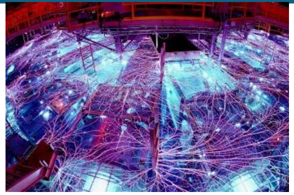
SAND2022-13662C

Sandia National Laboratories

Shapley Additive Explanations for Traveling Wave-based Protection on Distribution Systems

Miguel Jimenez Aparicio, Matthew J. Reno and Felipe Wilches Bernal



PRESENTED BY

Miguel Jimenez Aparicio

Electric Power Systems Research Department
Sandia National Laboratories
Albuquerque, NM

October 2022

U.S. DEPARTMENT OF ENERGY NNSA

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

About myself

Miguel Jimenez Aparicio

Member of Technical Staff @ Electric Power Systems Research Department

Sandia National Laboratories, Albuquerque NM

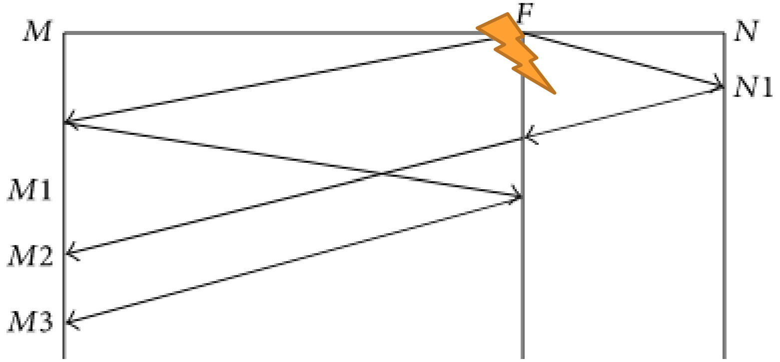
- Main area of research: Fast protection for power distribution systems
  - Designing signal-processing techniques for Traveling Wave detection and data extraction
  - Training Machine-Learning/ Deep-Learning (ML/DL) models for fast fault location

Introduction (I)

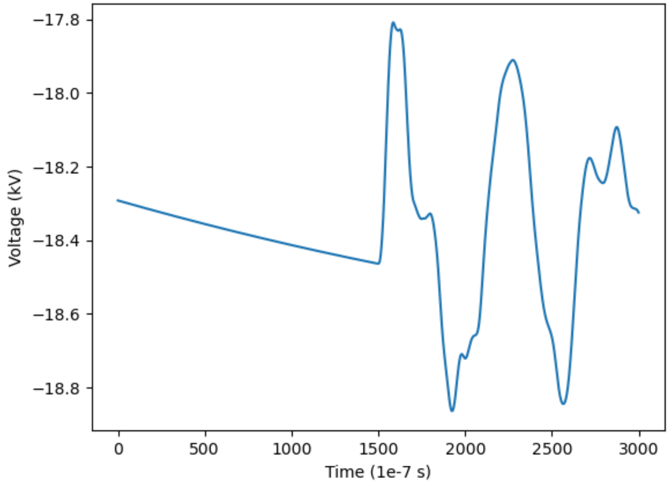
Traveling Waves facts:

- Propagation at almost the speed of light
- Wide frequency waves (from a few kHz to MHz)
- Attenuation and distortion due to:
  - Line impedance
  - System discontinuities, such as junctions, shunt elements, etc.

By analyzing the frequency components, we are able to back track the propagation path and get the fault location!



