CERN

MACHINE LEARNING MODELS FOR BREAKDOWN PREDICTION IN RF CAVITIES FOR ACCELERATORS

Christoph Obermair, TE-MPE-CB
Thomas Cartier-Michaud, TE-MPE-CB
Andrea Apollonio, TE-MPE-CB
Lorenz Fischl, TE-MPE-CB
Lee Millar, CLIC BE-RF
Franz Pernkopf, TU-Graz

Outline

3 main tasks for the successfull icoorporation of ML in TE:

1. Show potential → CLIC RF breakdown prediction
2. Build trust → Explainable AI + Experiments
3. Enable easy application → ML framework

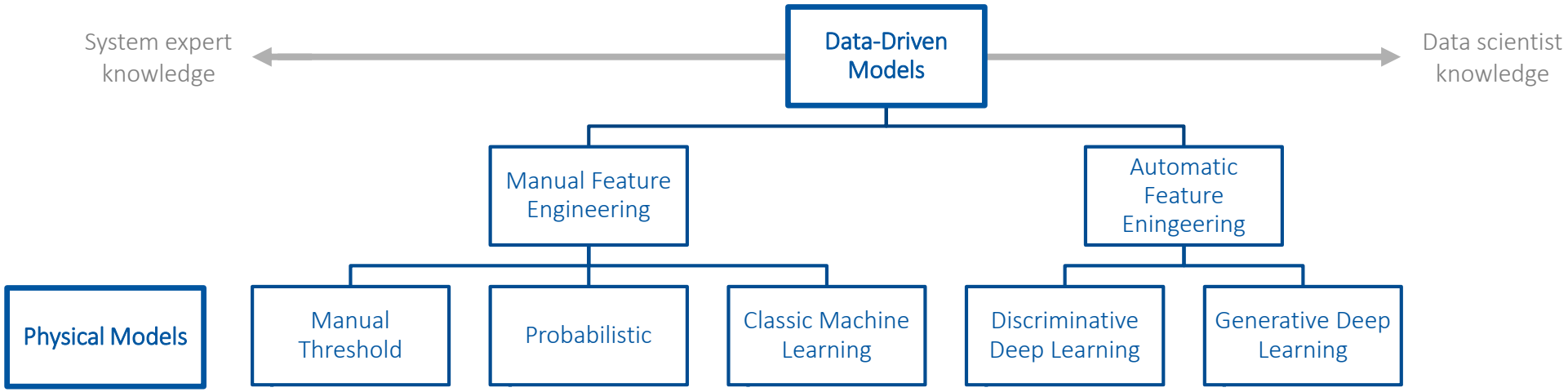
Outline:

1. Introduction
2. Methodology
3. Results trend data analysis
4. Results event data analysis
5. Conclusion
6. Outlook

1. Introduction

a. Overview of Data-Driven Models
b. RF Cavities
c. RF Cavity breakdowns
d. XBOX 2 Test Stand

1.a. Overview of Data-Driven Models



Approach 1:

Use physics knowledge to define model.

Requirement:

Deep understanding of given use case

Approach 2:

Use model to gain physics knowledge.

Requirement:

Data sets to measure performance e.g.: $D = \{(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_N, y_N)\}$

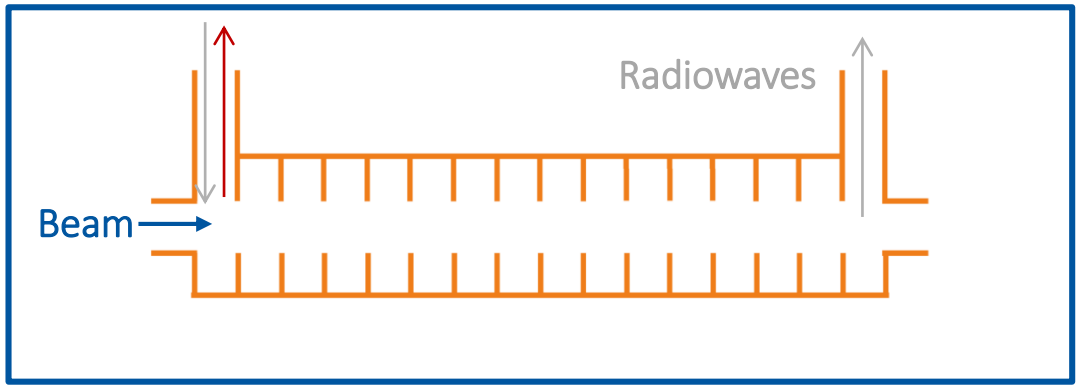
Content from M. Maciejewski and H. Fawaz et al., „Deep learning for time series classification“, 2019

1.b. RF Cavities

Radiofrequency (RF) cavities: To accelerate particles, the accelerators are fitted with metallic chambers containing an electromagnetic field known as radiofrequency (RF) cavities. Charged particles injected into this field receive an electrical impulse that accelerates them.

- In linear accelerators the peak gradient determines the maximum achievable energy
- We analyze the very high gradient (100MV/m) cavities for the Compact Linear Collider (CLIC) *
- CLIC RF cavities are tested in the XBOX test stands at CERN
- Several aplications beyond accelerators (e.g. cancer treatment)

RF cavity



CLIC concept image

* https://home.cern/science/accelerators/compact-linear-collider

1.c. RF Cavity Breakdowns

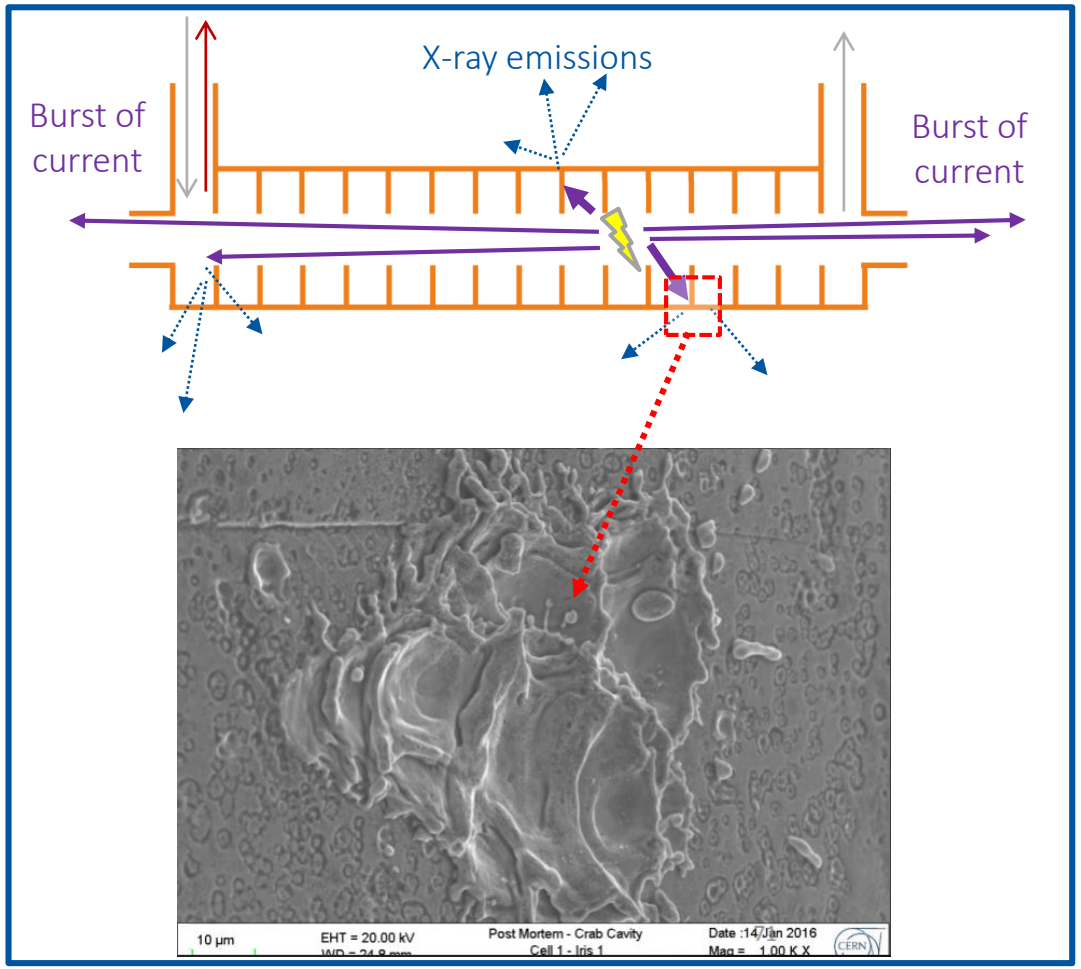
RF cavity breakdown: Small deformities cause field enhancement on surface which can lead to plasma formation (arc) and surface damage.

