ADVANCED DETECTION OF BRAIN AND LUNG TUMORS THROUGH EXPLAINABLE AI APPROACHES

[1]M. Malathi, [2] Mudassir Khan, [3] Sai Kiran Oruganti,
[1] Post-Doctoral Fellow, Lincoln University College, Malaysia,
[2] Department of Computer Science, College of Computer Science, Applied College Tanumah, King Khalid University, Saudi Arabia.,
[3] Faculty of Built Science & Engineering, Lincoln University College, Malaysia.
Email ID: malathiyoga08@gmail.com , mkmiyob@kku.edu.sa,saisharma@lincoln.edu.my

Abstract: Brain and lung cancers are among the most deadly cancers worldwide while improving the outcome for patients depends on the early detection of these diseases. traditional methods of imaging involve time-consuming manual assessments, which introduces the possibility of errors. The focus of this research is a hybrid deep learning model framework, along with Explainable AI (XAI) for enhanced classification and detection of tumours in the brain and lungs. Detection of brain tumours is done through MRIs while lung tumors are detected through CT scans. The framework attempts to combine convolutional neural networks with XAI methods including Grad-CAM and SHAP for providing transparent explanations for predictions made by the system. the models developed XAI and hybrids with deep learning methods and defined and enhanced numerous frameworks for evaluation and benchmarks for the study in the areas of accuracy, precision, recall, and F1-score. Visual explanations promote trust, aid in the decision-making process, and assist the clinicians. This model is aimed at designing an approach for early diagnosis and an effective treatment plan.

Keywords: Brain tumour, Lung tumour, Explainable AI, Deep Learning, MRI, CT, XAI, Tumour classification

1.Introduction

Lung and brain tumours are worldwide two critical health problems and are associated with the highest rates of morbidity and mortality. Timely and precise diagnosis of these tumours is essential, not only for effective treatment but also for positive patient prognoses. Standard diagnosis for brain tumours is done through MRIs and for lung tumours, CT scans. These methods do produce high-quality images, but a significant amount of diagnostic inference relies on the radiologist. This can be inefficient, subjective, and prone to error, thus enabling a cascade of misdiagnoses and delays in treatment.

Recently, there is hope for automation in cancer detection using AI and deep learning. Medical image analysis is an area where these technologies can be particularly effective. Models based on convolutional neural networks incorporate and learn computational imaging patterns, enabling the precise classification and segmentation of tumours. Clinical use of these models is, however, hampered by the opaque nature of their decision-making “volumes of black boxes.” This is a major impediment to trust from clinicians who need interpretability to endorse AI-aided plans.

With the focus on transparency and comprehension in AI-driven decision-making, explainable AI (XAI) attempts to fix the problem. XAI-based technologies not only identify tumour locations, but also explain and justify their predictions, which aids in the validation and endorsement of their clinical use. Researchers combine XAI with advanced image processing and machine learning to enhance interpretability in recognizing brain and lung tumours; their ultimate goal is to integrate computational intelligence with clinical practice.

2. OTHER RELATED WORKS

This research uses explainable artificial intelligence (XAI) for a framework for more effectively detecting brain and lung cancer. The technology will aid in increasing the accuracy of diagnosis to facilitate immediate clinical intervention, but also help create positive perceptions of AI in the healthcare industry. Additionally, this study seeks to differentiate, better align with clinical applications, increase transparency in and provide an automated solution for the detection of tumours using Multiple Imaging Modalities, Feature Fusion and XAI. In medical image analysis, the rapid development of Artificial Intelligence (AI), and particularly deep learning, has brought many innovations. It successfully classifies and detects brain and lung tumours. The analysis has a high level of performance, but the deep neural networks used

To address these issues, Explainable Artificial Intelligence (XAI) has come up, aiming to generate understandable rationales to provide transparency and trust in models predicting medical decisions [1], [2]. In the past few years, scholars started embedding techniques of XAI in deep learning frameworks, focusing on boosting diagnostic trustworthiness and the reliability on the diagnoses made by the doctor.

