

# Guidelines and Evaluation of Clinical Explainable AI in Medical Image Analysis

Weina Jin<sup>1</sup>, Xiaoxiao Li<sup>2</sup>, Mostafa Fatehi<sup>3</sup>, Ghassan Hamarneh<sup>1</sup>

1. Medical Image Analysis Lab, School of Computing Science, Simon Fraser University  
 2. Department of Electrical and Computer Engineering, University of British Columbia  
 3. Division of Neurosurgery, University of British Columbia

Published in  
*Medical Image Analysis* 84 (2023) 102684  
 DOI: 10.1016/j.media.2022.102684

HAMARNEH  
 Medical Image Analysis  
 Research Group

SFU

## Research Problem

How can developers choose an AI explainability technique that is optimal for a target clinical task?

## Our Solution

## Clinical Explainable AI Guidelines

No technical knowledge is required to understand the explanation

Explanation is relevant to clinical decision-making

Explanation should truthfully reflect model decision process

Human judgment on explanation plausibility may reveal decision quality

Computational speed is within clinical users' tolerable waiting time

Explainable AI algorithms

Guideline 1  
Understandable

Guideline 2  
Clinical relevant

Guideline 3  
Truthful

Guideline 4  
Informative plausibility

Guideline 5  
Fast

Suitable for clinical use

Evaluation results on 16 heatmap algorithms

G1 Passed

G2 Partially passed

G3 Not passed

G4 Not passed

G5 Mostly passed

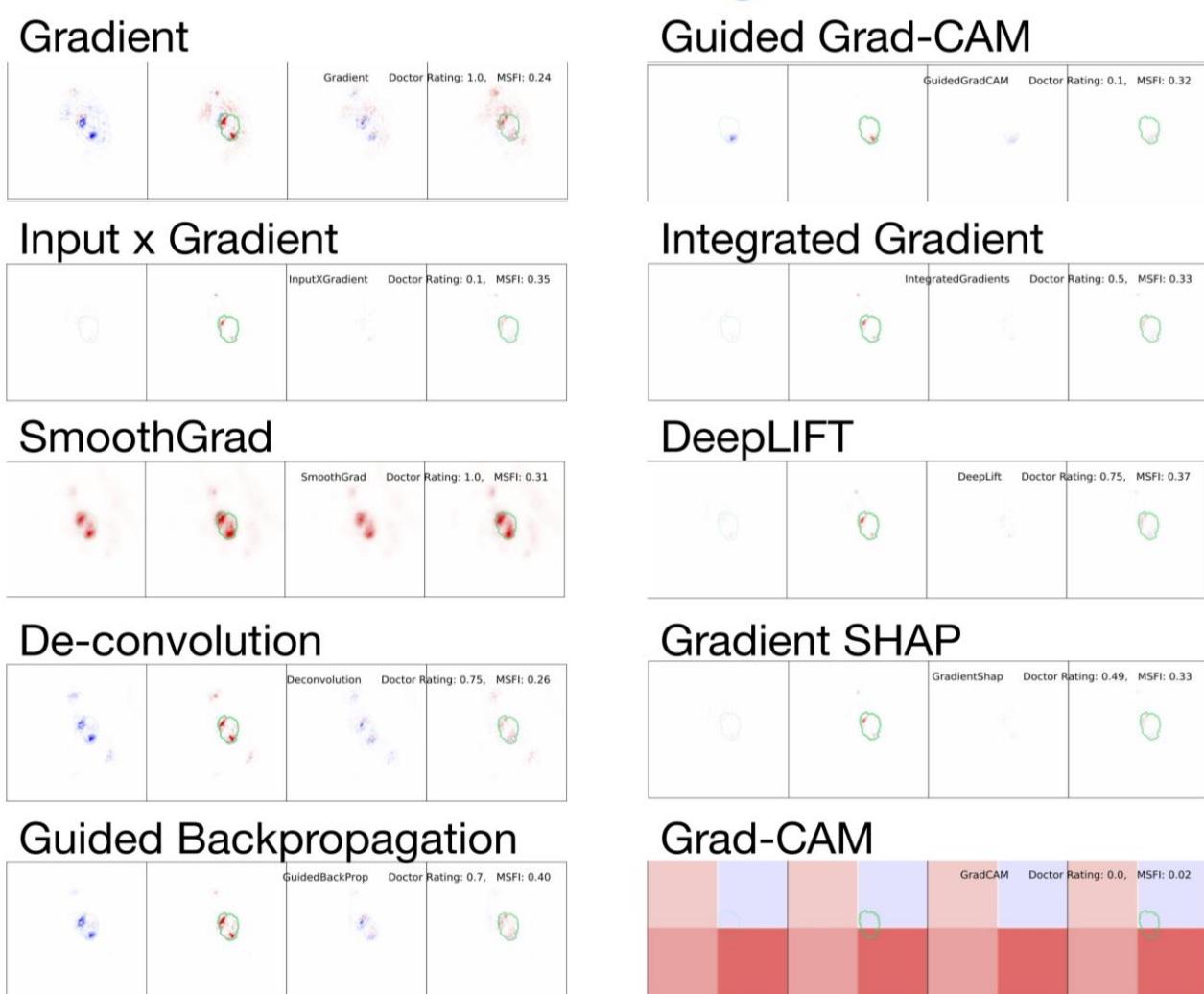
The evaluated heatmap methods did not meet G3 and G4, thus cannot be recommended for clinical use.