- Main limitation of high field in the CLIC RF cavities
- Conditioning necessary to minimize breakdown rate
- They manly occur in groups. The first breakdown in each group is called **primary breakdown**, the others are **follow-up breakdowns**.

Goals:

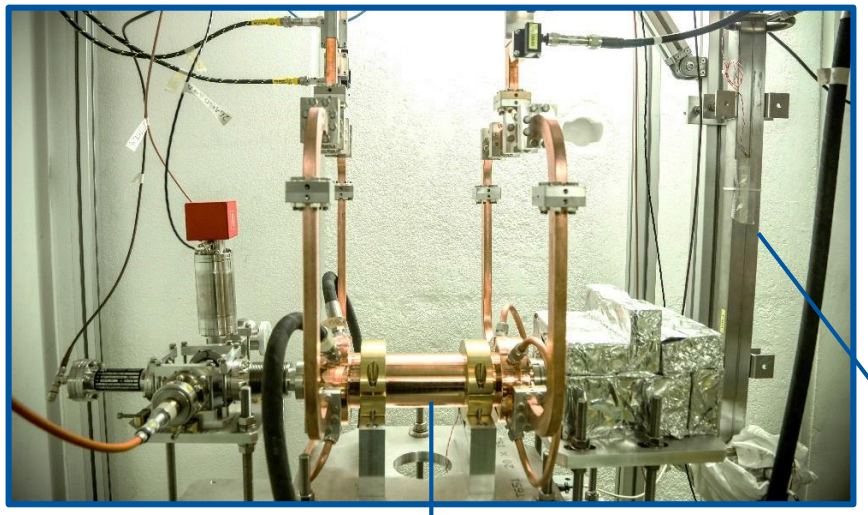
1. Investigate breakdown precursors in CLIC RF structures
2. Operational tool for breakdown reduction, in order to avoid their occurrence by suitable adjustment of the power in the cavity

RF cavity breakdown



Content provided by Lee Millar

1.d. XBOX 2 Test Stand



Abbreviations for analyzed signals
P … Power
K … Klystron
R … Reflected
I … Incident
S … Structure
E … Exit

PKR PKI

Solid State Amplifier

Modulator & klystron

Pulse compressor

RF Cavity

RF Load

PSR
PSI

PEI
PER

Ion Pump

Upstream Faraday Cup

Structure

Downstream Faraday Cup

Content provided by Lee Millar

2. Methodology

a. Analysis Workflow
b. Data Overview
c. Transformation
d. Exploration
e. Modeling
f. Explainable AI

2.a. Analysis Workflow

	Transformation →	Exploration	→	Modeling	→	Execution
Goal	Clean data	Explore data properties		Train & validate model		Implement model into test facility?
		Utilize **unsupervised** and **supervised** machine learning				
Progress	✓	✓		✓		**Future work**

Content from M. Maciejewski

2.b. Data Overview

Data available: 90 GB of Data from 2018 (6months of data)

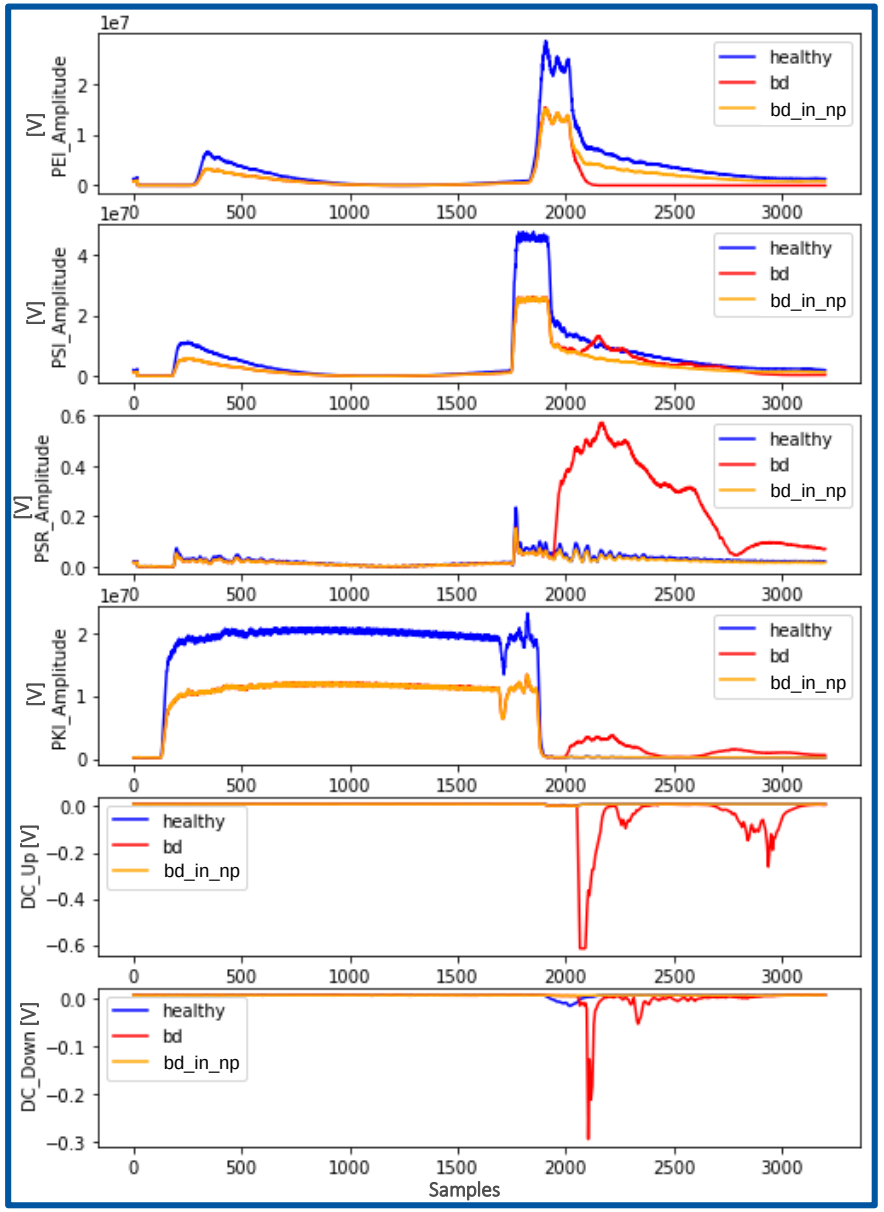
a.) Trend data:

- Context data about other machine parameters:
  i.e. Temperature, pressure (~1Hz sample rate)

b.) Event data:

- Data from test stand sensors: (healthy, breakdown, breakdown in next pulse) 16 Channels with up to 3200 sample points (1.6 GHz sample rate) RF signal is 12 GHz

	Bunker_WG_Temp	Klystron_Flange_Temp	Load_Temp	PC_Left_Cavity_Temp	PC_Right_Cavity_Temp
0	24.134285	24.116962	27.955711	30.285633	30.187632
1	24.082500	24.145723	27.876066	30.277182	30.208776
2	23.241980	24.400513	27.502401	30.253901	30.191576
3	18.783634	19.573864	27.900787	30.100540	30.056396

