

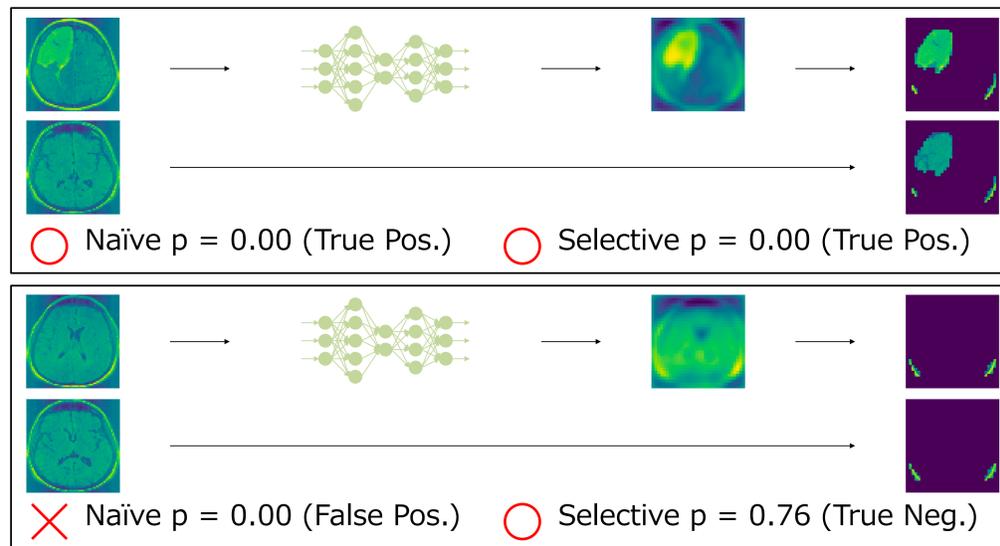
Abstract

深層学習を科学研究で活用する際には予測性能に加え、説明性・解釈性が重要である

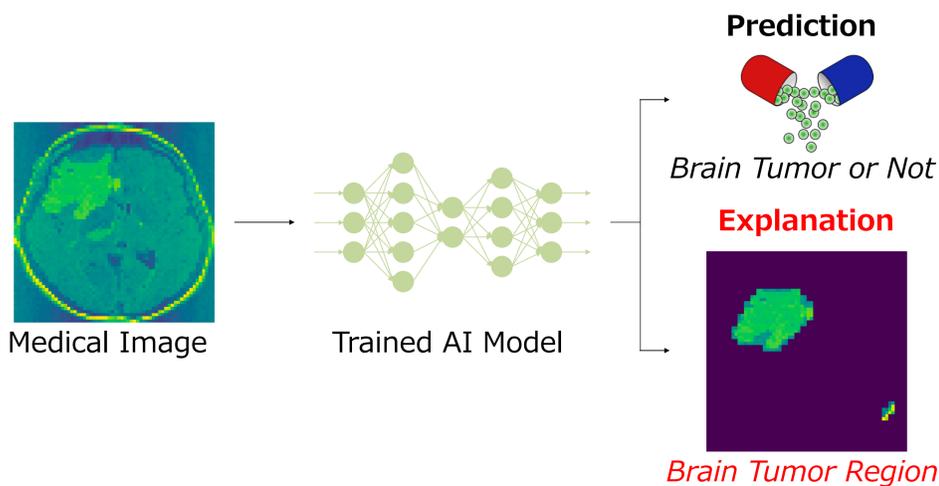
説明性・解釈性のための様々な可視化手法が提案されているが、信頼性評価が難しい

深層学習の説明・解釈に対する統計的信頼性を P 値として定量化する方法を提案する

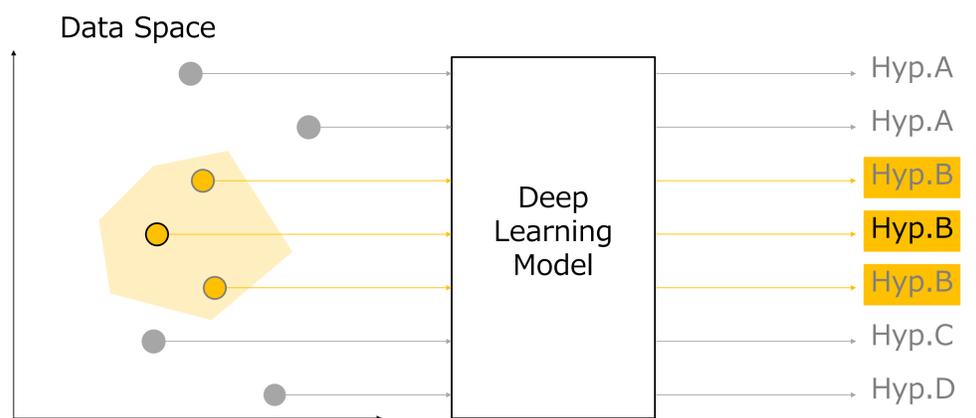
Naïve vs. Selective p-values



Prediction and Explanation in AI

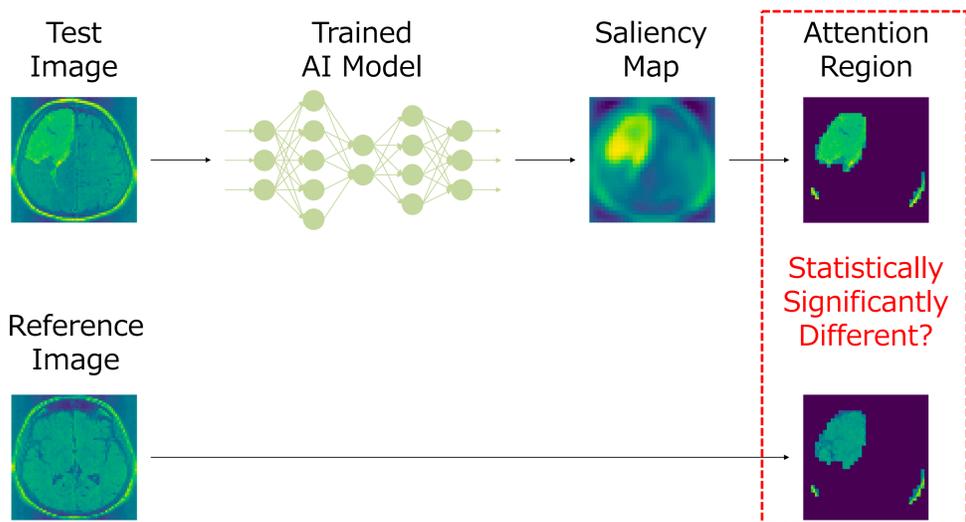


Selective Inference (Intuition)



Selective inference is conducted by conditioning on the selection of a hypothesis to avoid the hypothesis selection bias.

Statistical Test for Explanation



Selective Inference (Formulation)

Test & Reference Images

$$\begin{aligned} \text{Test Image: } X &= \underbrace{s}_{\text{signal}} + \underbrace{\varepsilon}_{\text{noise}}, & \varepsilon &\sim \mathcal{N}(\mathbf{0}, \Sigma) \\ & \text{Normally-distributed noise} \\ \text{Reference Image: } X^{\text{ref}} &= \underbrace{s^{\text{ref}}}_{\text{signal}} + \underbrace{\varepsilon^{\text{ref}}}_{\text{noise}}, & \varepsilon^{\text{ref}} &\sim \mathcal{N}(\mathbf{0}, \Sigma) \\ & \text{Normally-distributed noise} \end{aligned}$$

