Title: Explainable AI for Medical Image Classification and Segmentation

Planning Abstract: The advancement of Artificial Intelligence (AI) in medical imaging has significantly improved the accuracy of disease diagnosis and treatment planning. However, deep learning models, particularly in medical image classification and segmentation, are often criticized for their lack of transparency and interpretability. The black-box nature of these models raises concerns about reliability, and the potential risks associated with incorrect predictions in high-stakes medical applications. This challenge makes it difficult for medical professionals to trust and effectively utilize deep learning models. The primary goal is to explore how Explainable AI (XAI) techniques can enhance trustworthiness and interpretability in deep learning-based medical imaging models. Understanding how these models conclude their predictions is crucial for gaining acceptance among healthcare practitioners and ensuring reliable healthcare applications. This research is arguably significant because of its potential to bridge the gap between AI advancements and healthcare usage. With increasing reliance on AI models in medical imaging, enhancing their interpretability can lead to better diagnostic support, and improved patient outcomes. Exploring XAI in the medical domain aligns with ethical AI principles by promoting transparency and accountability. This research will include the application of activation based visual methods such as Grad-CAM, Grad-CAM++, and Score-CAM to generate heatmaps that highlight critical regions contributing to classification and segmentation tasks that helps understand more about the visual outputs. These heatmaps will separate the specific portions from inputs to distinguish between the parts that's helping the most in prediction task. Also, this research will include feature attribution methods similar to Layer-wise Relevance Propagation (LRP), and Shapley values to get a deeper understanding of which features influence predictions. The effectiveness will be analyzed by reviewing case studies and benchmarking XAI approaches. The result analysis will be shown using the latest work that integrated XAI methods with medical image segmentation. The findings will be combined to highlight best practices, challenges, and future research directions in the field. This research will also provide insights into how XAI can improve medical image analysis while maintaining high classification and segmentation performance. This research will contribute to the broader discussion on the role of XAI in medical imaging and its potential to improve trust and usability in clinical AI applications.