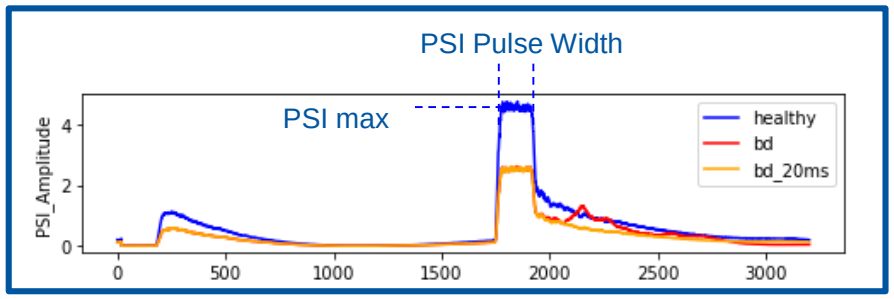


2.c. Transformation

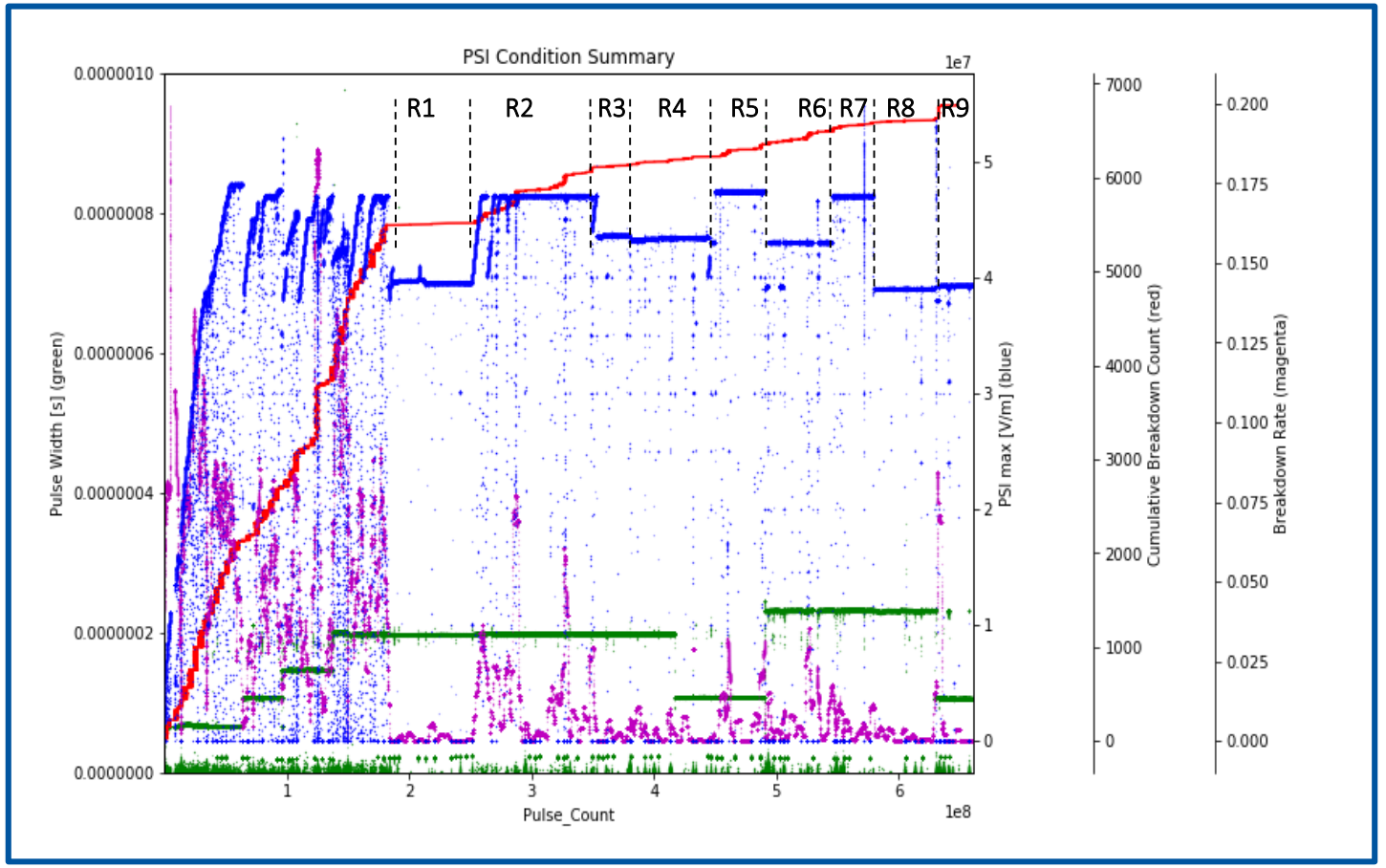
Goal: Clean data, fast queries, memory efficient storage

- Adding of extra features (e.g. pulse width, max, run 1-9)
- Transformation of data into .hdf5 (24 GB)

Extracted Features



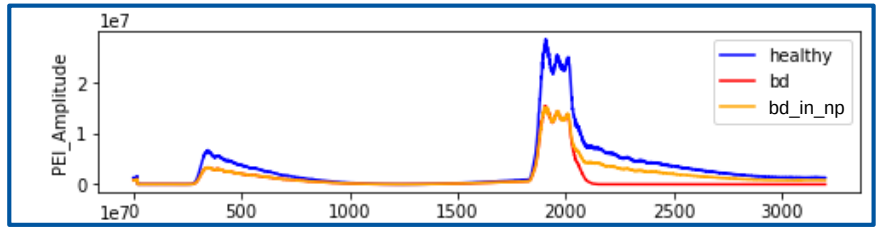
2018 Trend Data



2.c. Merging of Event Data with Trend Data

Input data:

- Different sampling time of event data (~20ms/1min) and trend data (~1s)
- Clean the data such that we do not give the answer (“high contrast”)



1min

Event data: o o o

Trend data: o o o o o o o o o o

1s

oox

Synchronized?

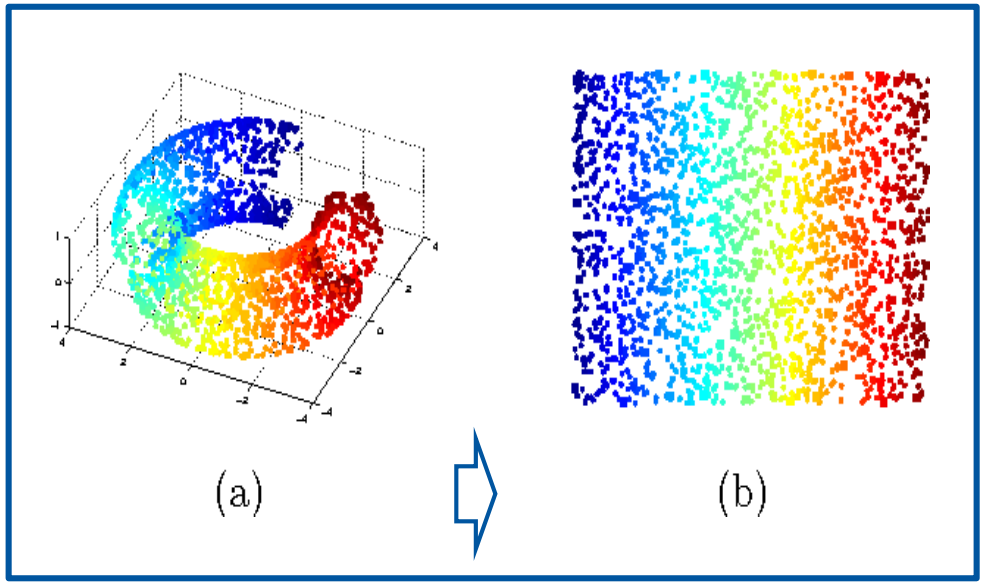
Trend Data Event Data

	Bunker_WG_Temp	Klystron_Flange_Temp	Load_Temp	PC_Left_Cavity_Temp	PC_Right_Cavity_Temp
0	24.134285	24.116962	27.955711	30.285633	30.187632

2.d. Exploration

Goal: General method to explore multidimensional dataset.