Algorithm (e.g., CAM, Grad-CAM, Attention)

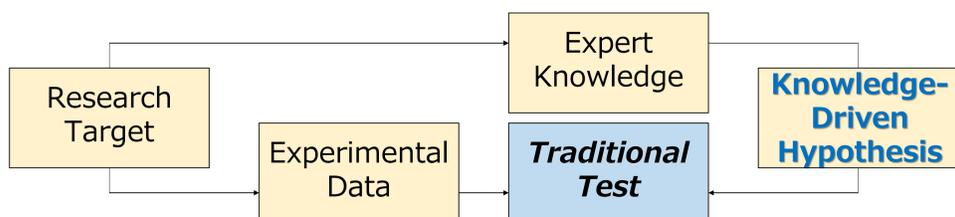
$$\underbrace{A}_{\text{algorithm}} : \underbrace{X}_{\text{image}} \mapsto \underbrace{\mathcal{M}_X}_{\text{attention region}}$$

Hypothesis Test

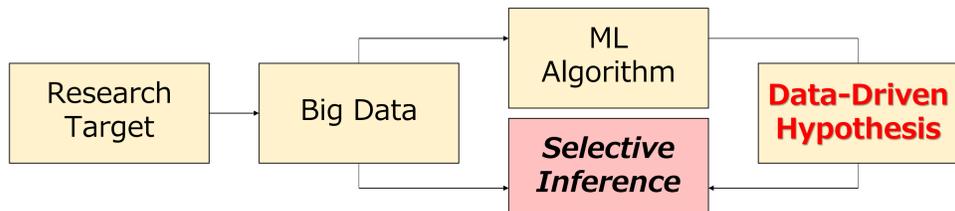
$$H_0 : s_i = s_i^{\text{ref}} \forall i \in \mathcal{M}_X \quad \text{vs.} \quad H_1 : s_i \neq s_i^{\text{ref}} \exists i \in \mathcal{M}_X$$

Knowledge-Driven vs. Data-Driven

Knowledge-Driven Science



Data-Driven Science



Selective Inference Literature

選択的推論は特徴選択（例：Lasso）にののための統計的推測法として2016年頃から研究が盛んになった

特徴選択に限らず、様々なデータ駆動仮説の統計的信頼性を評価できることが認知され始めている

問題によっては、検出力が低い、計算コストが高い、などの課題があり研究が進展中である

[1] Lee et al. Exact post-selection inference, with application to the lasso. Annals of Statistics (2016). [2] Duy et al. Quantifying Statistical Significance of Neural Network-based Image Segmentation by Selective Inference. NeurIPS (2022) [3] Miwa et al. Valid P-Value for Deep Learning-Driven Salient Region. ICLR (2023)