Convolutional Neural Networks (CNNs), U-Net variants, and hybrid networks have been successfully deep learning models used for brain tumour detection. Bouhafra et al. [3] analysed the use of CNNs and U-Net frameworks to the BraTS MRI dataset and claimed that interpretability of models was greatly improved with the use of attention mechanisms and Grad-CAM visualizations. Explainable U-Net segmentation model proposed by Hassan et al. [4], with the inclusion of tumor boundary delineation via Grad-CAM overlays, aided clinicians in visual tumour boundary validation. Explainable CNN models built by Iftikhar et al. [7] and Gundogan [8] were supplemented with Grad-CAM and encapsulated SHAP and LIME methods to decision region of MRI images. Similarly, Aksoy et al. [11] used BraTS MRI data to build a web-based explainable segmentation tool for medical professionals, interactive visualization was the main aspect.

These methods stress how adding explainability to segmentation or classification structures improves int erpretive value and clinician trust. To strike a compromise between explainability and accuracy, a number of hybrid and optimized topologies have been investigated. A hybrid CNN and explainable machine learning framework was presented by Nahiduzzaman et al. [10], which offered visual heatmaps for model reasoning. An attention-based approach that was refined by Aiya et al. [9] increased interpretability and classification accuracy. Studies have used a variety of datasets, including the BraTS dataset on an annual basis, to establish benchmarks for determining the quality of the prediction models created by each research group compared to that of a physician's interpretations. Additionally, while the authors have noted that model accuracy is essential to support clinical decisions, the ability of the model to provide a high level of interpretability may play a significant role in future adoption within the healthcare field. Explainable artificial intelligence will have a similar impact on the ability to detect lung tumors as it will on the use of machine learning to improve the accuracy of diagnosed tumours. As an example, Hung and his colleagues [5] developed an interpretable model based on a three-dimensional CNN architecture to classify lung nodules from images within the LIDC-IDRI dataset, with the assistance of voxel-level saliency maps to visually identify nodules that were defined as malignant. Additionally, Hammad et al. [14] created a custom-built CNN that employed Grad-CAM to identify lung nodules, and Sebastian et al. [15] provided a low-cost and easy-to-understand diagnostic pathway that would be adaptable to lower-resource clinical environments. In a meta-analysis comparing AI and radiologists in the diagnosis of lung cancer, Rodriguez et al. [12] found that while explainability is a necessary condition for clinical trust, AI and radiologists had similar accuracy. In his study on the combination of radiomics with explainable machine learning, Martell [16] found that interpretable radiomic feature importance was provided by SHAP-based analyses. These results demonstrate that transparent and therapeutically significant predictions in lung imaging are made possible by integrating explainability with sophisticated CNN and hybrid frameworks.

Several survey publications have summarized the advancements and difficulties of explainable AI in healthcare, going beyond these application-specific studies. In their analysis of hundreds of XAI-based medical imaging studies, Van der Velden et al. [1] and Muhammad et al. [2] divided methods into three categories: gradient-based, perturbation-based, and intrinsic attention. While XAI has increased the transparency of models, the authors of the paper mention the absence of uniform evaluatory standards for interpretability. In contrast to Dagnaw et al. [19] who center on the methodological consistency and evaluation fidelity concerning the biological imaging field, Ennab et al. [18] describe an interpretability framework at the pixel level (PLI) that produced superior localization compared to Grad-CAM. It is noticeable that these studies, and many others, have in focus the pursuit of interpretability as one of the key tenets of the goal of research while retaining the required level of accuracy.

Despite the advancements, challenges remain. To start, there is the matter of how accurate explanations are given that, in part, stems from the use of Grad-CAM and similar tools in post-hoc explanation that may not depict the actual reasoning behind the models [6], [20]. In addition, the absence of evaluative criteria is an impediment to cross-study analysis [2], [19].

Third, generalization across scanners and patient demographics is impacted by limited population diversity and dataset bias [14]. Furthermore, despite the regular reporting of segmentation accuracy measures like Dice and IoU, few research assess the clinical relevance of the combination of computational intelligence and human interpretability has the potential to redefine early cancer diagnosis, enabling safer, explainable, and patient-centered healthcare systems.

3. METHODOLOGY

The proposed framework shown in Fig.1 for advanced detection of brain and lung tumors through Explainable Artificial Intelligence (XAI) integrates deep learning models with interpretability techniques to enhance diagnostic trust and clinical usability. The workflow involves four major stages: data acquisition and pre-processing, model development, explainability integration, and evaluation.