G1 Understandability & G2 Clinical relevance

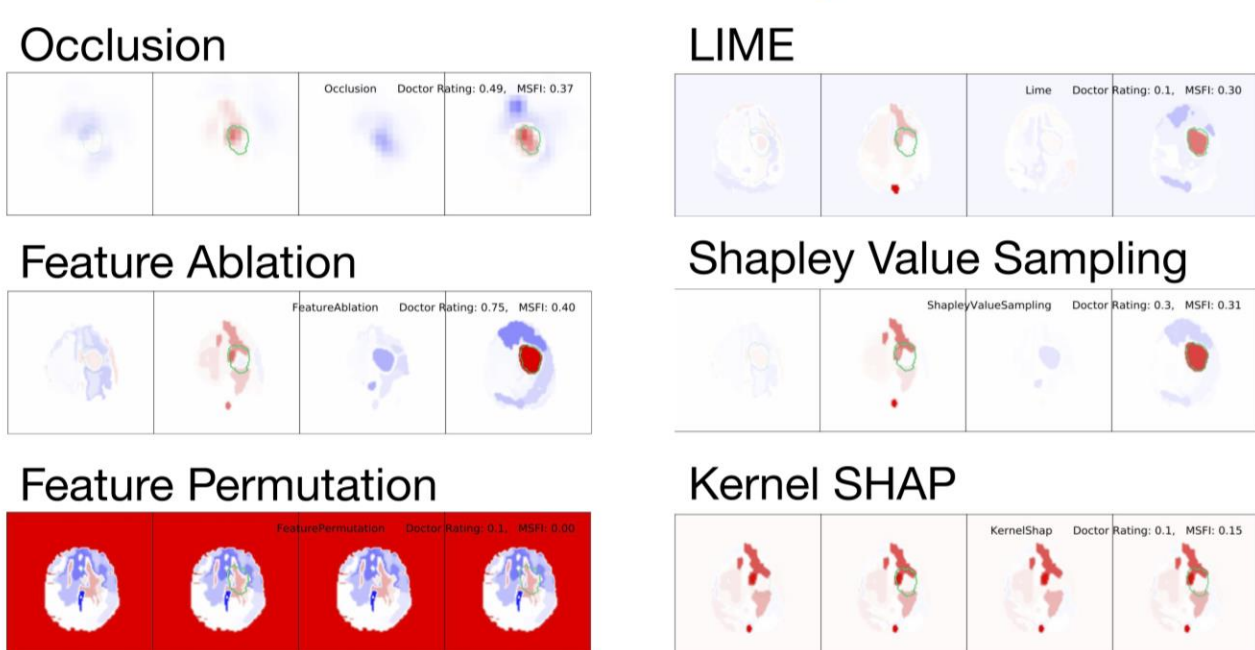
G3 Truthfulness

### Systematic evaluation on 16 