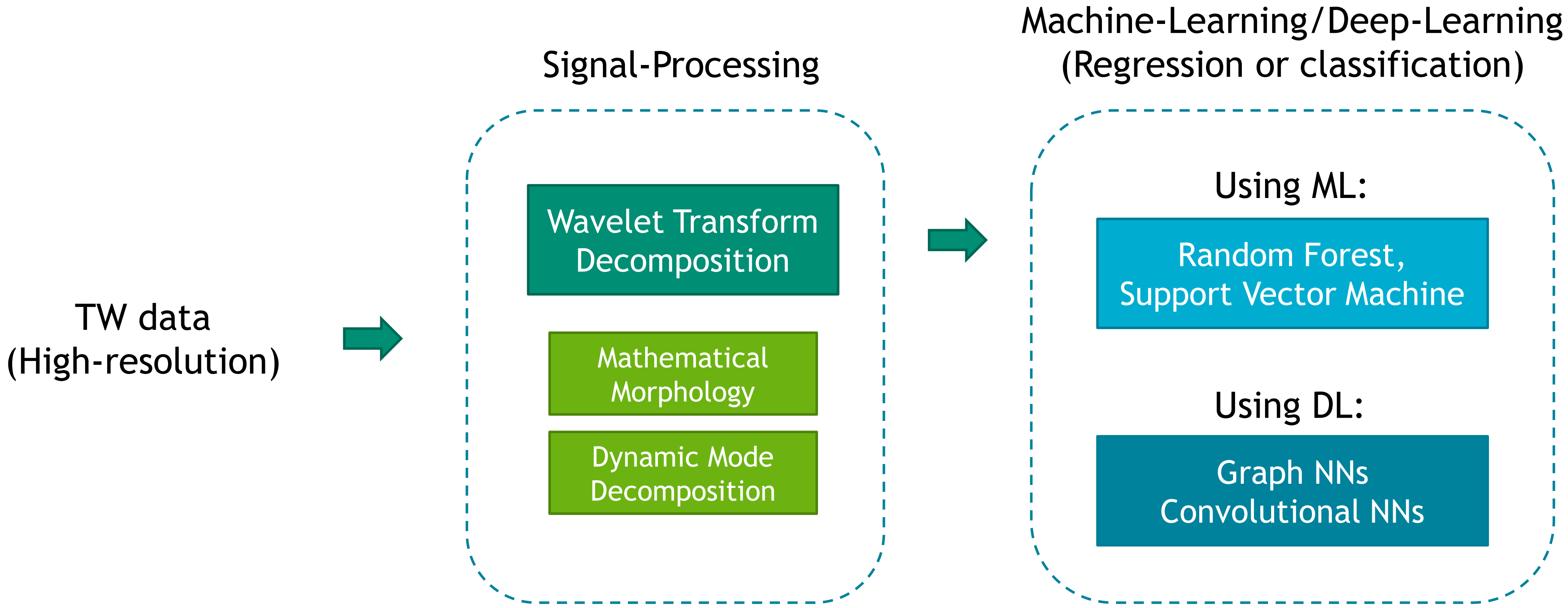
Traveling Waves are generated by moving charges in a fault



Traveling Wave at a sampling frequency of 10 MHz

Introduction (II)

Fault location using TWs



Introduction (III)

We can train ML models for estimating fault location, but …

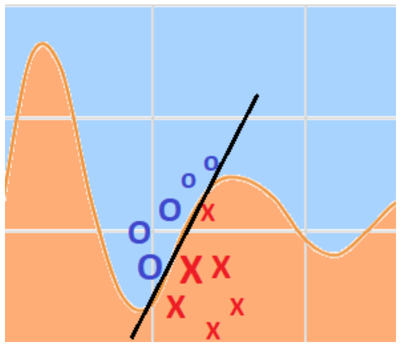
- How does a model use data to give an individual prediction in a certain fault case?
- How does a model's prediction change when TW propagation conditions are different?

Essentially, how does a ML fault locator behave, and how it can be quantified?

→ **Interpretation/explainability of ML models is one of the biggest challenges ahead**

Background (I)