- 2D t-SNE (t-Distributed Stochastic Neighbor Embedding)
- Find projection to 2D space by retaining most important information
- Axes lose their physical meaning

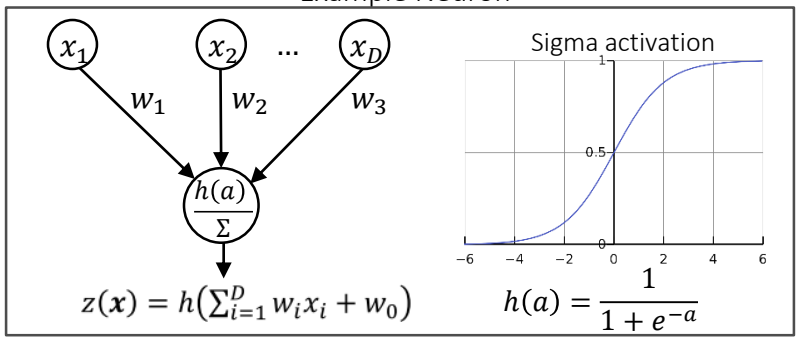


2.e. Modeling

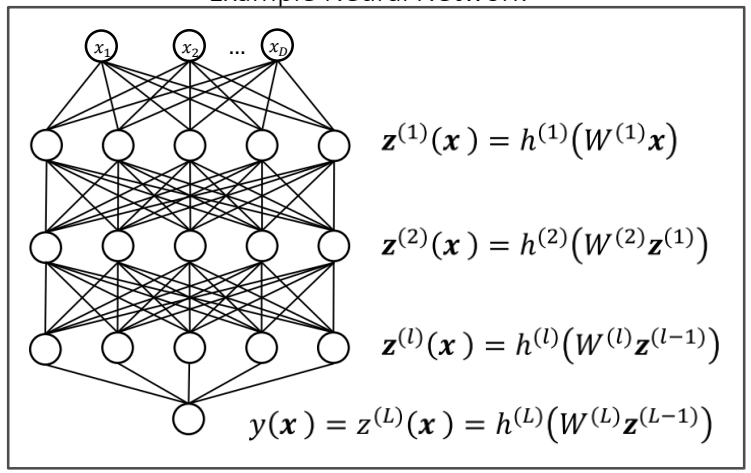
Neuron:

Main component of neural network
x … input samples
w … neuron weights

Example Neuron



Example Neural Network



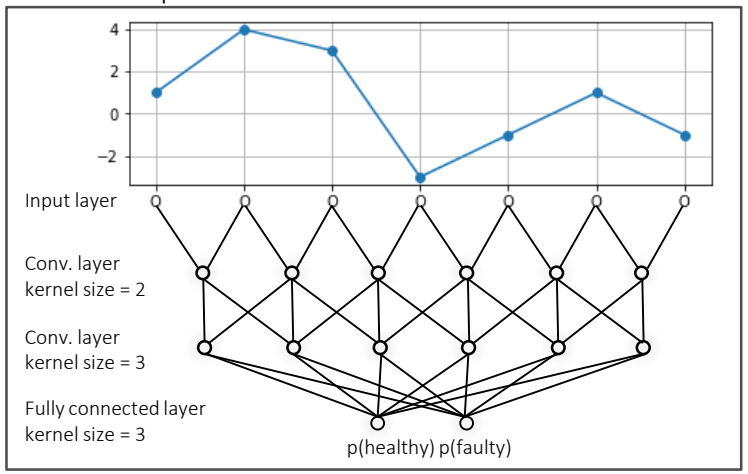
Neural networks:

Many layers of neurons makes it possible to find nonlinear properties (deep learning)

Convolutional neural networks:

A neuron is not connected to every single neuron in the next layer,
but only to neighbouring neurons (convolution).*

Example CNN for Time Series Classification



Explainable AI

Analyze w_i to reverse engineer which sample was important for classification.
"SHAP values"** are calculated and state the importance of each input x_D.

*H. Fawaz et al., „Deep learning for time series classification", 2019
**Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." NeurIPS. 2017.

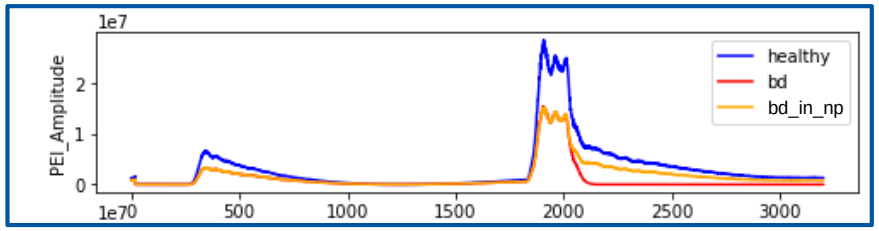
2.e. Modeling

	Bunker_WG_Temp	Klystron_Flange_Temp	Load_Temp	PC_Left_Cavity_Temp	PC_Right_Cavity_Temp
0	24.134285	24.116962	27.955711	30.285633	30.187632

Two modeling stages:

a. Train on trend data + extracted features (scalars)

b. Train on event data (time series signals)



Class imbalance (within runs):

"healthy": 124448, "faulty": 479 (250 are follow-up breakdowns) → 2.5% of "healthy" signals taken + weight for "faulty"

Classification:

Assumption from trend data analysis: Precursor occurs only in the previous pulse

- Classification problem: calculate probability of a **breakdown in next pulse → used label**

Computation time:

- a. Computiation time on HTCondor with GPU: 1min, locally with only CPU: 10 min
- b. Computiation time on HTCondor with GPU: 1h, locally with only CPU: 24h

2.f. Explainable AI

- Often ML models are treated as black boxes.
- To build trust, the reason behind prediction is often more important than prediction:
  - E.g. medical diagnosis, predictive maintenance, …
  - Trust in model vs. trust in prediction

Additive Feature Attribution Methods:

- For linear models, the feature importance simply the weight
- Define a local linear model $g(z')$ that approximates the output $f(x)$.

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z_i$$

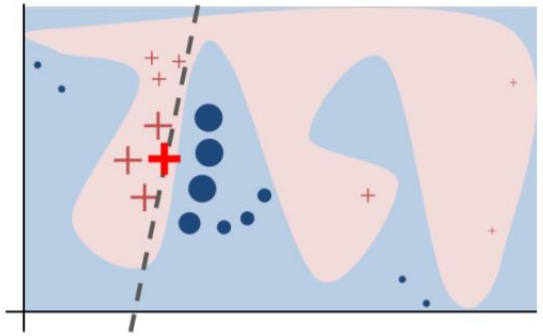


Figure 3: Toy example to present intuition for LIME. The black-box model's complex decision function f (unknown to LIME) is represented by the blue/pink background, which cannot be approximated well by a linear model. The bold red cross is the instance being explained. LIME samples instances, gets predictions using f, and weighs them by the proximity to the instance being explained (represented here by size). The dashed line is the learned explanation that is locally (but not globally) faithful.

Definitions:
x… original representation of instance
x'… interpretable representation (=humanly understandable feature) i.e. biniary vector which states presence or absence of e.g. a word
z'… uniformly sampled instance of x'
$g(x)$… explanation model
$f(x)$… original model
ϕ … effect of feature

Marco Tulio Ribeiro et al. "Why should i trust you?: Explaining the predictions of any classifier". 2016

2.f. SHAP

How would you determine the importance of features:

- Remove feature from data set – if accuracy changes: feature is important
- Problem: **Order** of features highly influences accuracy
  - **Graph theory:** How to fairly allocate credit? – Game theory by Lioyd **Shap**ley (Nobel Pize in 2012)

For Additive Feature Attribution Methods there is a single unique solution if the following properties are fulfilled:

- Local accuracy: $f(x) = g(x')$ Sum of feature importance has to equal the accuracy difference
- Missingness: $x_i' = 0 \rightarrow \phi_i = 0$ If a model is good without a feature, this feature has no importance
- Consistency: If model changes after adding feature, this features is important

Shap Values: $$\phi_i(f, x) = \sum_{z' \subseteq x'} \frac{|z'|!(M-|z'|-1)!}{M!} [f_x(z') - f_x(z' \backslash \mathrm{i})]$$

- Applicable to all Additive Feature Attribution Methods

Definitions:
$f_x(z') = f(h_x(z'))$
$x = h_x(x'))$
$z' \backslash \mathrm{i}$ … $z_i' = 0$
M … Number of simpified input features

Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." NeurIPS. 2017.