heatmap algorithms

#### Gradient-based algorithms

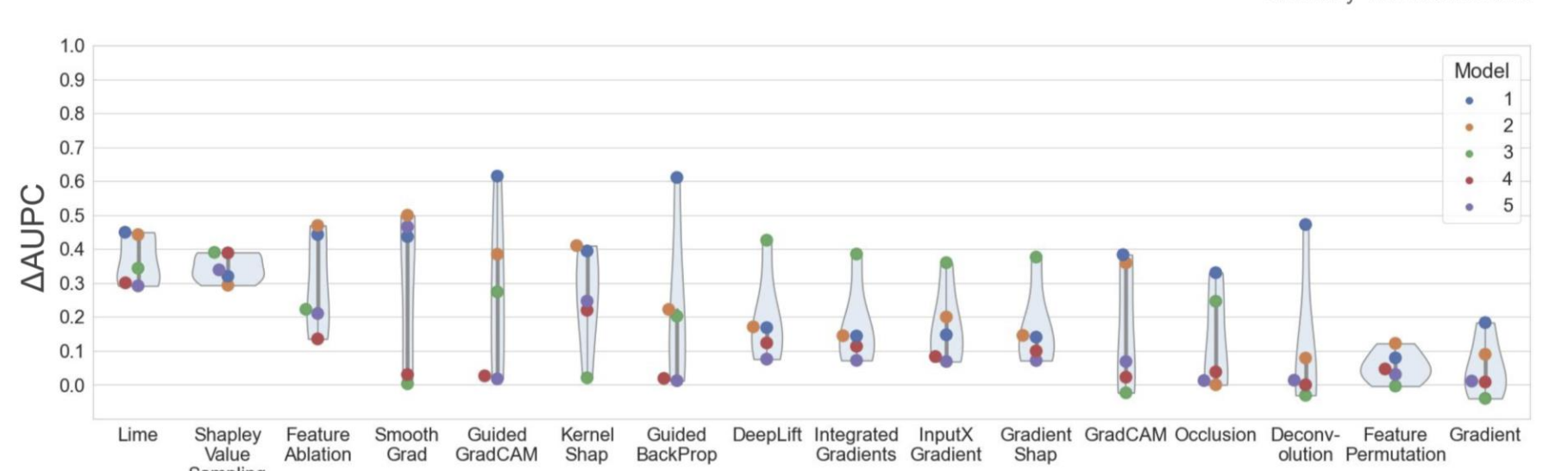
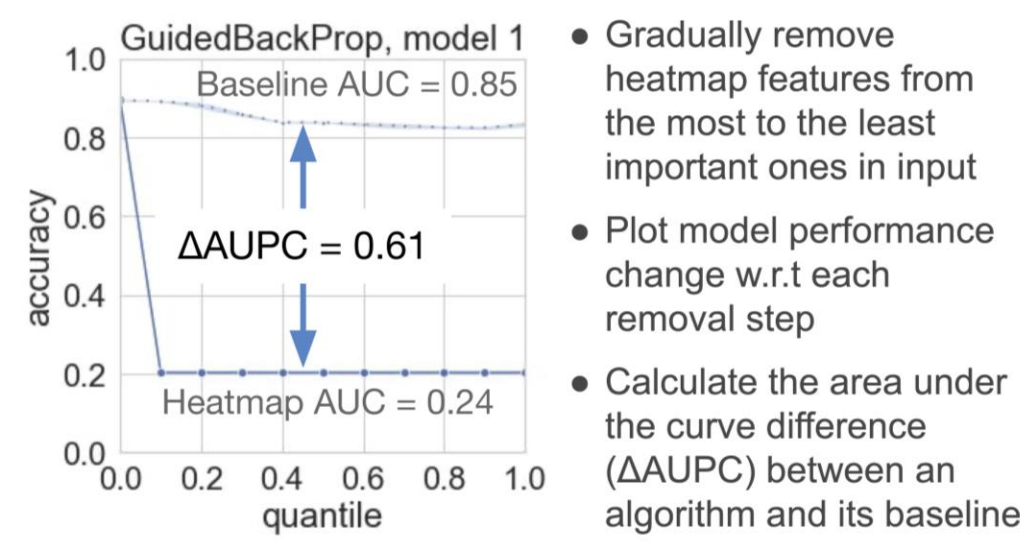