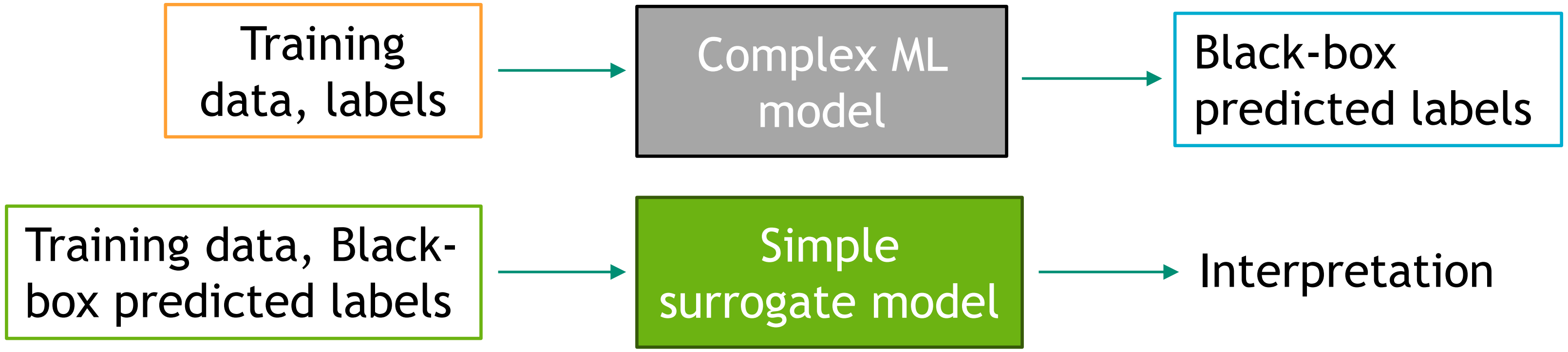
LIME



Model-agnostic interpretation techniques:

- Individual Conditional Expectation
- Local Interpretable Model-agnostic Explanations (LIME)
- Surrogate models
- SHapley Additive exPlanations (SHAP)

Surrogate model



Background (II)

SHapley Additive exPlanations (SHAP):

- Based on game theory
- Applicable to any ML model
- For individual predictions

Number of coalitions *M*

Subset of features including feature *j*

$$\phi_j = \frac{1}{M} \sum_{m=1}^{M} (f(x^m_{+j}) - f(x^m_{-j}))$$

Shapley value for a feature *j*

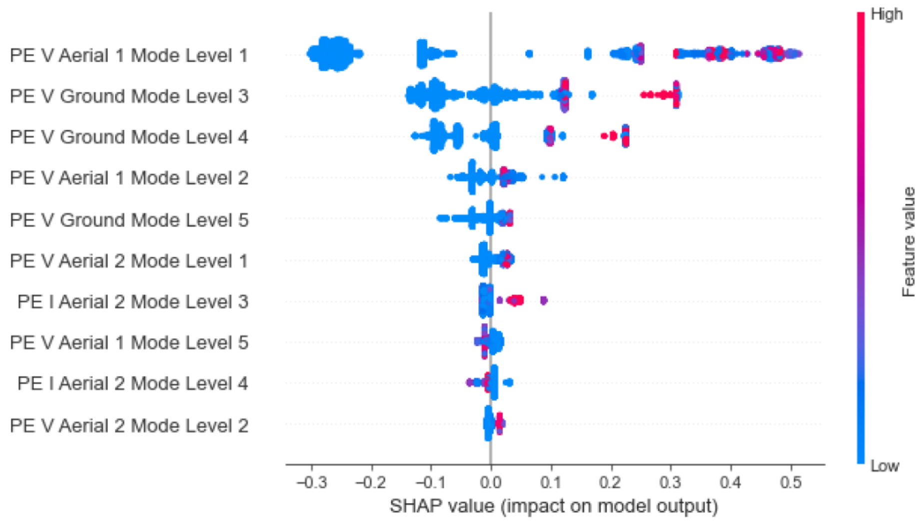
Subset of features excluding feature *j*

"For a given prediction, SHAP studies what is the average contribution of each feature in the result under several coalitions of features."

Background (III)

Insights provided by SHAP:

