



Evaluating the Fidelity of Explanations for Convolutional Neural Networks in Alzheimer's Disease Detection

Bjarne C. Hiller

2025-03-09

University of Rostock

Who are we?

University of Rostock: Institute for Visual and Analytic Computing

Bjarne C. Hiller

Martin Becker

Thomas Kirste

German Center for Neurodegenerative Diseases (DZNE)

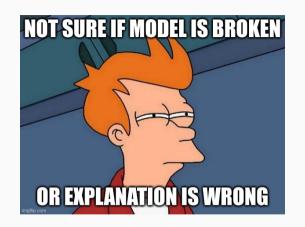
- Devesh Singh, MSc
- Dr. Martin Dyrba

Deep Learning for Medical Image processing

Data

Model

eXplainable AI and Feature Attribution Methods



Perturbation Tests: Insertion and Deletion

Attribution Maps

Relevance per ROI

results

Hiller (Uni Rostock) Fidelity of Explanations for AD Classification from MRI

Take-Away

- 1. Perturbation tests offer a model-agnostic fidelity metric.
- 2. The baseline should be chosen carefully w.r.t. to context.
- 3. Attribution Maps need interpretation to actually explain anything.