Explainable AI (XAI)

Interpretability Needs

Clinical trust and adoption. Regulatory requirements and liability concerns

Saliency Maps

Grad-CAM, attention maps. Highlight important image regions

Counterfactual Explanations

What changes would flip prediction. Minimal perturbations for opposite class

Concept Attribution

High-level concept importance. Relate to clinical features (texture, shape, density)

Trust Building

Consistent explanations aligned with clinical knowledge. Detect spurious correlations