- Global feature importance
- Feature importance on individual predictions

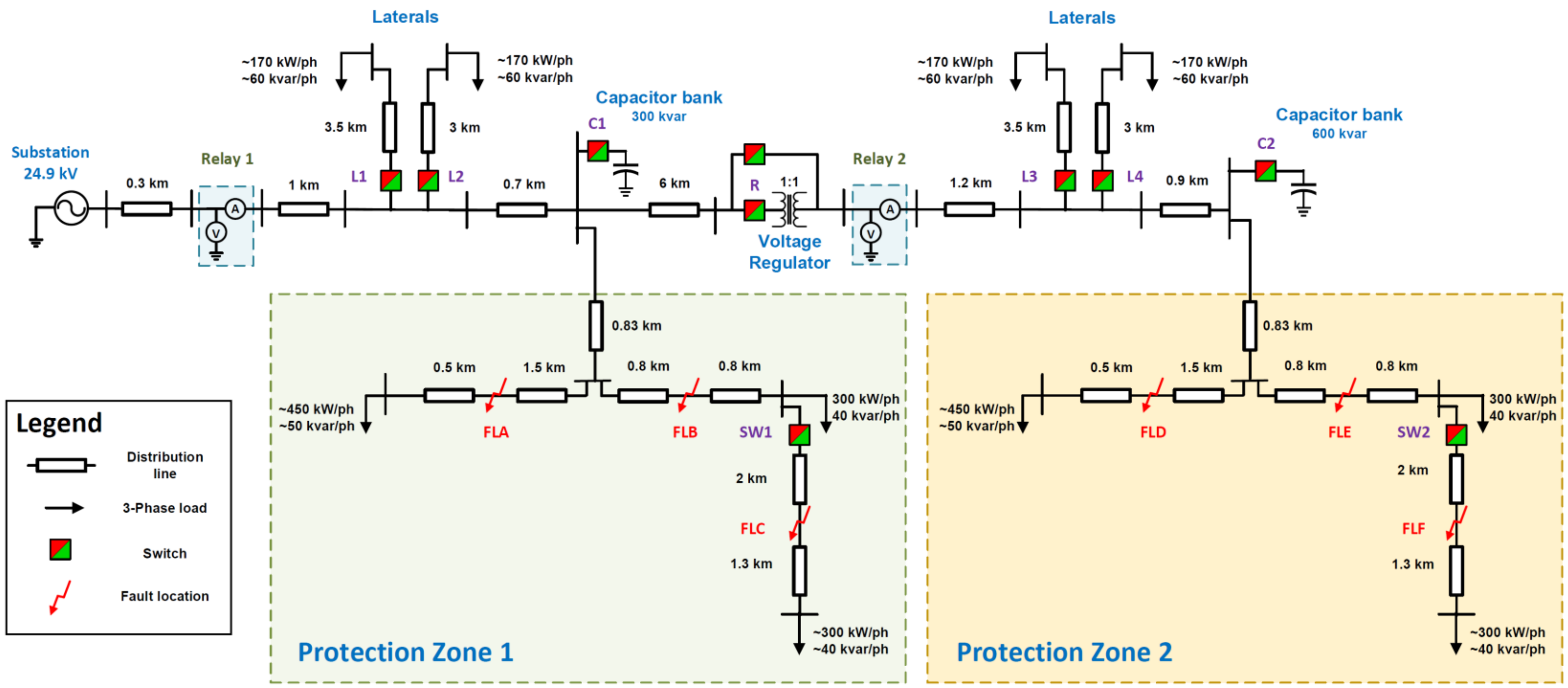


The Use Case

Studying TW fault locator on…

Task 1: Area protection → Predict fault location on Protection Zone 1 or 2 from Relay 1
Task 2: Is the extrapolation of the ML model in PZ1 Relay 1 to PZ2 Relay 2 feasible?

Simplified distribution system



6,440 fault simulations:

- 7 fault types
- 5 fault resistance values
- 6 fault locations
- 3 laterals combinations
- 3 extra branches combinations
- 2 regulator combinations
- 4 capacitor banks combinations

