



Intelligent
Embedded Systems



Adaptive Shapley:

Using Explainable AI with Large Datasets to Quantify the
Impact of Arbitrary Error Sources

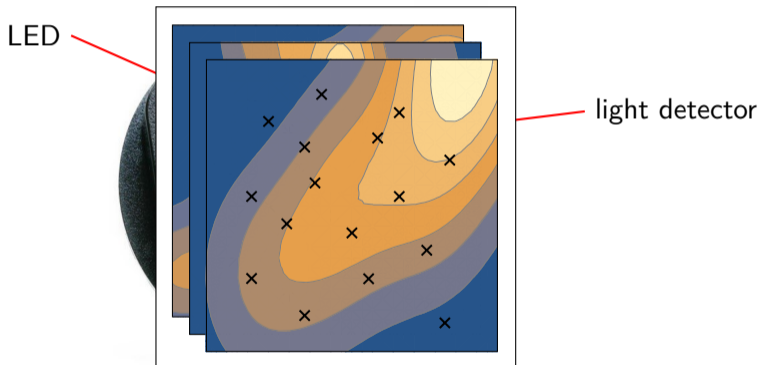
MSRRS-based Sensors



MSRRS-based Sensors



MSRRS-based Sensors



MSRRS-based Sensors

Production Difficulties



MSRRS-based Sensors

Production Difficulties



MSRRS-based Sensors

Production Difficulties



MSRRS-based Sensors

Production Difficulties



Goal: Understand influence of quality criteria on neural network prediction accuracy

Shapley Values

An Introduction

Given a **coalitional** game

Shapley values represent the contribution of one **player** to a **game**

Shapley Values

An Introduction

Shapley values represent the contribution of one **player** to a **game**

Player
?

Game
?

Shapley Values

Players

Player

?

Game

?

Shapley Values

Players

Player

detectors

Game

?

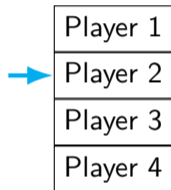
Shapley Values

Players

Player 1
Player 2
Player 3
Player 4

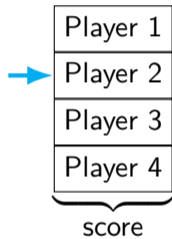
Shapley Values

Players



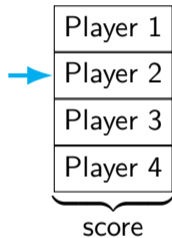
Shapley Values

Players



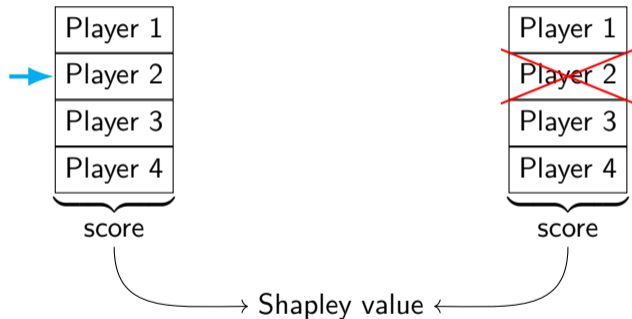
Shapley Values

Players



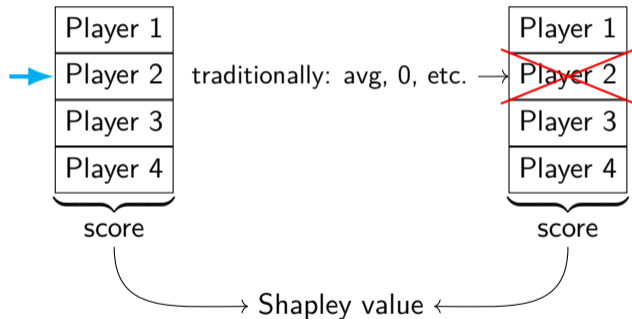
Shapley Values

Players



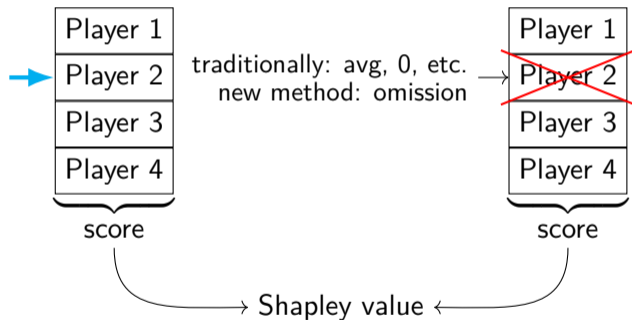
Shapley Values

Players



Shapley Values

Players



Shapley Values

Game Score

Player	Game
detectors	?

