Gradient-Based Attribution

Input Image

Suspicious region

Sensitivity Analysis

Region A: High

Region B: Medium

Region C: Low

Neural Network

Convolutional

Feature Maps

Classification

∇ **Gradient Backprop**

Prediction

Malignant
0.87

Saliency Map

High gradient magnitude

Attribution Score

$A\_i = |\partial f/\partial x\_i|$
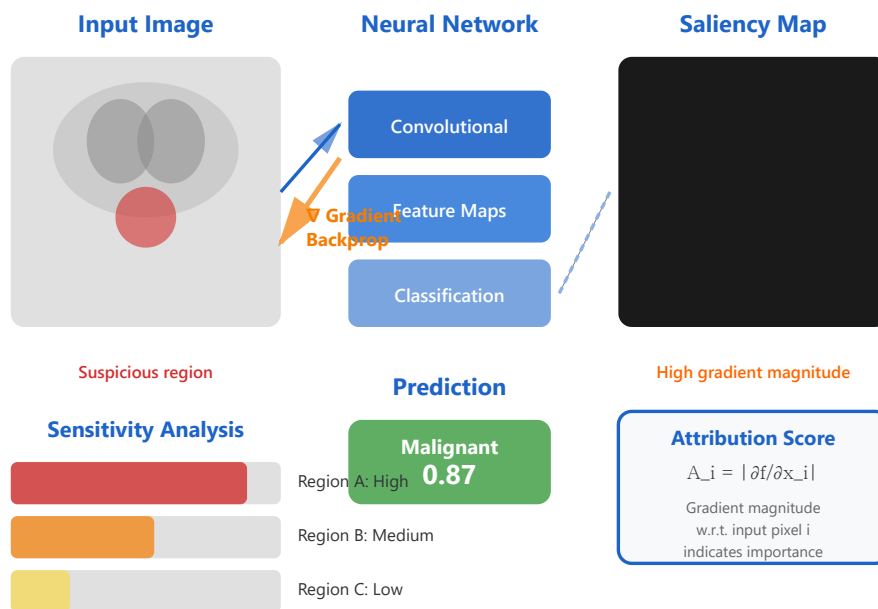
Gradient magnitude
w.r.t. input pixel i
indicates importance

📈 **Gradient Analysis**

Using gradients to measure input feature importance

🎯 **Sensitivity Mapping**

Showing which inputs affect the output most significantly

🗺️ **Saliency Maps**

Visualizing important pixels and regions in images

⚕️ **Clinical Interpretation**

Understanding what factors drive model decisions