tumor types and grades. Additionally, exploring advanced interpretability techniques beyond Grad-CAM to provide deeper insights into model decision-making. Finally, investigating alternative architectures could further improve tumor detection accuracy while maintaining computational efficiency.

6. Ethics Statement

The dataset utilized in this study was sourced from a publicly available repository hosted on Kaggle (https://www.kaggle.com/datasets/murtozalikhon/brain-tumor-multimodal-image-ct-and-mri). The images were fully anonymized and did not contain any identifiable patient information. Since the study relied exclusively on secondary analysis of publicly accessible data and involved no direct interaction with human subjects, ethical approval and informed consent were not required.

7. References

[1] R. Zhang, D. M. Wang, Y. L. Liu, M. L. Tian, L. Zhu, J. Chen, and J. Zhang, "Symptom management in adult brain tumours: A literature review," Nursing Open, vol. 10, no. 8, pp. 4892–4906, 2023.

[2] K. KK., M. S. Rajan, K. Hegde, S. Koshy, and A. Shenoy, "A comprehensive review on brain tumor," Int. J. Pharmaceutical, Chemical & Biological Sciences, vol. 3, no. 4, 2013.

[3] A. S. Peddinti, S. Maloji, and K. Manepalli, "Evolution in diagnosis and detection of brain tumor—Review," in Proc. J. Phys.: Conf. Series, vol. 2115, no. 1, p. 012039, Nov. 2021.

[4] A. A. Akinyelu, F. Zaccagna, J. T. Grist, M. Castelli, and L. Rundo, "Brain tumor diagnosis using machine learning, convolutional neural networks, capsule neural networks and vision transformers, applied to MRI: A survey," J. Imaging, vol. 8, no. 8, p. 205, 2022.

[5] S. Ali, J. Li, Y. Pei, R. Khurram, K. U. Rehman, and T. Mahmood, "A comprehensive survey on brain tumor diagnosis using deep learning and emerging hybrid techniques with multi-modal MR image," Arch. Comput. Methods Eng., vol. 29, no. 7, pp. 4871–4896, 2022.

[6] M. Arabahmadi, R. Farahbakhsh, and J. Rezazadeh, "Deep learning for smart healthcare—A survey on brain tumor detection from medical imaging," Sensors, vol. 22, no. 5, p. 1960, 2022.

[7] B. H. Van der Velden, H. J. Kuijf, K. G. Gilhuijs, and M. A. Viergever, "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis," Med. Image Anal., vol. 79, p. 102470, 2022.

[8] A. B. Arrieta et al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Information Fusion, vol. 58, pp. 82–115, 2020.

[9] M. S. I. Khan et al., "Accurate brain tumor detection using deep convolutional neural network," Comput. Struct. Biotechnol. J., vol. 20, pp. 4733–4745, 2022.

[10] M. M. Mijwil, R. Doshi, K. K. Hiran, O. J. Unogwu, and I. Bala, "MobileNetV1-based deep learning model for accurate brain tumor classification," Mesopotamian J. Comput. Sci., pp. 29–38, 2023.

[11] B. Badjie and E. D. Ülker, "A deep transfer learning based architecture for brain tumor classification using MR images," Information Technology and Control, vol. 51, no. 2, pp. 332–344, 2022.

[12] A. Younis, L. Qiang, C. O. Nyatega, M. J. Adamu, and H. B. Kawuwa, "Brain tumor analysis using deep learning and VGG-16 ensembling learning approaches," Applied Sciences, vol. 12, no. 14, p. 7282, 2022.

[13] H. R. Almadhoun and S. S. Abu-Naser, "Detection of brain tumor using deep learning," 2022.

[14] V. Anand, S. Gupta, D. Gupta, Y. Gulzar, Q. Xin, and S. Juneja, "Weighted average ensemble deep learning model for stratification of brain tumor in MRI images," Diagnostics, vol. 13, no. 7, p. 1320, 2023.

[15] A. Gurunathan and B. Krishnan, "Detection and diagnosis of brain tumors using deep learning convolutional neural networks," Int. J. Imaging Syst. Technol., vol. 31, no. 3, pp. 1174–1184, 2021.

[16] M. Murtoza, "Brain tumor multimodal image (CT and MRI) dataset," Kaggle. Available: https://www.kaggle.com/datasets/murtozalikhon/brain-tumor-multimodal-image-ct-and-mri

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778.

[18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 4510–4520.

[19] 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 Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017, pp. 618–626.