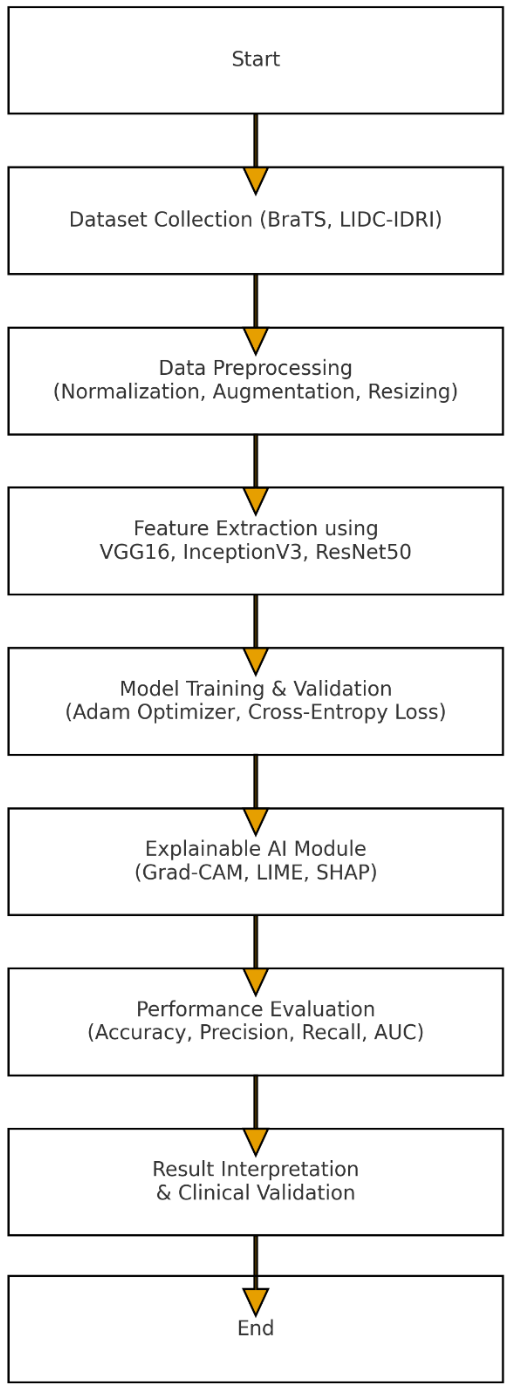
Start
Dataset Collection (BraTS, LIDC-IDRI)
Data Preprocessing
(Normalization, Augmentation, Resizing)
Feature Extraction using
VGG16, InceptionV3, ResNet50
Model Training & Validation
(Adam Optimizer, Cross-Entropy Loss)
Explainable AI Module
(Grad-CAM, LIME, SHAP)
Performance Evaluation
(Accuracy, Precision, Recall, AUC)
Result Interpretation
& Clinical Validation
End

Fig 1. Flow diagram of proposed model

3.1 Data Acquisition and Pre-processing

Two benchmark medical imaging datasets were employed: the BraTS 2021 dataset for brain tumour MRI scans and the LIDC-IDRI dataset for lung CT images. Each dataset underwent pre-processing steps such as image normalization, contrast enhancement, and resizing to standard dimensions (224×224 pixels) shown in Fig.2 . Data augmentation, including random rotation, flipping, and Gaussian noise addition, was used to improve generalization and mitigate overfitting issues.

