UNIVERSITY OF CAMBRIDGE

Getting a CLUE: A Method for Explaining Uncertainty Estimates

Showcase Talk, CCAIM 2021

Javier Antorán, Umang Bhatt, Tameem Adel,

Adrian Weller, and José Miguel Hernández-Lobato

About Us

Javier Antorán
ja666@cam.ac.uk

Umang Bhatt
usb20@cam.ac.uk

Tameem Adel
tah47@cam.ac.uk

Adrian Weller
aw665@cam.ac.uk

José Miguel Hernández-Lobato
jmh233@cam.ac.uk

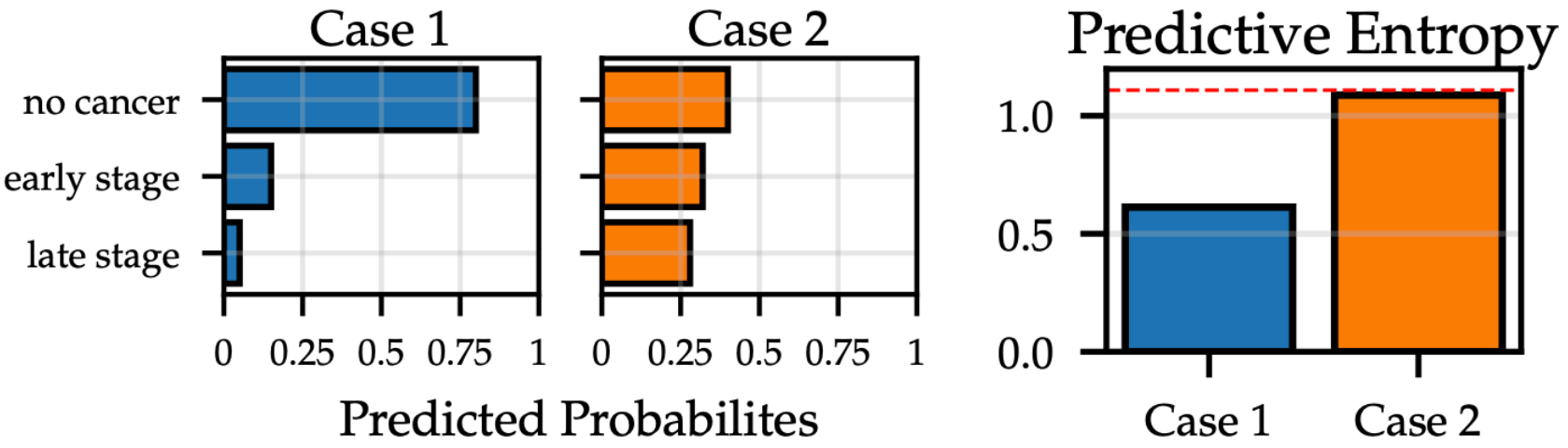




Computational and Biological Learning Lab

UNIVERSITY OF CAMBRIDGE

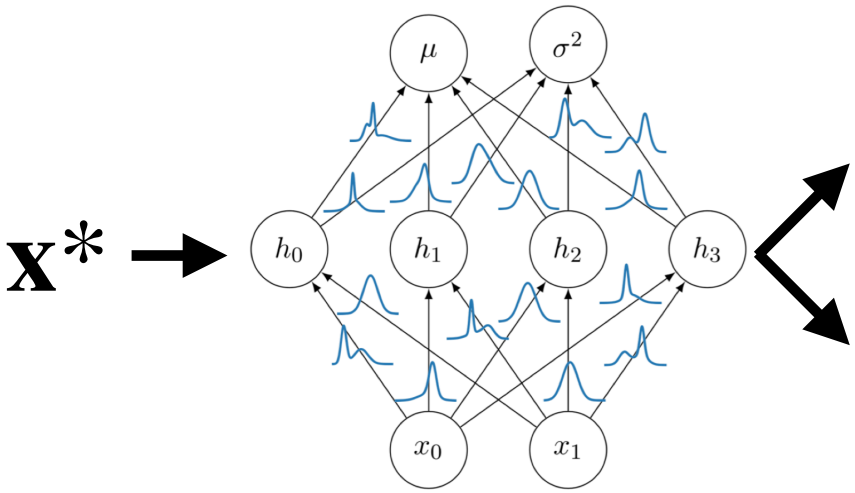
Data driven decision making with reliable uncertainty



Motivation: More Transparency in Deep Learning via Uncertainty

- LIME [Ribeiro et al., 2017]
- SHAP [Lundberg et al., 2017]
- Integrated Gradients [Sundararajan et al., 2017]
- FIDO [Chang et al., 2017]

ML Practitioner Workflow:



Accept a Certain Prediction → **Get Explanation**

Reject an Uncertain Prediction → ?

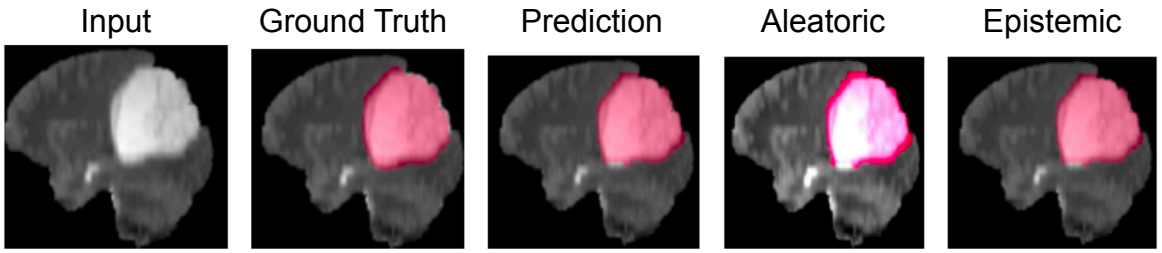
Are Uncertainty Aware Systems Interpretable?