#### Perturbation-based algorithms



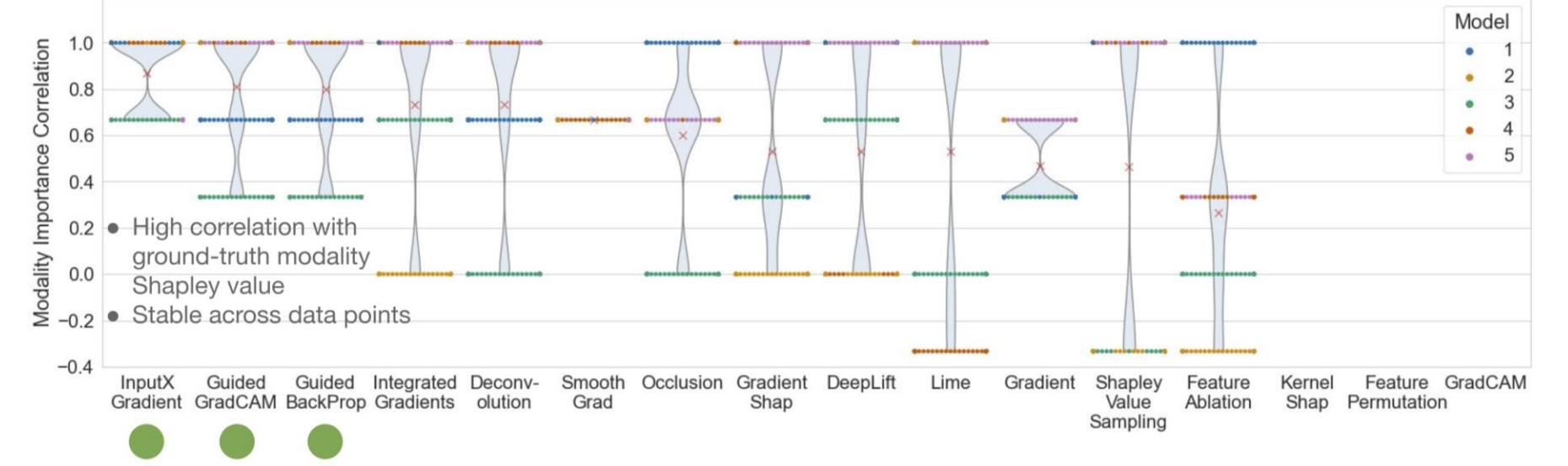
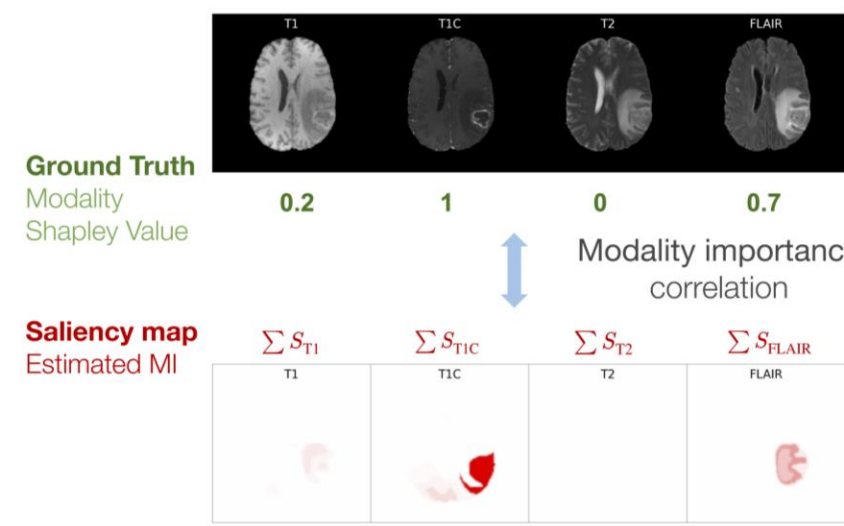
### 1. Gradual feature removal - feature level