Adding variability to study the classifier behavior under different conditions

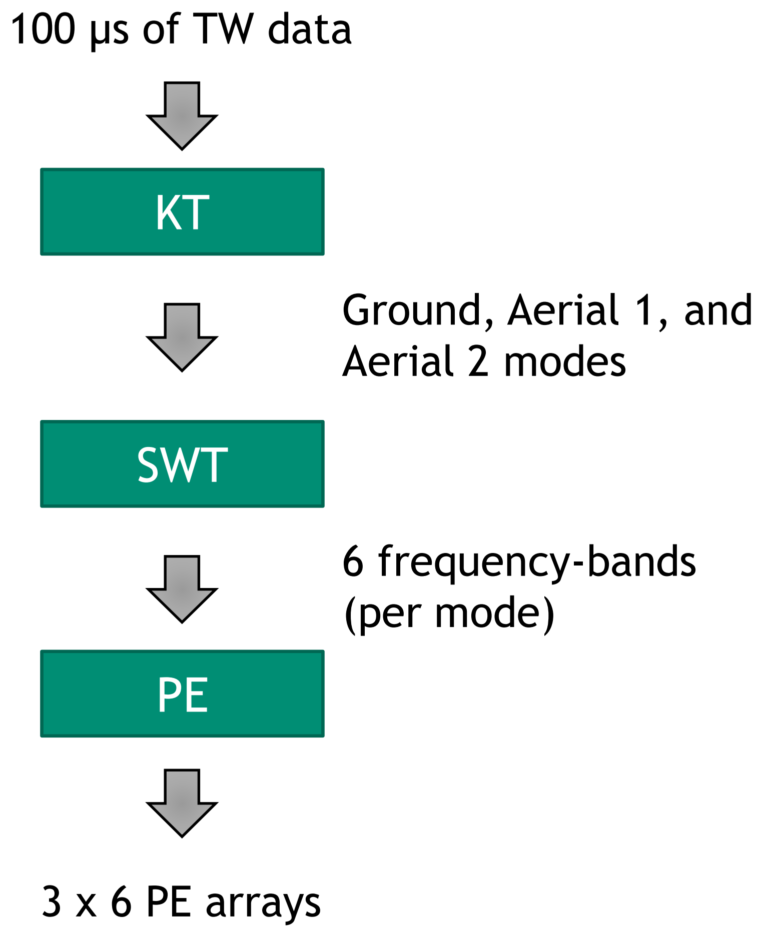
Signal-processing stage and ML model (I)

Signal-processing stage:

1. Karrenbauer Transform (KT)
2. Stationary Wavelet Transform (SWT)
3. Parseval's Energy (PE)

TABLE I: SWT Boundaries for Frequency Bands

Decomposition Level	Lower Frequency	Upper Frequency
1	2.5 MHz	5 MHz
2	1.25 MHz	2.5 MHz
3	625 kHz	1.25 MHz
4	312.5 kHz	625 kHz
5	156.25 kHz	312.5 kHz
6	78.125 kHz	156.25 kHz



Signal-processing stage and ML model (II)

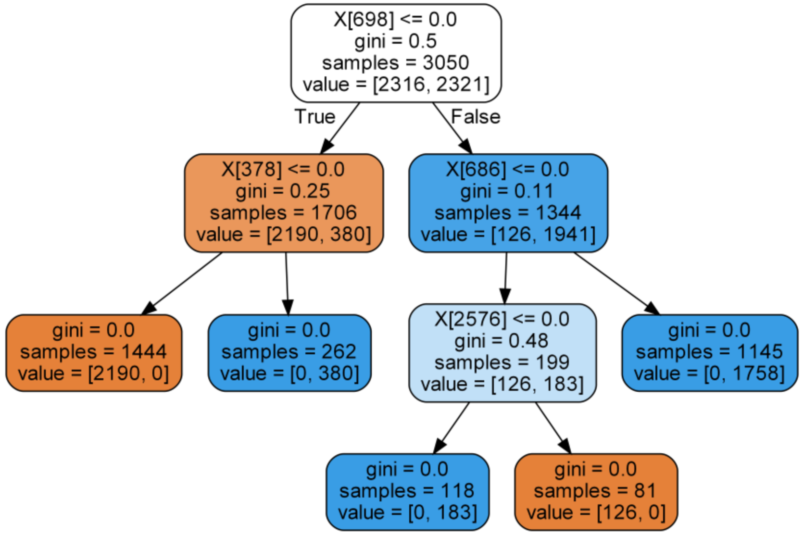
Models are based on Random Forest (RF)

Internally, a RF is composed of several decision-tree estimators that select relevant energy values in the time-series

80% data for training, 20% data for testing

3 approaches:
A) 1 model, 1 mode (ground mode)
B) 1 model, 3 modes
C) 3 models, 3 modes (one per mode)

Example of one RF estimator



Results (I)