Shapley Values

Game Score

Player

detectors

Game

prediction error

Shapley Values

Game Score

$$\text{prediction error} = |\text{ground truth} - \text{prediction}|$$

Shapley Values

Game Score

$$\text{prediction error} = |\text{ground truth} - \text{prediction}|$$

Shapley Values

Game Score

$$\text{prediction error} = |\text{ground truth} - \text{prediction}|$$

↑
approximation
from omitting
detector

Shapley Values

Player	Game
detectors	prediction error

Shapley values represent the contribution of one **detector** to the **measurement error**

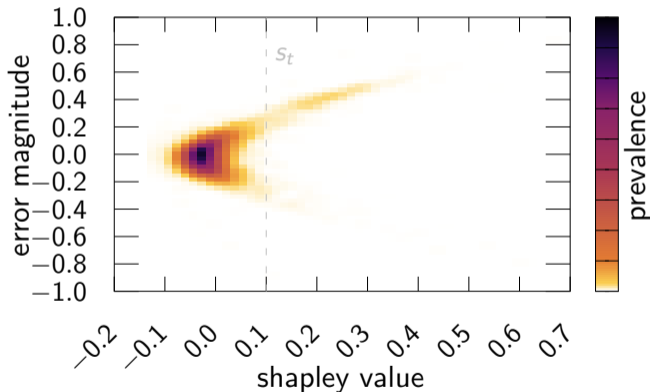
Player	Game
detectors	prediction error

Shapley values represent the contribution of one **detector** to the **measurement error**

→ Calculate Shapley values for a large dataset and average for each sensor

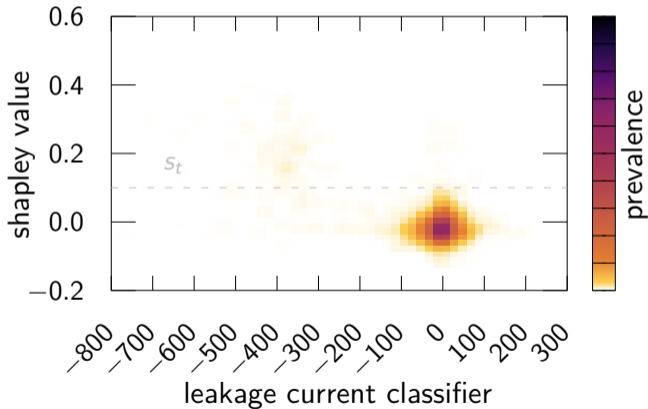
Shapley Value Interpretation

Error Magnitude



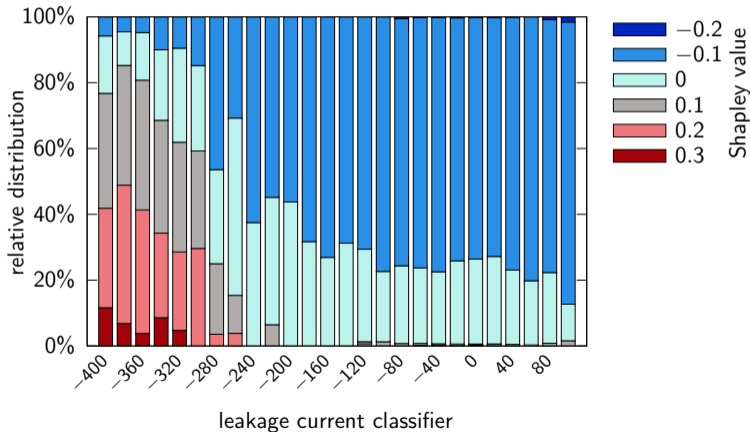
Shapley Value Interpretation

Leakage Current Classifier Heatmap



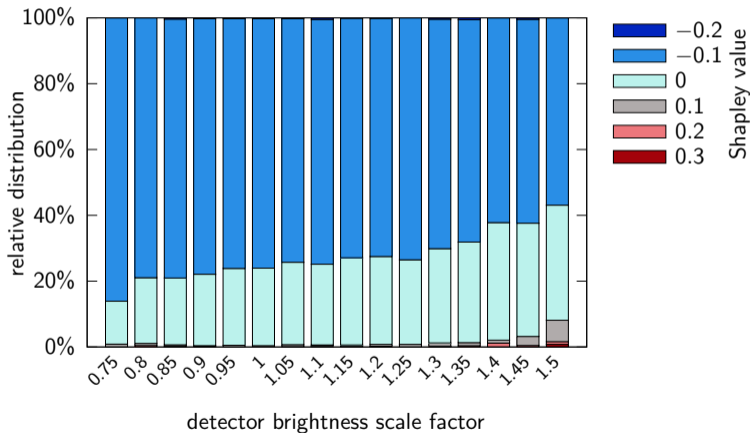
Shapley Value Interpretation

Leakage Current Classifier Histogram



Shapley Value Interpretation

Detector Scale



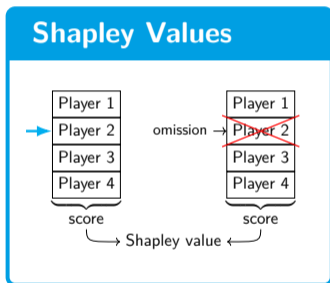
Applying Shapley Values



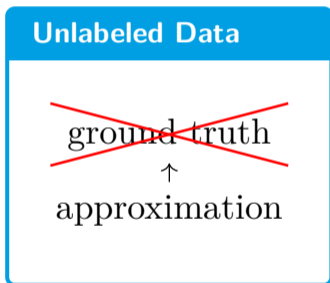
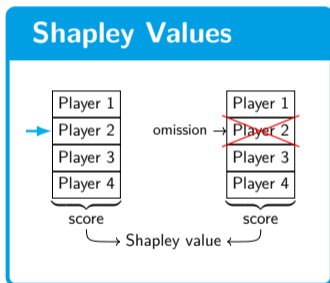
Applying Shapley Values



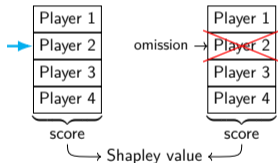
Conclusion



Conclusion



Shapley Values



Unlabeled Data

~~ground truth~~
↑
approximation

Results

