Article type: Perspective

Home Page URL: https://www.alliedacademies.org/neuroinformatics-and-neuroimaging/

Explainable AI models for clinical decision support in neuroimaging-based diagnostics.

Elias Papadopoulos*

Department of Neuroimaging, University of Athens, Greece.

***Correspondence to:** Elias Papadopoulos**,** Department of Neuroimaging, University of Athens, Greece, E-mail: elias.papadopoulos@braininformatics.gr

Received: 03-Jan-2025, Manuscript No. AANN-25-169301; **Editor assigned:** 04-Jan-2025, PreQC No. AANN-25-1693015(PQ); **Reviewed:** 18-Jan-2025, QC No AANN-25-1693015; **Revised:** 21-Jan-2025, Manuscript No. AANN-25-1693015(R); **Published:** 28-Jan-2025, DOI:10.35841/aann -10.2.195

Introduction

Explainable artificial intelligence (XAI) models are increasingly being recognized as essential tools for clinical decision support in neuroimaging-based diagnostics. While artificial intelligence (AI) has demonstrated remarkable capabilities in detecting patterns and predicting outcomes from complex neuroimaging data, the opaque nature of many deep learning algorithms has hindered their adoption in clinical practice. Clinicians require transparency and interpretability to trust and act upon AI-generated recommendations, especially when these decisions impact patient diagnosis and treatment. Explainable AI bridges this gap by offering models that not only make accurate predictions but also provide human-understandable rationales for their outputs. In neuroimaging, where MRI, CT, PET, and other modalities generate vast amounts of high-dimensional data, XAI enables the visualization and interpretation of the specific brain regions, features, or connectivity patterns influencing the model's decisions, thereby fostering clinical confidence and regulatory compliance [1].

The implementation of XAI in neuroimaging-based diagnostics typically involves incorporating interpretability techniques into established AI frameworks. For example, convolutional neural networks (CNNs), widely used for image classification and segmentation, can be enhanced with visualization methods such as saliency maps, Grad-CAM (Gradient-weighted Class Activation Mapping), or layer-wise relevance propagation (LRP). These methods highlight regions of interest in brain scans that contribute most strongly to the model's predictions, allowing radiologists to compare these AI-derived attention maps with their own expert assessments. In the context of neurodegenerative diseases like Alzheimer's or Parkinson's, XAI models can pinpoint subtle structural or metabolic changes that may not be obvious to the human eye, enabling earlier detection. Additionally, feature attribution methods for structured neuroimaging-derived biomarkers—such as volumetric measurements, diffusion tensor imaging metrics, or functional connectivity scores—help explain model reasoning in a more quantitative manner, making it easier to integrate into clinical reporting [2].

Beyond the technical visualization of decision-making processes, explainable AI also plays a role in improving diagnostic accuracy and reducing bias in neuroimaging models. Black-box AI systems are vulnerable to learning spurious correlations or overfitting to non-clinical artifacts, such as scanner-specific noise or demographic imbalances in the training dataset. XAI techniques allow researchers and clinicians to inspect whether the model is making decisions based on clinically meaningful features rather than irrelevant confounders. This interpretability supports iterative refinement of both the model and the data, improving robustness and

generalizability. In diseases with overlapping imaging signatures, such as differentiating between frontotemporal dementia and Alzheimer's disease, XAI can reveal the nuanced features that guide the model toward one diagnosis over another, offering valuable insights for both machine learning engineers and medical practitioners [3].

The integration of XAI into clinical decision support systems for neuroimaging requires a careful balance between model complexity and interpretability. While highly complex deep learning architectures often achieve state-of-the-art accuracy, they can be challenging to interpret without sophisticated post hoc explanation techniques. In contrast, simpler machine learning models, such as decision trees or generalized linear models, are inherently more interpretable but may lack the predictive power of deep networks. Hybrid approaches, combining the strengths of both, are gaining attention—for example, using deep learning for feature extraction followed by interpretable classifiers for decision-making. Furthermore, user-centered design principles are critical when deploying XAI tools in clinical settings, ensuring that the explanations are presented in formats that align with the workflows and cognitive needs of radiologists and neurologists. Visualization dashboards, interactive heatmaps, and textual justifications tailored to medical decision-making can significantly enhance usability and adoption [4].

Despite these advances, several challenges remain before XAI can be fully integrated into routine neuroimaging-based diagnostics. Standardizing explanation methods across different AI architectures is essential for consistency, as variations in interpretability outputs can lead to confusion or misinterpretation. The clinical validation of XAI tools requires large, diverse datasets and rigorous testing to ensure that explanations are reliable across patient populations and imaging modalities. There are also legal and ethical considerations: explanations must be accurate and actionable to avoid misleading clinicians, and patients should be informed about the role of AI in their care. Additionally, the computational demands of generating explanations for high-resolution neuroimaging data can be significant, necessitating efficient algorithms and optimized infrastructure. Overcoming these hurdles will require collaboration among AI researchers, clinicians, regulatory bodies, and industry partners to establish best practices and guidelines for explainable neuroimaging AI [5].

Conclusion

Explainable AI represents a crucial advancement for integrating artificial intelligence into neuroimaging-based clinical decision support systems. By making the decision-making processes of AI models transparent, interpretable, and clinically meaningful, XAI bridges the trust gap between machine predictions and medical expertise. These models not only improve diagnostic accuracy but also enhance clinician understanding of disease-specific imaging patterns, enabling earlier detection and more personalized treatment strategies. While challenges related to standardization, validation, and computational efficiency remain, ongoing research and interdisciplinary collaboration are poised to overcome these barriers. As explainable AI becomes more mature and accessible, it has the potential to transform neuroimaging diagnostics into a more transparent, trustworthy, and effective component of modern healthcare.

References

1. Di Costanzo A, Trojsi F, Tosetti M, et al. High-field proton MRS of human brain. Eur J Radiol. 2003;48(2):146-53.
2. Soher BJ, Dale BM, Merkle EM. A review of MR physics: 3T versus 1.5 T. Magn Reson Imaging Clin N Am. 2007;15(3):277-90.
3. Rovira A, Cordoba J, Sanpedro F, et al. Normalization of T2 signal abnormalities in hemispheric white matter with liver transplant. Neurology. 2002;59(3):335-41.
4. Rovira A, Grivé E, Pedraza S, et al. Magnetization transfer ratio values and proton MR spectroscopy of normal-appearing cerebral white matter in patients with liver cirrhosis. Am J Neuroradiol. 2001;22(6):1137-42.
5. Hajnal JV, Baudouin CJ, Oatridge A, et al. Design and implementation of magnetization transfer pulse sequences for clinical use. J Comput Assist Tomogr. 1992;16(1):7-18.

Citation: Papadopoulos E. Explainable AI models for clinical decision support in neuroimaging-based diagnostics. J NeuroInform Neuroimaging. 2025;10(2):195.