Accuracy results for Task 1 and 2

Model	Overall	SLG	LL	3P
Area Protection				
1 model/ 1 mode	100%	100%	100%	100%
1 model/ 3 modes	100%	100%	100%	100%
3 models/ 3 modes	100%	100%	100%	100%
Extrapolation to PZ2 R2				
1 model/ 1 mode	61.0%	86.2%	39.9%	48.9%
1 model/ 3 modes	63.4%	81.2%	50.0%	50.0%
3 models/ 3 modes	62.8%	80.8%	49.3%	50.0%

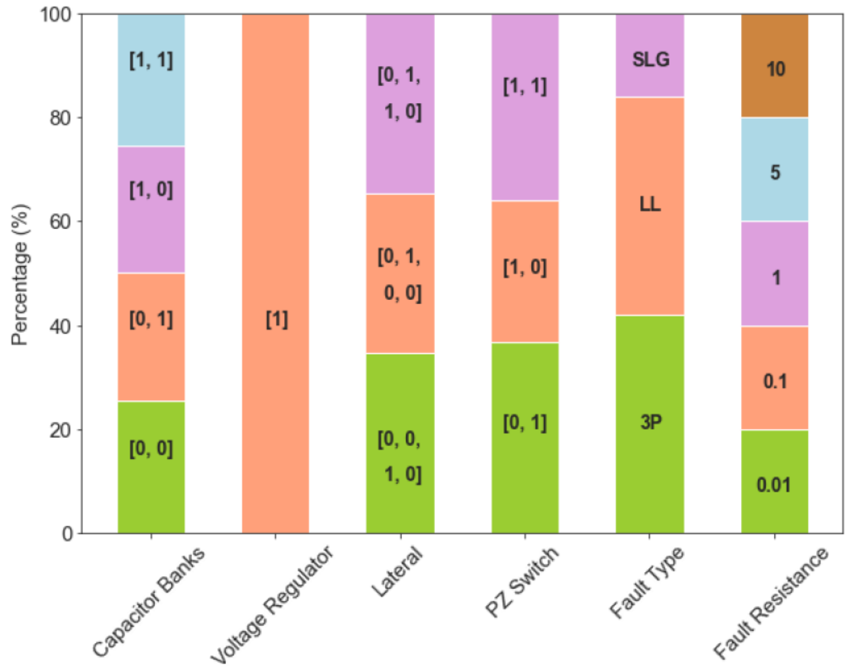
Task 1: Predict fault location on Protection Zone 1 or 2 from Relay 1

100% accuracy, distance between zones is key

Task 2: Is the extrapolation of the ML model in Relay 1 to Relay 2 feasible?

There are differences that lead to prediction failures

Distribution of incorrect prediction per factor (for Extrapolation to PZ2 R2 of 1 model/ 3 modes approach)



Analysis:

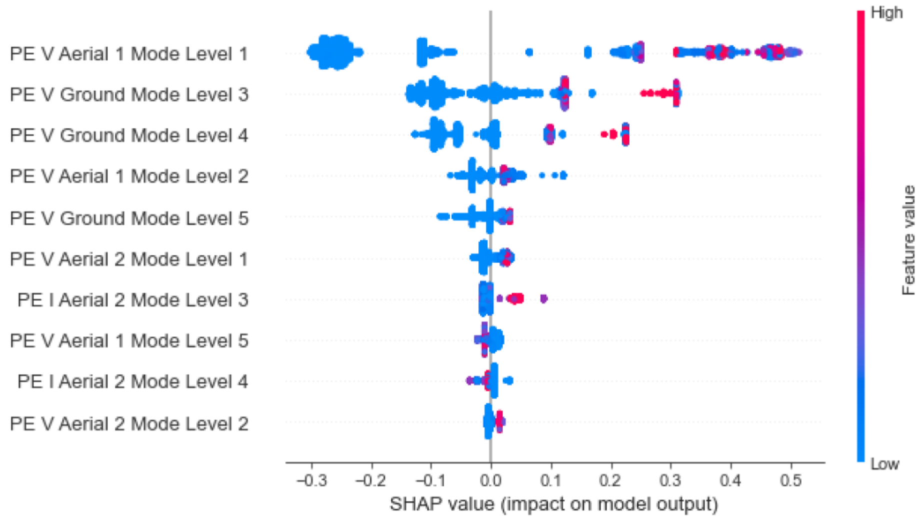
- What are the reasons behind the failures?
- How does the classifier behavior change when deployed on PZ2 Relay 2?