All algorithms failed



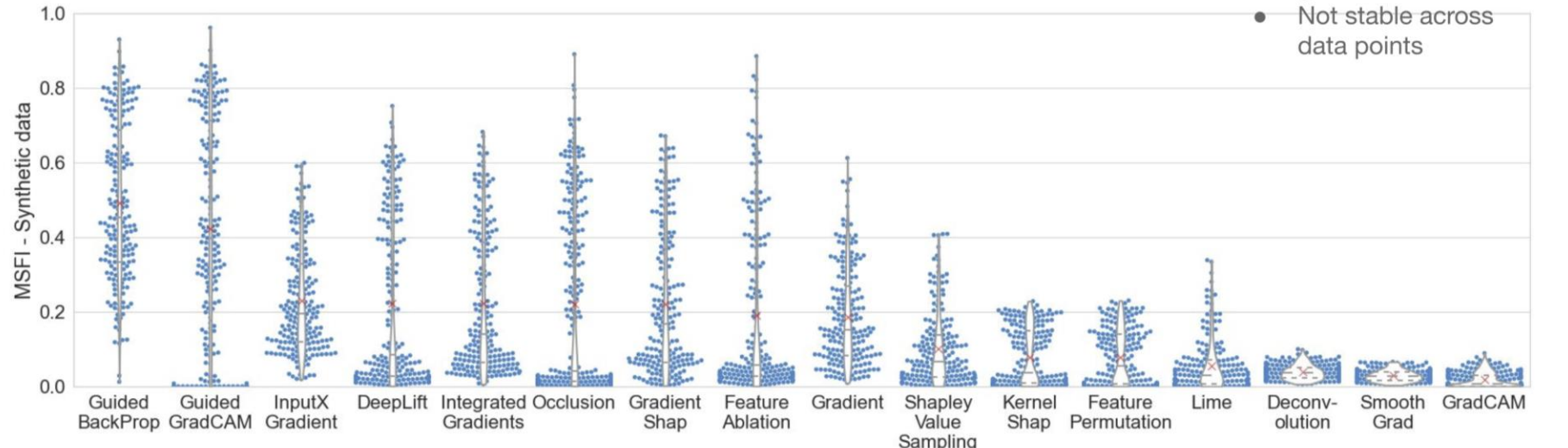
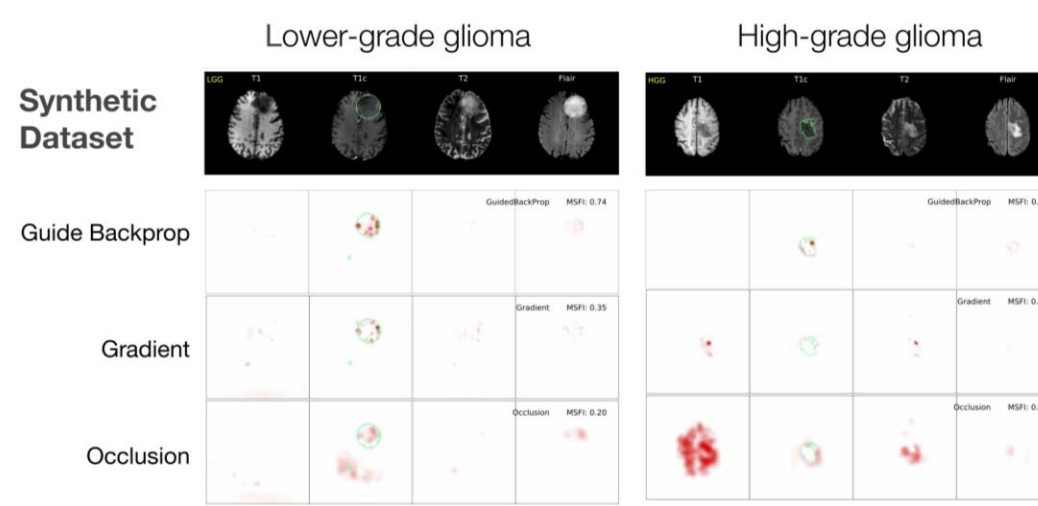
### 2. Modality importance correlation - modality level