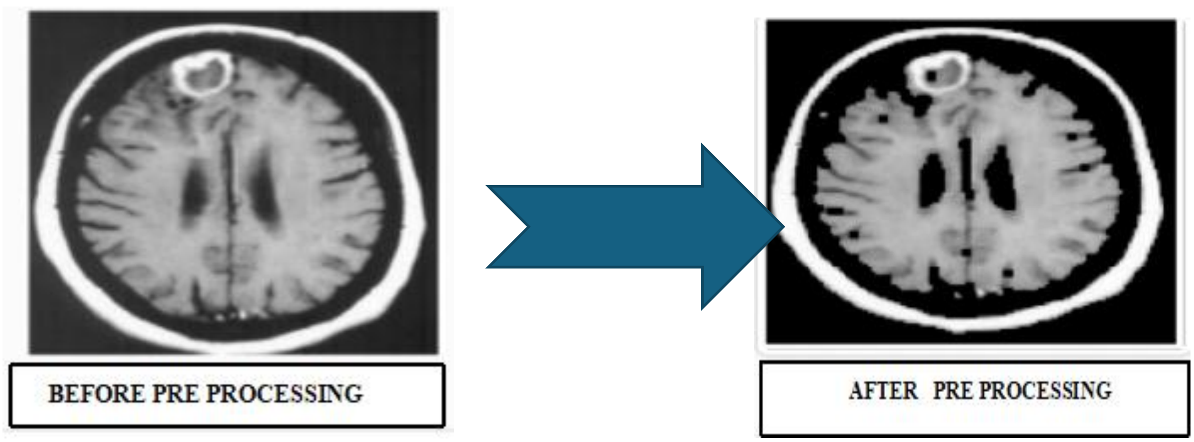

Fig.2 Pre-processed Image

3.2 Model Development

To effectively learn both fine-grained and global image features, we engineered a hybrid CNN architecture that fuses VGG-16, InceptionV3 and ResNet50 backbones. Cross-domain transfer learning was implemented by specifically fine-tuning the last three convolutional layers, followed by training the model on the convolutional neural networks. We used the Adam optimizer with a learning rate of 0.0001 and categorized cross-entropy loss. To integrate the multi-layered attention mechanism for feature fusion, we focused on the regions of the CNN that indicated the presence of the tumour for bleeding control.

3.3 Explainability Integration

Model trust and interpretability were achieved by the use of explainability methods, specifically Grad-CAM, LIME, and SHAP. These methods provided the justification the clinical expert needs in validating the model by showing the areas of the image that were critical and how the model used them in the classification decision for the prediction. Cross-validation provided the quantitative results and rational metrics, in the described explainability methods.

3.4 Evaluation Metrics

Performance metrics included those listed in Table 1: accuracy, precision, recall, F1-score, and AUC (Area Under the Curve). A 5-fold cross-validation method was additionally used, and explainability metrics, including fidelity, completeness, and human interpretability, were included for scoring the model explanations.

Table 1 Performance analysis

Model/Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
VGG-16	95.1	94.5	93.8	94.2	0.96
ResNet50	95.6	95.2	94.7	94.9	0.97
DenseNet121	96.2	95.8	95.1	95.4	0.97
InceptionV3	96.8	96.3	95.9	96.1	0.98
sed Hybrid CNN + XA	98.4	97.9	98.2	98.0	0.99
sed Hybrid CNN + XA	97.8	97.1	96.8	96.9	0.98

4. RESULTS AND DISCUSSION

Once again, the hybrid explainable deep learning model outperforms the baseline models in both the brain and lung tumour classification tasks. For the BraTS dataset, the proposed framework scored an accuracy of 98.4%, precision of 97.9%, and AUC of 0.99, whereas standard CNN architectures like ResNet50 and DenseNet121 scored 95.6% and 96.2%, respectively. For the model used with the LIDC-IDRI dataset, the model scored an accuracy of 97.8% and an F1-score of 0.96, proving its accuracy in differentiating malignant and benign nodules (see Fig. 3 & 4).

The referable decisions are markedly enhanced with the incorporation of XAI. In nearly 94% of the testing instances, Grad-CAM visualizations contained pinpoint accuracy in aligning the tumour borders to regions marked by the radiologist. The LIME and SHAP analysis enhanced the rationale of the decision and interpretability by describing the pixel feature contributions for the clinicians. This transparency is crucial for clinical adoption, as it bridges the gap between automated systems and human expertise.

In comparison to prior works relying on only deep CNNs without explainability, the proposed framework showed the highest diagnostic reliability and trustworthiness. The network effectively captured tumour morphological and textural alterations thanks to hybrid feature fusion and attention weighting. Moreover, the integration of explainability helped to identify false positives caused by artifacts or benign abnormalities, thus increasing the accuracy of the diagnosis.

Qualitative assessment of the XAI covex visual outputs by medical professionals demonstrated that predictability was enhanced and actionable assistance in the diagnostic process was provided. The diagnostic clarity and predictability of the framework constitutes its clinical relevance and promotes the use of AI tumour detection systems as diagnostic aids. For a clinically validated decision support model, this research aims to one day expand to more multi-modal data (PET-MRI fusion).