Results (II)

Possible reasons are:

- Attenuation due to capacitor banks
- Attenuation due to the regulator
- Additional TW reflections due to additional laterals in the main backbone
- Additional TW reflections due to additional laterals inside the Protection Zones

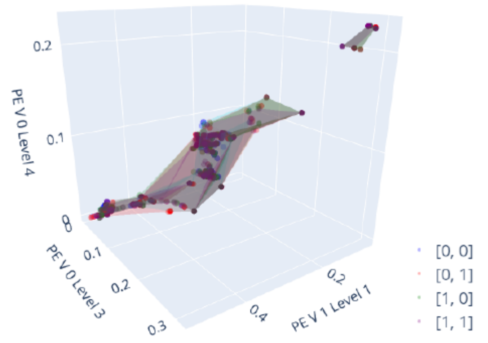
Determining the most important factors → The effect of a factor is more relevant if produces larger changes on the features' importance (given by SHAP)



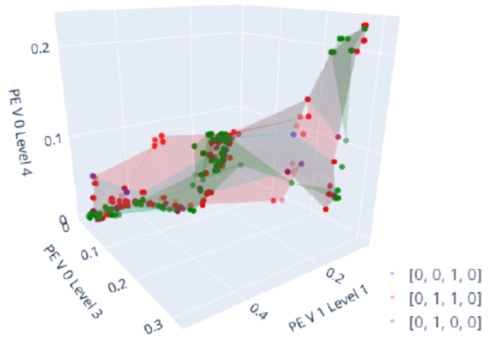
Top 10 most important features according to SHAP (1 model/ 3 modes approach)

Results. Analysis of SHAP values (I)

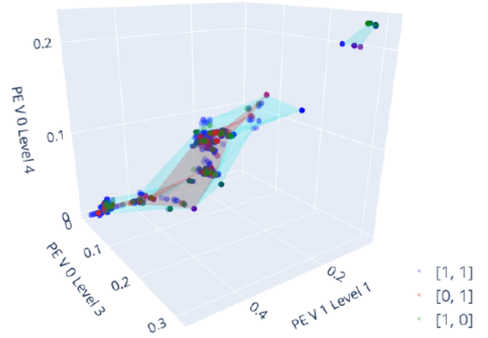
Area protection experiment (High accuracy)



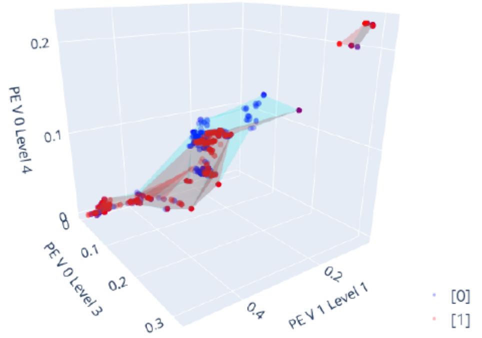
(a) Capacitor banks



(b) Laterals



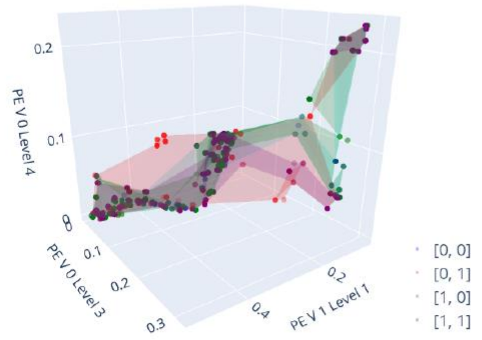
(c) PZ extra branch



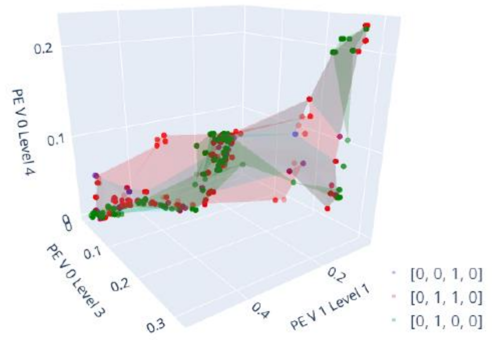
(d) Voltage regulator

Fig. 4: 3D analysis of SHAP distributions per combinations of each factor for PZ1 and PZ2 R1