3. Results trend data analysis

a. Exploration
b. Modeling
c. Explainable AI
d. Experimental validation

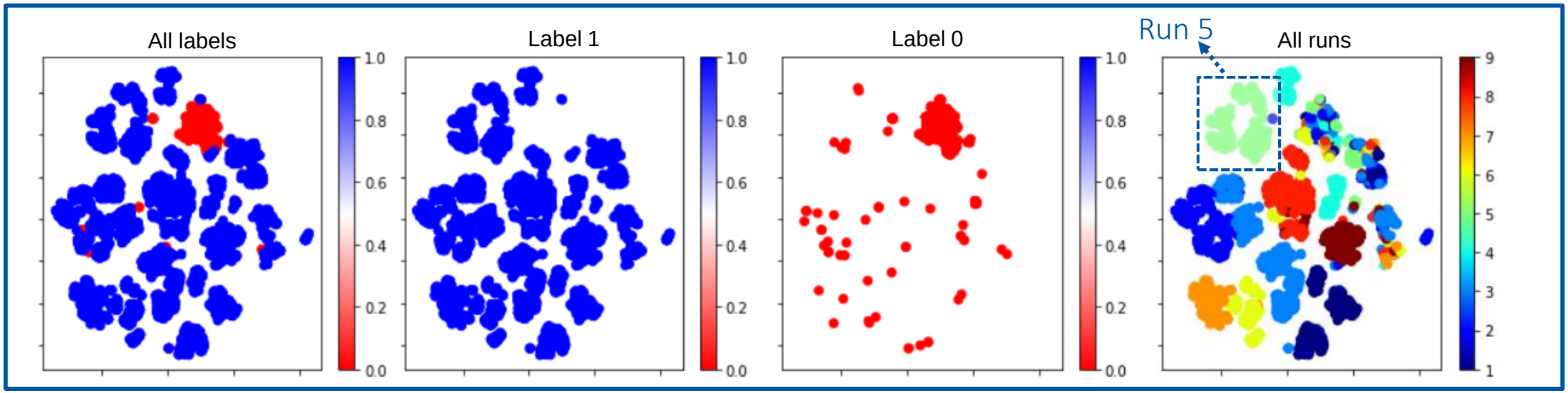
	Bunker_WG_Temp	Klystron_Flange_Temp	Load_Temp	PC_Left_Cavity_Temp	PC_Right_Cavity_Temp
0	24.134285	24.116962	27.955711	30.285633	30.187632
1	24.082500	24.145723	27.876066	30.277182	30.208776
2	23.241980	24.400513	27.502401	30.253901	30.191576
3	18.783634	19.573864	27.900787	30.100540	30.056396

3.a. Exploration

Goal: General method to explore multidimensional dataset.

1 … healthy
0 … faulty

2D t-SNE of XBOX 2 Trend Data (during stable operation)

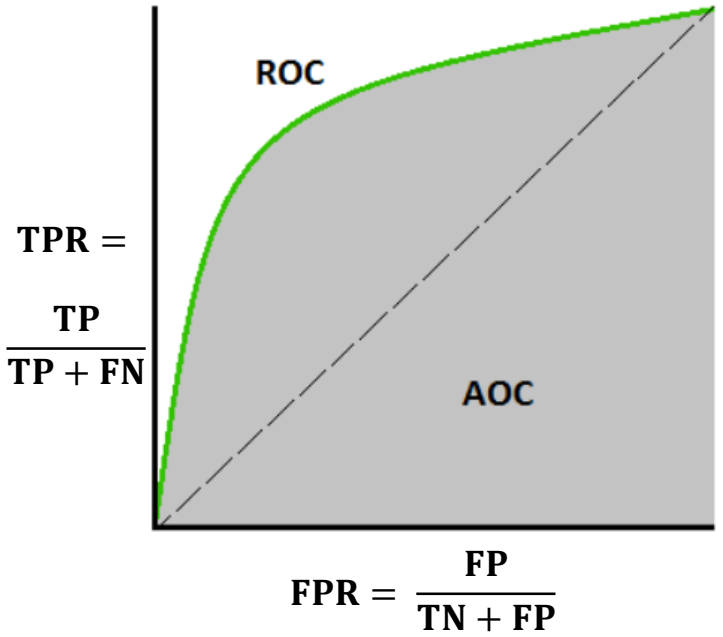


3.b. Modeling

Used Performance Measure:

AUC ROC Score: **A**rea **U**nder **C**urve of **R**eceiver **O**perating **C**haracteristic curve

		AUC – ROC Score				
	Labels	t-CNN	FCN	FCN2d	ResNet	Incept.
a.) trend data + features	All breakdowns	91.2%	94.0%	**94.2%**	93.5%	94.0%
	Follow-up breakdowns	96.3%	**99.9%**	99.9%	99.8%	99.9%
	Primary breakdowns	74.8%	80.4%	81.1%	78.7%	**80.9%**



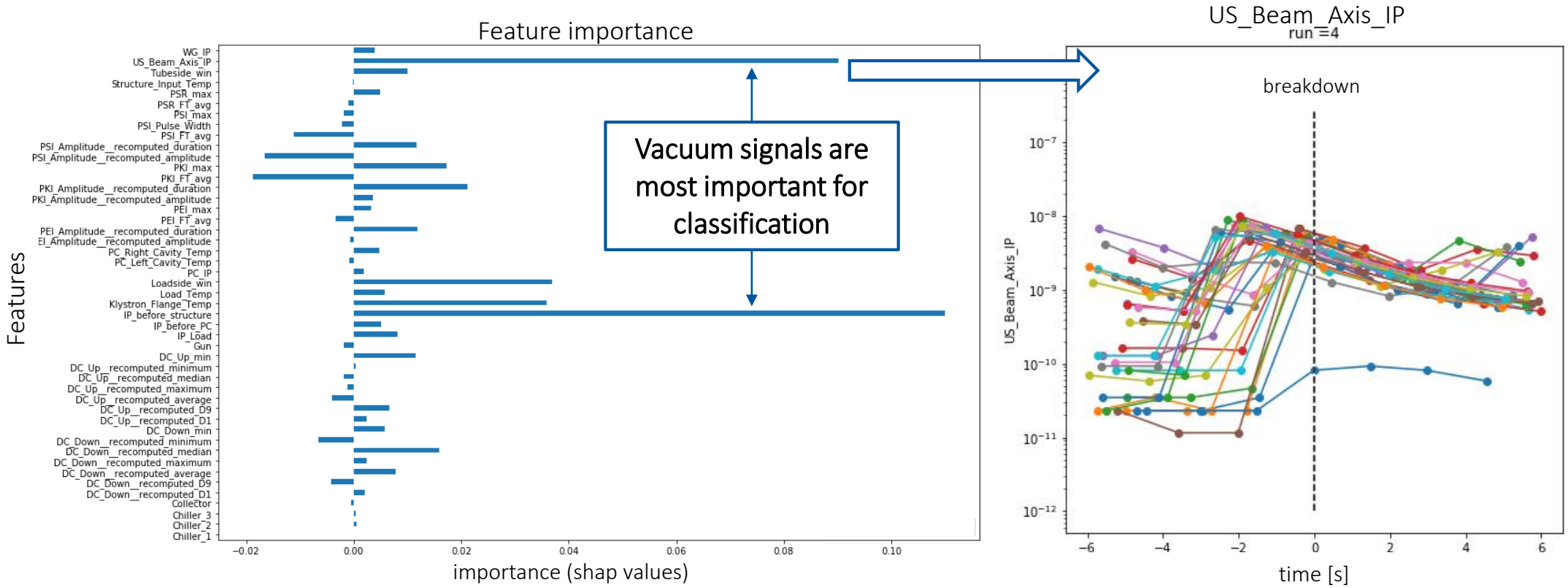
Why do we reach such high AUC ROC score?

→ **Explainable AI**

3.c. Explanation

Why do we reach such high AUC ROC score?