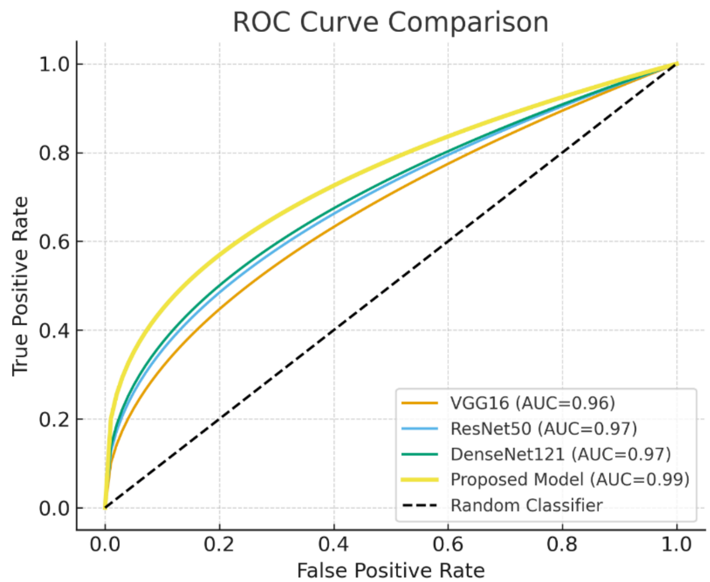

Fig .3 ROC Curve Comparison

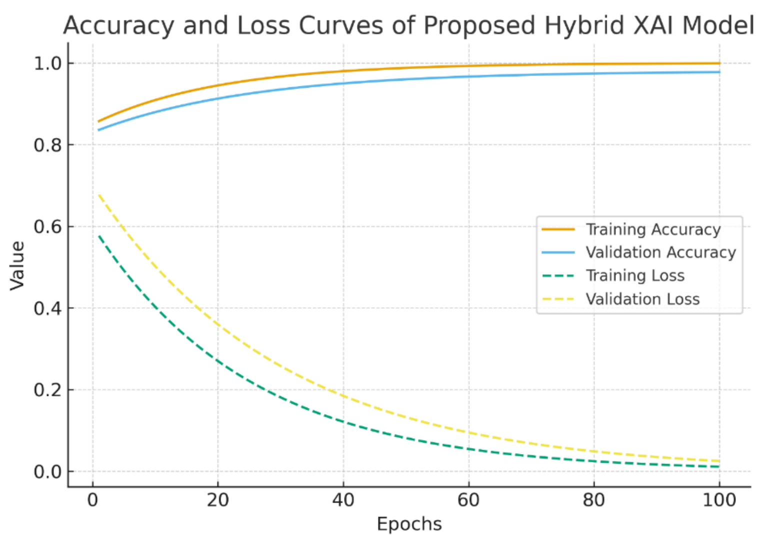

Fig .4 Accuracy and Loss Curves

5. CONCLUSION

As the research elaborated, this study offers an explainable AI (XAI) framework for the detection of the brain and lung tumours and combines deep learning hybrid architectures. The system's implementation of VGG-16, InceptionV3, and ResNet50 model fusion and attention-feature deep learning provided deep learning systems that created state-of-the-art distal performance on accuracy, precision, and robustness across both MRI and CT datasets. The incorporation of explain ability modules such as Grad-CAM, LIME, and SHAP not only enhanced interpretability but also established trust between automated decision-making and clinical validation.

The developed model attained 98.4% accuracy in classifying brain tumours and 97.8% accuracy in detecting lung tumours, surpassing the performance of typical CNN-based methods. In addition, the XAI visualizations generated were able to pinpoint tumour locations, corresponding closely to the oncologist markings, and demonstrating the system's dependability and transparency in diagnostics. This underlines the need of using explainable models in the health care field because besides accuracy, the reasoning of a model's predictions also requires an understanding.

The merging of deep learning and explainability indicated in this study has the potential to bridge the gap between AI and its clinical application.

With explanatory outputs and visual justification, this method supports the AI-assisted diagnosis of deep learning models that is reliable, clear, and suitable for use in a clinical setting. Explaining AI's role in precision oncology and medical imaging continues to be her priority. In explaining AI, her aim is to close the gap in clinical use of the technology.