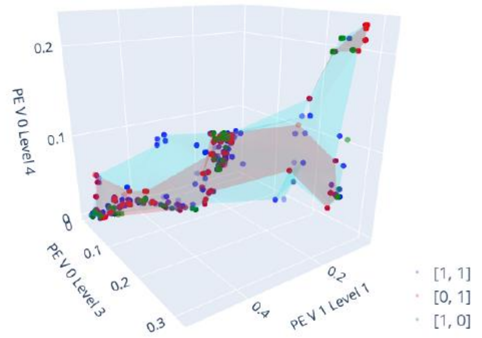
Extrapolation to PZ2 R2 experiment (Low accuracy)



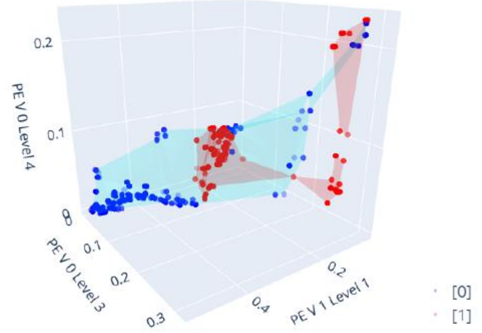
(a) Capacitor banks



(b) Laterals



(c) PZ extra branch



(d) Voltage regulator

Fig. 5: 3D analysis of SHAP distributions per combinations of each factor for PZ2 R2

Results. Analysis of SHAP values (II)

Jensen-Shannon Divergence:
This metric quantifies the similarity between two probability distributions (X and Y)

$$JSD(X||Y) = H\left(\frac{X+Y}{2}\right) - \left(\frac{H(X)+H(Y)}{2}\right)$$

Shannon Entropy

The voltage regulator has a large effect on the classifier behavior for both the "Area Protection"…

and "Extrapolation to PZ2 R2" experiments

TABLE VII: Top Divergences per Level PZ1 R1

PE Level	Factor	Comb. 1	Comb. 2	$\sqrt{\text{JSD}}$
V Aer. 1 Lev. 1	Lateral	[0, 0, 1, 0]	[0, 1, 0, 0]	0.772
V Gnd. Lev. 3	Lateral	[0, 0, 1, 0]	[0, 1, 1, 0]	0.606
V Gnd. Lev. 4	Lateral	[0, 0, 1, 0]	[0, 1, 0, 0]	0.526

TABLE VIII: Top Divergences per Level PZ2 R1

PE Level	Factor	Comb. 1	Comb. 2	$\sqrt{\text{JSD}}$
V Aer. 1 Lev. 1	Regulator	[1]	[0]	0.616
V Gnd. Lev. 3	Regulator	[1]	[0]	0.644
V Gnd. Lev. 4	Lateral	[0, 0, 1, 0]	[0, 1, 0, 0]	0.593

TABLE IX: Top Divergences per Level PZ2 R2

PE Level	Factor	Comb. 1	Comb. 2	$\sqrt{\text{JSD}}$
V Aer. 1 Lev. 1	Regulator	[1]	[0]	0.819
V Gnd. Lev. 3	Regulator	[1]	[0]	0.805
V Gnd. Lev. 4	Regulator	[1]	[0]	0.757

Conclusions

- The behavior of TW-based fault location classifier has been analyzed
- SHapley Additive exPlanations (SHAP) provides the features' contribution that led to a prediction
- JSD quantifies how similar (or not) the RF classifier behaves for faults under different circumstances

- Two tasks were considered:
  - Area protection: The classifier's behavior is consistent for different fault and system configurations. The effect of the voltage regulator is noticeable in faults from PZ2, but the classifier can cope with this variability.
  - Extrapolation to PZ2 R2: The effect of the voltage regulator between PZ1 and PZ2 significantly disturbs the classifier behavior, and accuracy is much lower.