Spike in pressure several seconds before breakdown, **not after** as generally assumed → precursor



3.d. Experimental validation

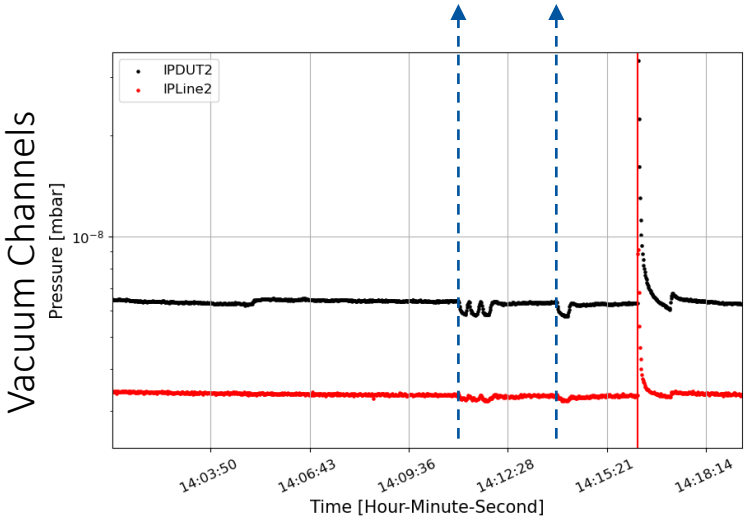
Idea:

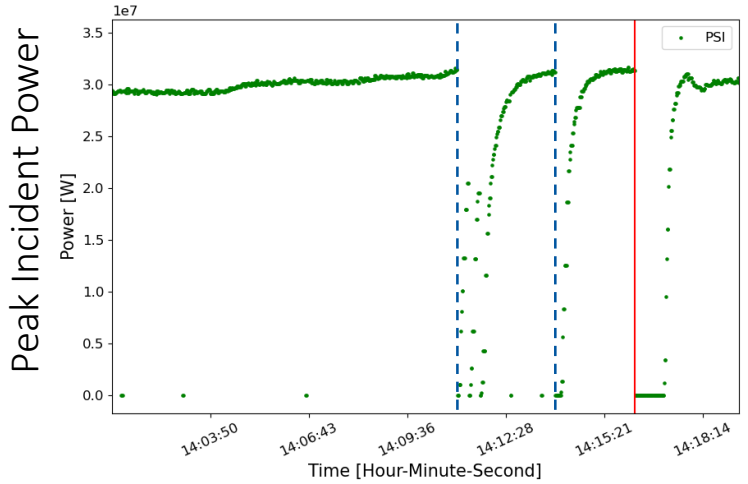
- Turn off RF power if vacuum signal exceeds threshold
- Surface deformities leading local field enhancement in RF cavities remain
- Breakdown should directly occur after turing on power again.

Details:

- The RF was pulsed at 50Hz, the vacuum levels are checked at 600Hz
- 3h of measurement, 15 breakdowns occurred in total (5 groups with 4, 1, 4, 2 and 4 breakdowns).
- In **two** breakdown groups, the vacuum precursors were **clearly observed**.
- Other groups could have occurred due to change of settings during the test.

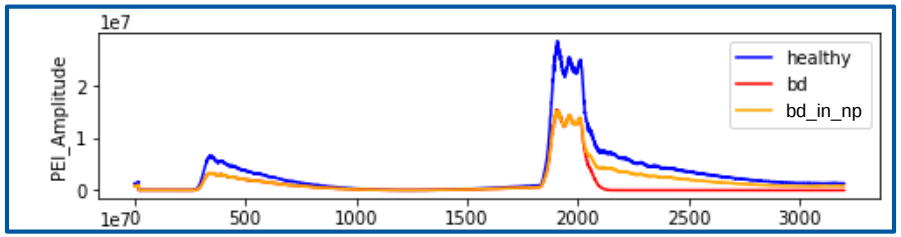
Vacuum interlocks, we then remove the threshold, ramp back up and a breakdown occurs.





4. Results event data analysis

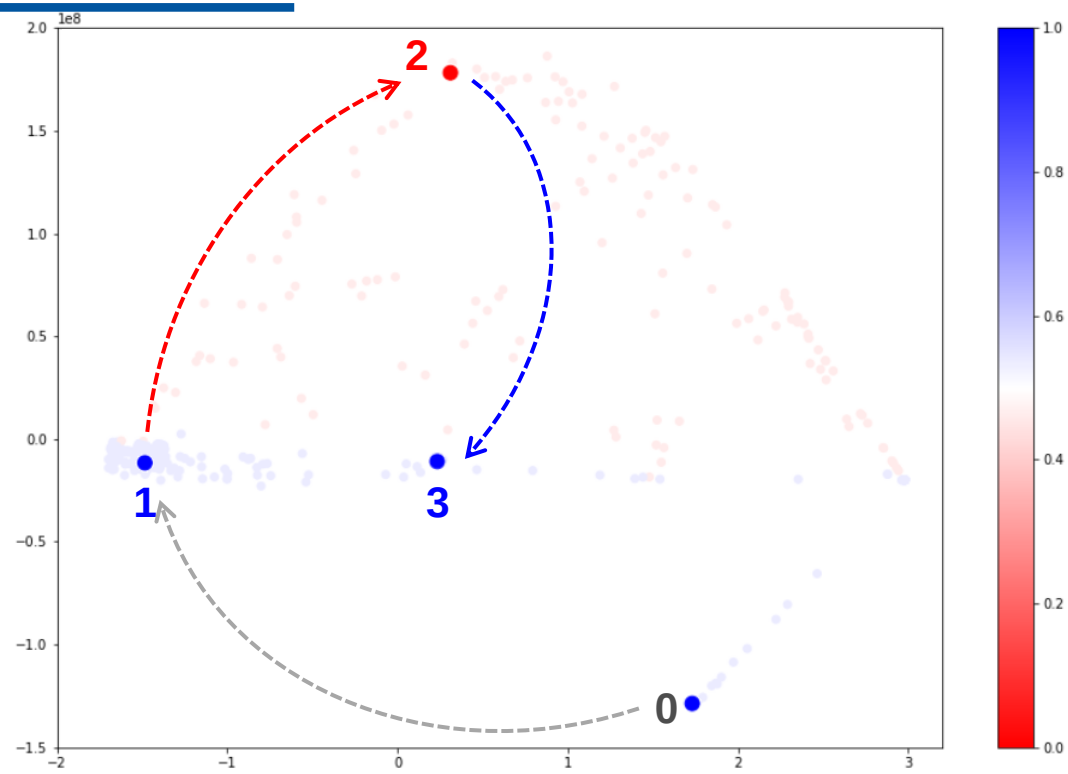
a. Exploration
b. Modeling
c. Explainable AI



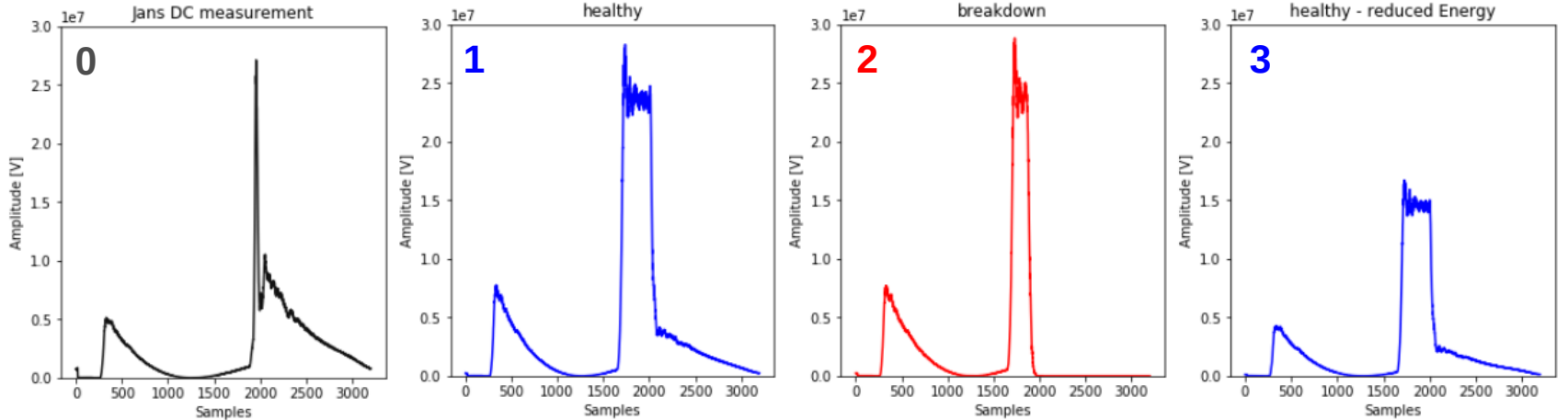
4.a. Exploration

Data:

- Event data signal (timeseries):
- Date: 20-26.01.2018
- Axes represent arbitrary



	label_CLIC	timestamp
0	1	2018-01-26 14:52:30.443347
1	1	2018-01-26 14:59:50.549269
2	0	2018-01-26 15:26:23.099318
3	1	2018-01-26 15:26:28.481433

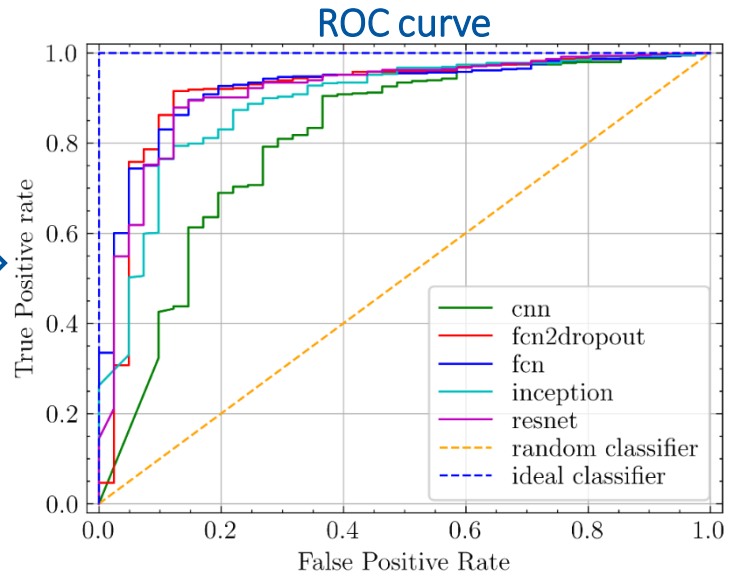


4.b. Modeling

Used Performance Measure:

AUC ROC Score: **A**rea **U**nder **C**urve of **R**eceiver **O**perating **C**haracteristic curve

		AUC – ROC Score				
	Labels	**t-CNN**	**FCN**	**FCN2d**	**ResNet**	**Incept.**
b.) event data	All breakdowns	84.7%	**89.3%**	85.9%	85.9%	86.8%
	Follow-up breakdowns	93.2%	99.3%	98.6	**99.3%**	98.7%
	Primary breakdowns	64.0%	**73.9%**	71.2	71.4%	70.8%



Why do we reach such high AUC ROC score on predicting follow-up breakdowns?

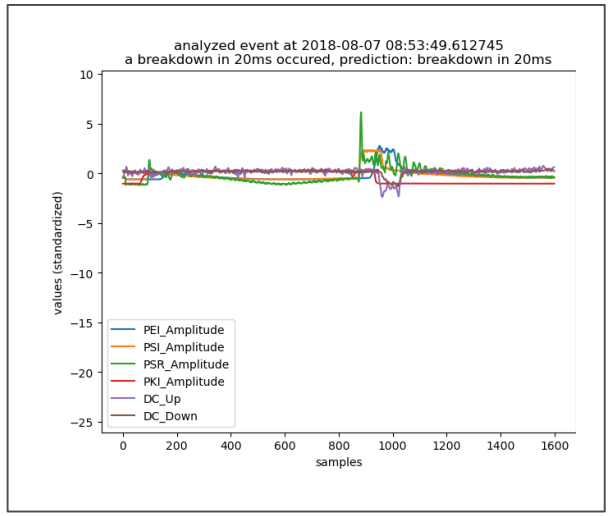
→ **Explainable AI**

4.c. Explanation

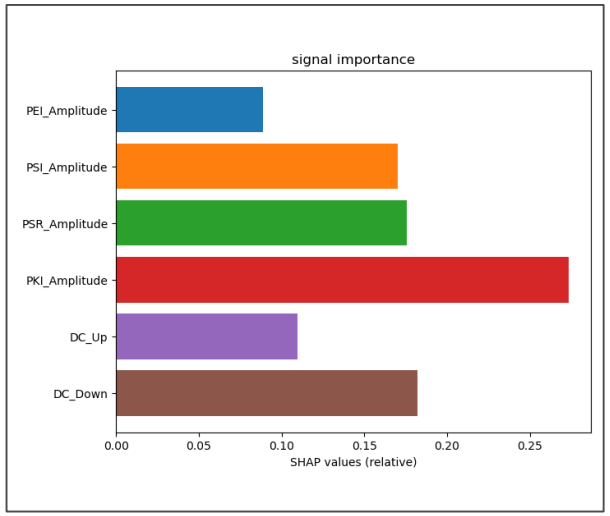
Why do we reach such high AUC ROC score for follow-up breakdowns?:

Explainable AI dashboard **shows reason** for model decisions and allows reverse engineering of physical properties.

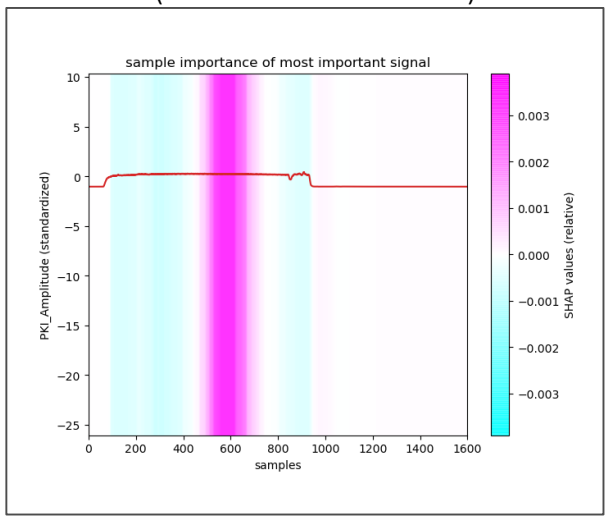
Signals used for current prediction



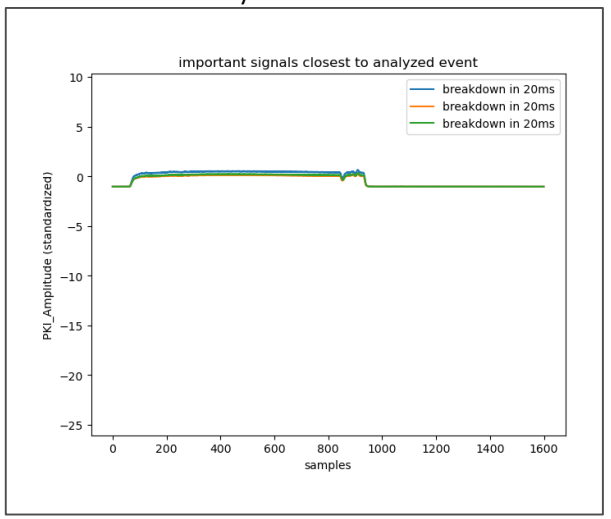
Signal importance (relative SHAP values)



Sample importance (relative SHAP values)



Signals which are taken as reference by network



By eye: The middle region seems to be not important, as there is no change in the signal

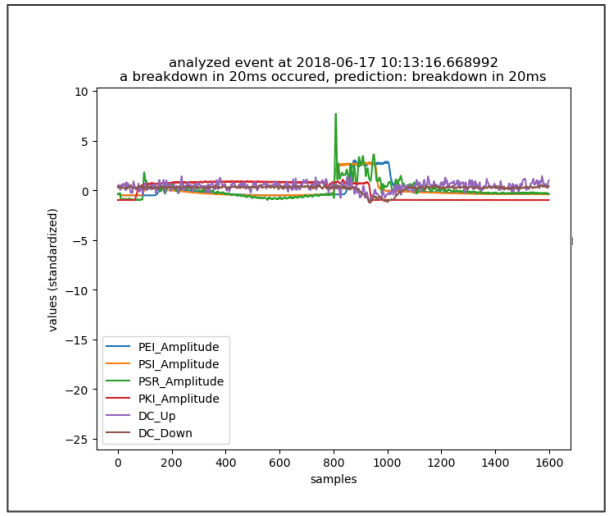
Expert explanation: The midle region of the PSR and PKI Amplitude is an important factor for the pulse compressor tuning. The high sample importance in this region might indicate a **wrong pulse compressor tuning during this time period.**