Some algorithms passed



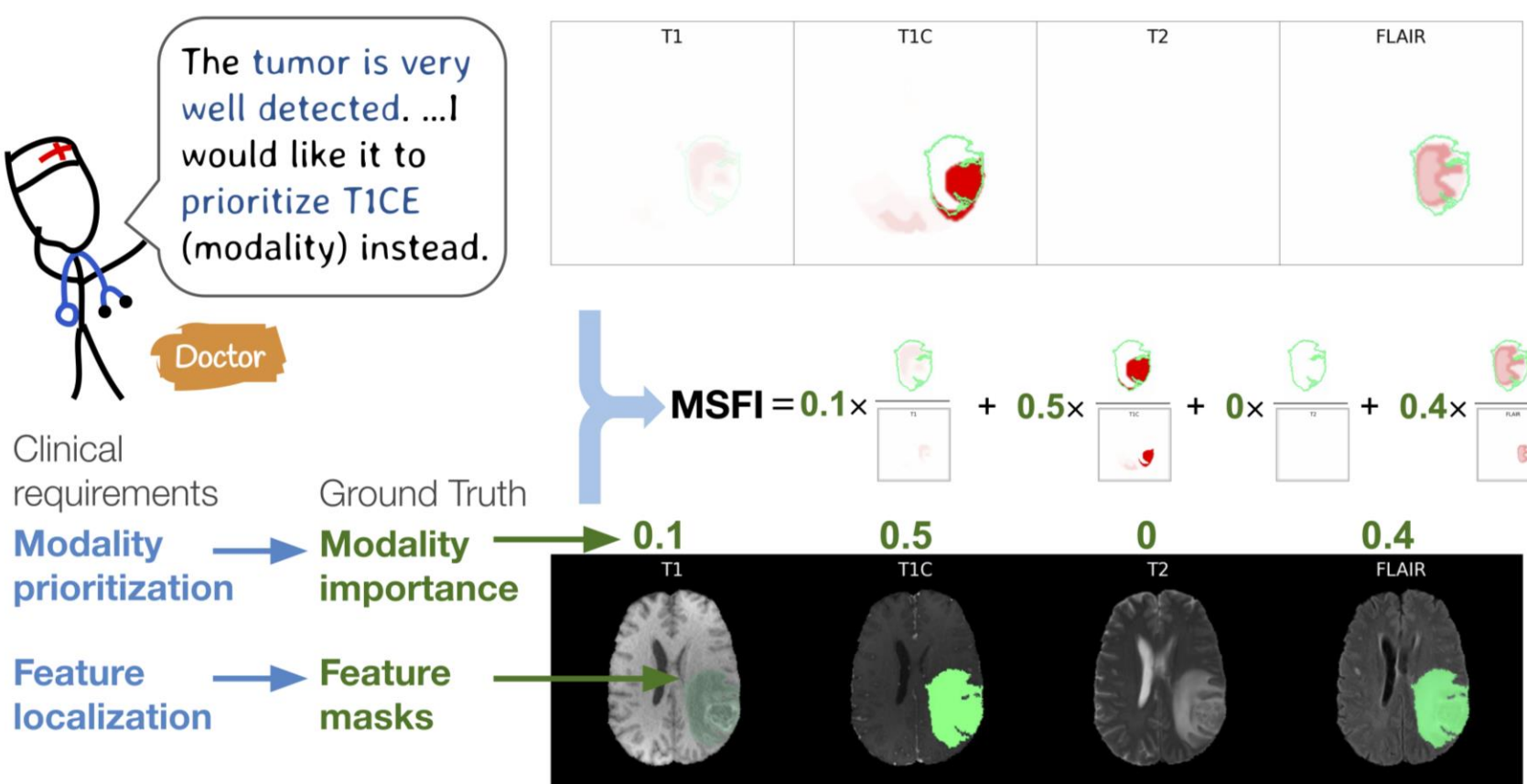
### 3. Synthetic data experiment - feature level