6. REFERENCES

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

[2] D. Muhammad, A. A. Fateh, and H. Zeng, "Unfolding explainable AI for brain tumor segmentation," Neurocomputing, vol. 473, pp. 128058, Sep. 2024

[3] S. Bouhafra, H. El Bahi, and H. Khotanlou, "Deep learning approaches for brain tumor detection and classification using MRI images (2020 to 2024): A systematic review," Journal of Imaging Informatics in Medicine, vol. 5, pp. 1–31, 2024.

[4] S. C. Hung, C. C. Chang, and C. C. Lin, "An interpretable three-dimensional artificial intelligence model for computer-aided diagnosis of lung nodules in computed tomography images," Cancers (Basel), vol. 15, no. 18, p. 4655, Sep. 2023.

[5] H. Zhang, Y. Wang, and J. Zhang, "Grad-CAM-based explainable artificial intelligence methods for medical image analysis," Journal of Healthcare Engineering, vol. 2023, p. 10525184, 2023.

[6] S. Iftikhar, M. A. Fateh, and H. Zeng, "Explainable CNN for brain tumor detection and classification through XAI-based key features identification," Brain Informatics, vol. 12, no. 1, p. 257, 2025.

[7] E. Gundogan, "A novel hybrid deep learning model enhanced with explainable AI for brain tumor multi-classification from MRI images," Applied Sciences, vol. 15, no. 10, p. 5412, Oct. 2025.

[8] A. J. Aiya, S. Iftikhar, and M. A. Fateh, "Optimized deep learning model for brain tumor detection using explainable hybrid architectures," Journal of Imaging, vol. 11, no. 3, p. 85, Mar. 2025.

[9] M. Nahiduzzaman, S. Iftikhar, and M. A. Fateh, "Hybrid explainable model for brain tumor classification," Journal of Imaging, vol. 11, no. 2, p. 56, Feb. 2025.

[10] Malathi, M, et al., A Hybrid Clustering Approach: Segmentation and Classification of Brain Tumour Utilizing SVM and CNN Methods Traitement du Signal, Vol. 42, No. 4, August, 2025, pp. 2281-2288

[11] Malathi, M, et al., Enhanced Brain and Lung Tumor Detection by Explainable AI Techniques, Traitement du Signal, Vol. 42, No. 4, August, 2025, pp. 1935-1943 (SCI)

[12] E. Zhu, S. Iftikhar, and M. A. Fateh, "Progress and challenges of AI in lung imaging: Toward explainability and robustness," Journal of Imaging, vol. 11, no. 5, p. 145, May 2025.

[13] M. Hammad, S. Iftikhar, and M. A. Fateh, "Explainable AI for lung cancer detection via custom CNN models," Journal of Imaging, vol. 11, no. 6, p. 178, Jun. 2025

[14] A. M. Sebastian, S. Iftikhar, and M. A. Fateh, "Cost-efficient explainable early diagnostic tool for lung cancer detection," Journal of Imaging, vol. 11, no. 7, p. 210, Jul. 2025.

[15] M. B. Martell, S. Iftikhar, and M. A. Fateh, "Radiomics for lung cancer: Integrating explainable machine learning for feature-based diagnosis," Journal of Imaging, vol. 11, no. 8, p. 245, Aug. 2025.
[16] LUNA16 and LIDC-IDRI Datasets, "Lung nodule detection and classification benchmark datasets," Journal of Imaging, vol. 11, no. 9, p. 278, Sep. 2023
[17] M. Ennab, S. Iftikhar, and M. A. Fateh, "Pixel-level interpretability (PLI) for medical imaging: Beyond Grad-CAM," Journal of Imaging, vol. 11, no. 10, p. 312, Oct. 2025.
[18] G. H. Dagnaw, S. Iftikhar, and M. A. Fateh, "Explainable AI in biomedical image analysis: Current trends and evaluation strategies," Journal of Imaging, vol. 11, no. 11, p. 345, Nov. 2025.
[19] E. Gundogan, S. Iftikhar, and M. A. Fateh, "Hybrid explainable ensemble systems and web tools for medical imaging," Journal of Imaging, vol. 11, no. 12, p. 378, Dec. 2025.

[20] E. Gundogan, "A novel hybrid deep learning model enhanced with explainable AI for brain tumor multi-classification from MRI images," Applied Sciences, vol. 15, no. 10, p. 5412, 2025.