Explainable AI in MRI-Based Cancer Diagnosis: Improving Accuracy and Clinical Trust

Shrawan Kumar Yadav, Nims University Rajasthan, Jaipur, shrawan11kumar@rediffmail.com

Abstract

Magnetic Resonance Imaging (MRI) is a cornerstone in cancer diagnosis due to its superior soft tissue contrast and non-invasive nature. The increasing application of Artificial Intelligence (AI), particularly deep learning models, has enhanced the accuracy and efficiency of cancer detection in MRI scans. However, the opaque nature of these models poses challenges for clinical integration, as healthcare professionals often hesitate to rely on systems they cannot interpret or verify. This is where Explainable AI (XAI) becomes crucial—offering transparency in model decision-making by identifying the key features or regions influencing predictions. XAI techniques, such as saliency maps, SHAP (SHapley Additive exPlanations), and LIME (Local Interpretable Model-agnostic Explanations), help bridge the gap between AI-driven analytics and clinical judgment by making AI outputs interpretable and trustworthy.

This study highlights the importance of integrating XAI into MRI-based cancer diagnosis to improve diagnostic reliability, promote clinical adoption, and ensure ethical AI deployment. By enhancing transparency, XAI not only boosts clinician confidence but also enables collaborative decision-making, ultimately leading to better patient outcomes. Furthermore, explainability fosters regulatory acceptance and aligns with medical standards, making AI systems more viable for real-world application. Thus, incorporating XAI into diagnostic workflows is a critical step toward building intelligent, safe, and human-centered healthcare solutions.

Keywords: Explainable AI, MRI, Cancer Diagnosis, Deep Learning, Clinical Trust

Introduction

Magnetic Resonance Imaging (MRI) plays a pivotal role in cancer diagnosis due to its ability to capture detailed soft tissue contrast without ionizing radiation. However, interpreting MRI scans can be complex and prone to inter-observer variability among radiologists. To address this, Artificial Intelligence (AI), particularly deep learning, has emerged as a powerful tool to assist in identifying malignant patterns with higher speed and consistency. Despite its promise, the "black-box" nature of many AI models raises concerns among clinicians regarding reliability, accountability, and diagnostic transparency. This is where Explainable AI (XAI) becomes critical—by offering insights into how AI systems make decisions, XAI bridges the gap between complex algorithms and clinical applicability. Through visualizations like heatmaps or attention maps and interpretable feature extraction, XAI can highlight specific regions or characteristics within an MRI that influence the model's diagnosis, allowing radiologists to better understand and trust the system's outputs.

Incorporating XAI into MRI-based cancer diagnostics not only enhances clinical trust but also improves diagnostic accuracy by enabling collaborative decision-making. By making AI-driven predictions more interpretable, clinicians can validate or question the model's conclusions, leading to a more robust diagnostic workflow. Moreover, XAI facilitates regulatory approval and integration into real-world clinical settings by ensuring that AI models meet transparency and accountability standards. This increased interpretability is particularly valuable in high-stakes decisions, such as distinguishing between benign and malignant tumors or planning treatment strategies. As research in XAI evolves, techniques such as saliency maps, SHAP (SHapley Additive exPlanations), and LIME (Local Interpretable Model-agnostic Explanations) are being actively explored and adapted for medical imaging. Ultimately, the fusion of explainability and AI not only empowers clinicians with deeper insights but also fosters patient confidence in machine-assisted diagnoses, paving the way for more reliable, ethical, and effective cancer care.

Need of the Study

The need for this study arises from the increasing reliance on Artificial Intelligence (AI) in medical imaging, particularly in MRI-based cancer diagnosis, where precision and trust are paramount. While AI models, especially deep learning networks, have demonstrated remarkable accuracy in detecting and classifying cancerous lesions, their opaque decision-making processes limit clinical adoption. Physicians are often hesitant to rely solely on AI outputs without understanding the rationale behind the predictions, which can affect diagnostic confidence and patient outcomes. Therefore, integrating Explainable AI (XAI) is essential to make these systems more transparent, interpretable, and aligned with clinical reasoning. XAI not only helps in validating the AI's predictions but also enhances collaborative diagnosis, where radiologists can corroborate AI findings with their expertise. This study is crucial to explore how XAI tools like saliency maps, SHAP, or LIME can be effectively applied to MRI scans, highlighting relevant image features that influence diagnostic decisions. Furthermore, as AI becomes increasingly embedded in healthcare, there is a pressing need to ensure ethical deployment, regulatory compliance, and user acceptance. Thus, this study addresses a significant gap in current medical AI research—developing systems that are not only accurate but also explainable and trustworthy for safe integration into routine cancer diagnostic workflows.

Conclusion

The integration of Explainable AI (XAI) into MRI-based cancer diagnosis marks a significant advancement in the field of medical imaging, bridging the gap between high-performing AI models and the clinical need for transparency and trust. While traditional AI systems offer impressive diagnostic capabilities, their black-box nature poses a challenge to widespread clinical adoption. This limitation is particularly critical in cancer diagnosis, where the consequences of misinterpretation can be severe. By providing interpretable outputs and visual explanations, XAI empowers clinicians to understand, verify, and trust AI-assisted decisions, ultimately improving diagnostic confidence and patient care outcomes.

XAI facilitates better collaboration between AI systems and medical professionals by offering insights into the decision-making process through tools such as saliency maps, attention mechanisms, SHAP, and LIME. These techniques enable healthcare providers to pinpoint the image regions or features influencing a model's prediction, making it easier to validate results and identify potential errors. Moreover, the use of XAI contributes to regulatory approval and ethical deployment by promoting accountability and reducing the risks associated with opaque algorithms. It also plays a vital role in patient communication, allowing clinicians to explain diagnoses and treatment decisions more clearly.

In conclusion, the adoption of Explainable AI in MRI-based cancer diagnosis not only enhances the accuracy of disease detection but also builds clinical trust and supports ethical AI deployment. As the field evolves, continued research into more robust, user-friendly XAI methods will be essential for integrating AI seamlessly into routine clinical practice. The future of cancer diagnostics lies not only in intelligent algorithms but also in transparent, interpretable systems that align with medical expertise, ensuring safer, more reliable, and human-centric healthcare delivery.

References

1. Hassan, M. R., Islam, M. F., Uddin, M. Z., Ghoshal, G., Hassan, M. M., Huda, S., & Fortino, G. (2022). Prostate cancer classification from ultrasound and MRI images using deep learning based Explainable Artificial Intelligence. *Future Generation Computer Systems*, *127*, 462-472.
2. Intelligence, D. A. (2025). From black box AI to XAI in neuro-oncology: a survey on MRI-based tumor detection. *Discover*, *5*, 30.
3. Qian, J., Li, H., Wang, J., & He, L. (2023). Recent advances in explainable artificial intelligence for magnetic resonance imaging. *Diagnostics*, *13*(9), 1571.
4. Deshpande, N. M., Gite, S., Pradhan, B., & Assiri, M. E. (2022). Explainable artificial intelligence–a new step towards the trust in medical diagnosis with AI frameworks: a review. *CMES-Computer Modeling in Engineering and Sciences*.
5. Mittal, P. (2025). From black box AI to XAI in neuro-oncology: a survey on MRI-based tumor detection. *Discover Artificial Intelligence*, *5*(1), 1-21.
6. Ariful Islam, M., Mridha, M. F., Safran, M., Alfarhood, S., & Mohsin Kabir, M. (2025). Revolutionizing Brain Tumor Detection Using Explainable AI in MRI Images. *NMR in Biomedicine*, *38*(3), e70001.