All algorithms failed

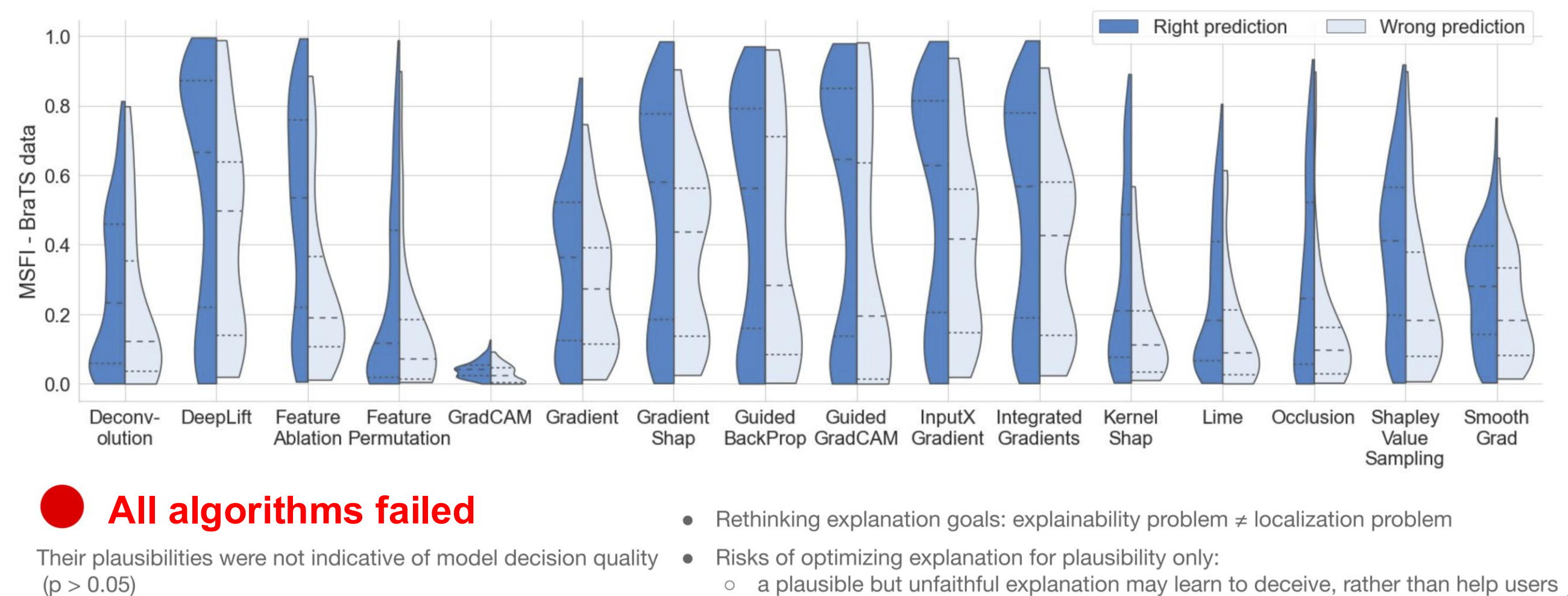


## G4 Informative plausibility

### Step 1. Plausibility quantification



### Step 2. Test for plausibility relation with prediction correctness



Project Webpage  
[weina.me/clinical\\_xai\\_guideline](http://weina.me/clinical_xai_guideline)



Project Code  
[github.com/weinajin/multimodal\\_explanation](https://github.com/weinajin/multimodal_explanation)