Research Proposal

Deep learning has achieved remarkable success across diverse domains, from image recognition and natural language processing to healthcare and finance. However, the reliance on complex, black-box models raises concerns regarding robustness, interpretability, and trust. These limitations hinder the deployment of deep learning systems in critical applications where reliability and transparency are paramount. This research proposal aims to address these challenges by developing novel machine learning methodologies that enhance both the robustness and interpretability of deep learning models, focusing on medical image analysis for cancer detection.

Robustness: Deep learning models are susceptible to adversarial attacks and distribution shifts, leading to significant performance degradation. This vulnerability is particularly problematic in safety-critical applications like autonomous driving and medical diagnosis.

Interpretability: The lack of transparency in deep learning models makes it difficult to understand their decision-making process, hindering trust and accountability. This is especially crucial in domains where explainability is legally mandated or ethically required.

Motivation:

To develop robust deep learning models that are resilient to adversarial attacks and distribution shifts, ensuring reliable performance in real-world scenarios.

To enhance the interpretability of deep learning models, enabling users to understand and trust their predictions, facilitating informed decision-making.