Polyp segmentation:

(a) input image (c) EFCN-8 prediction (e) EFCN-8 uncertainty

Wickstrøm, et. al., 2019

MRI brain tumour segmentation:



Sagar, 2020

Opening the Black Box with Counterfactual Explanations

Input
(99.9%)

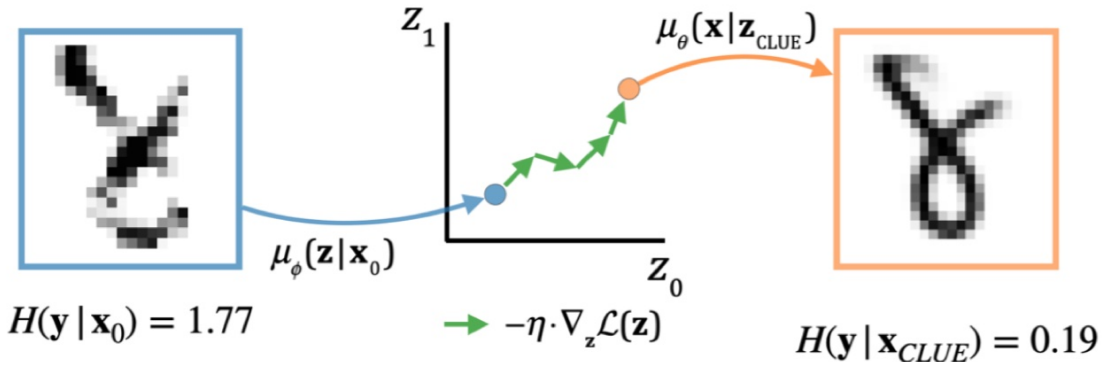
CA
(25.9%)

Chang et. al., 2018

Getting a CLUE

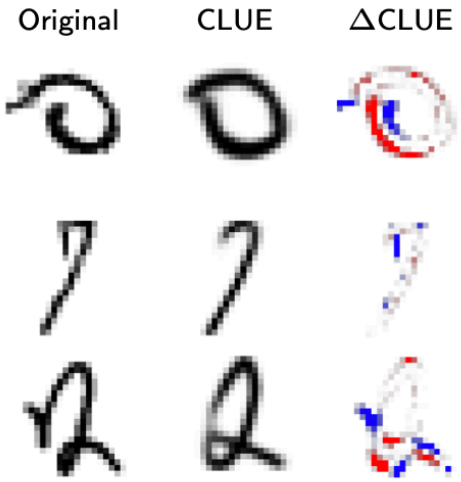
Counterfactual **L**atent **U**ncertainty **E**xplanations

"What is the **smallest change** we need to make to an input such that our model produces more **certain predictions**"



Showing CLUEs to Users

$$\Delta \mathbf{x} = \mathbf{x}_{\text{CLUE}} - \mathbf{x}_0$$



(a) MNIST

	Original
LSAT	41.0
UGPA	3.7
Race	Asian
Sex	Female

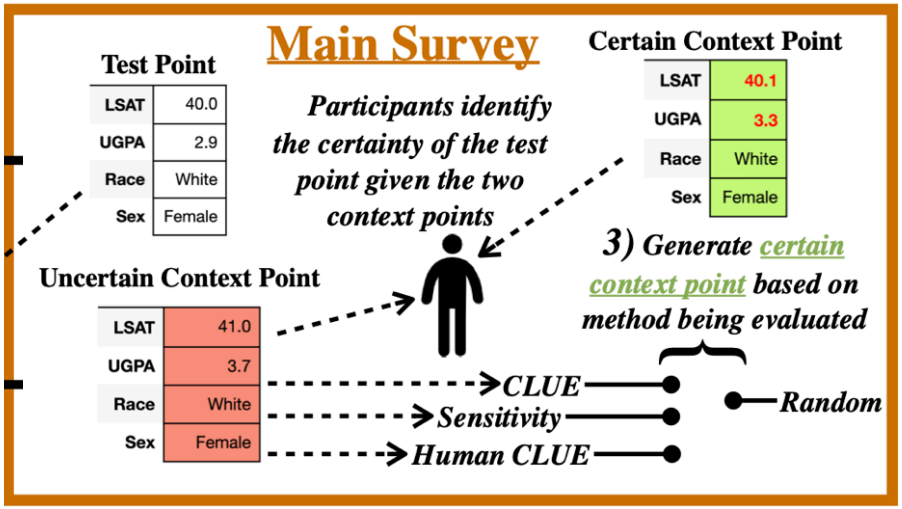
	CLUE
LSAT	41.0
UGPA	3.0
Race	White
Sex	Female

(b) LSAT

Figure 5: Example image and tabular CLUEs.

User Study: Setup

Human Simulability: Users are shown context examples and are tasked with predicting model behaviour on new datapoint.



User Study: Results

Tasks:

- **COMPAS (Criminal Recidivism Prediction, 7 dim)**
- **LAST (Academic Performance Prediction, 4 dim)**

Users:

- **University Students with ML experience**
- **10 Users per approach (total: 40), 10 Questions per Dataset**

Method	N. participants	Accuracy (%)
Random	10	61.67
Sensitivity	10	52.78
Human	10	62.22
CLUE	10	**82.22**

Summary

- Predictive Uncertainty makes ML systems safer and more reliable
- We introduce CLUE, a method to answer the question: "**How should we change an input such that our model produces more certain predictions?**"
- Our user study finds that CLUEs help users understand the sources of a model's uncertainty.
- Future work will apply CLUE to more diverse and interesting data

Read The Full Paper at:
arxiv.org/abs/2006.06848