4.c. Explanation

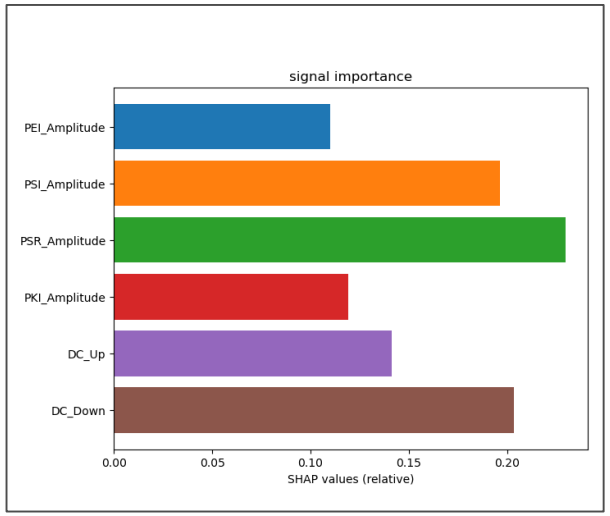
Why do we reach such high AUC ROC score for follow-up breakdowns?:

Explainable AI dashboard **shows reason** for model decisions and allows reverse engineering of physical properties.

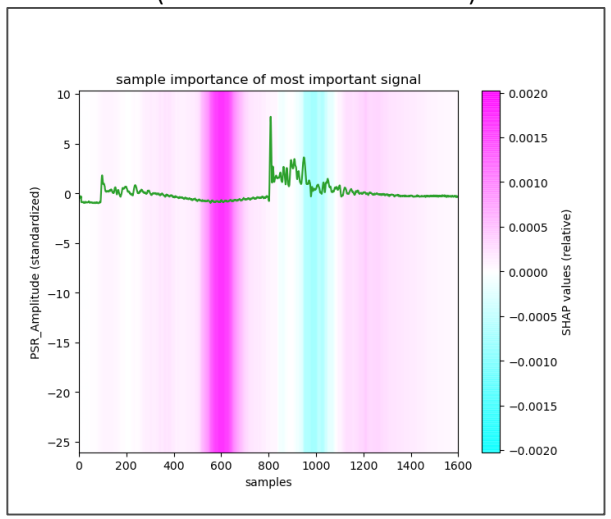
Signals used for current prediction



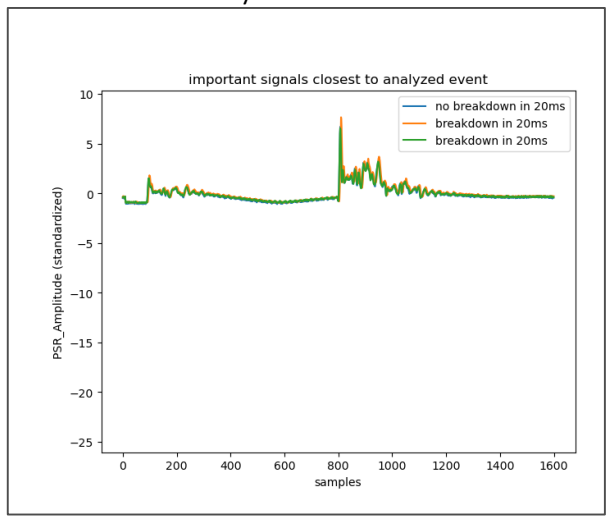
Signal importance (relative SHAP values)



Sample importance (relative SHAP values)



Signals which are taken as reference by network



By eye: The middle region seems to be not important, as there is no change in the signal

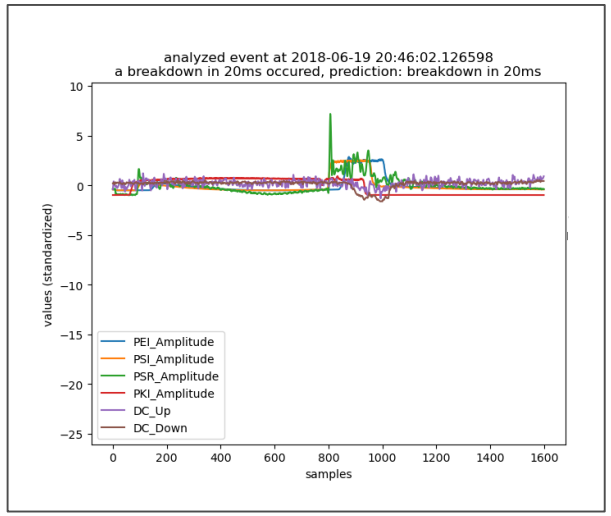
Expert explanation: The midle region of the PSR and PKI Amplitude is an important factor for the pulse compressor tuning. The high sample importance in this region might indicate a **wrong pulse compressor tuning during this time period.**

4.c. Explanation

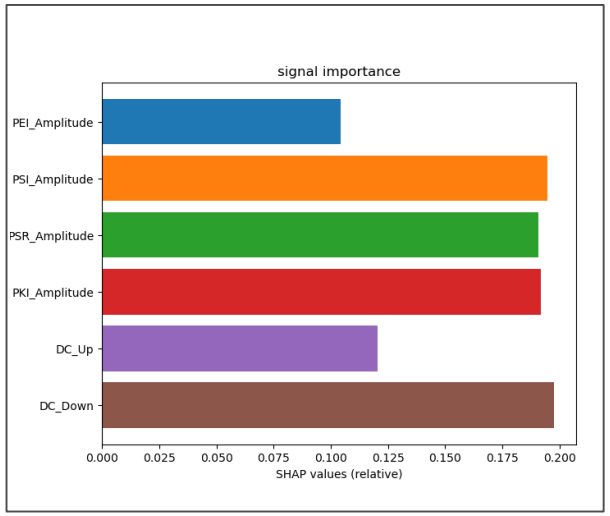
Why do we reach such high AUC ROC score for follow-up breakdowns?:

Explainable AI dashboard **shows reason** for model decisions and allows reverse engineering of physical properties.

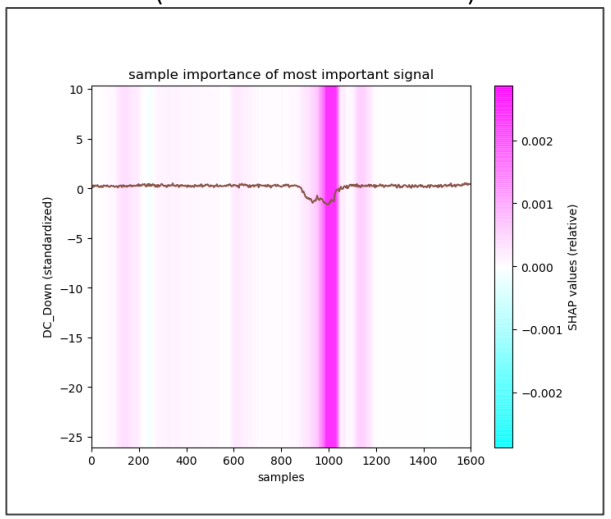
Signals used for current prediction



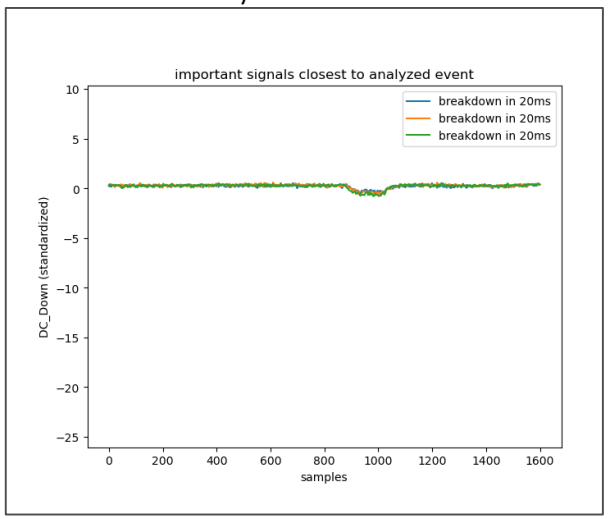
Signal importance (relative SHAP values)



Sample importance (relative SHAP values)



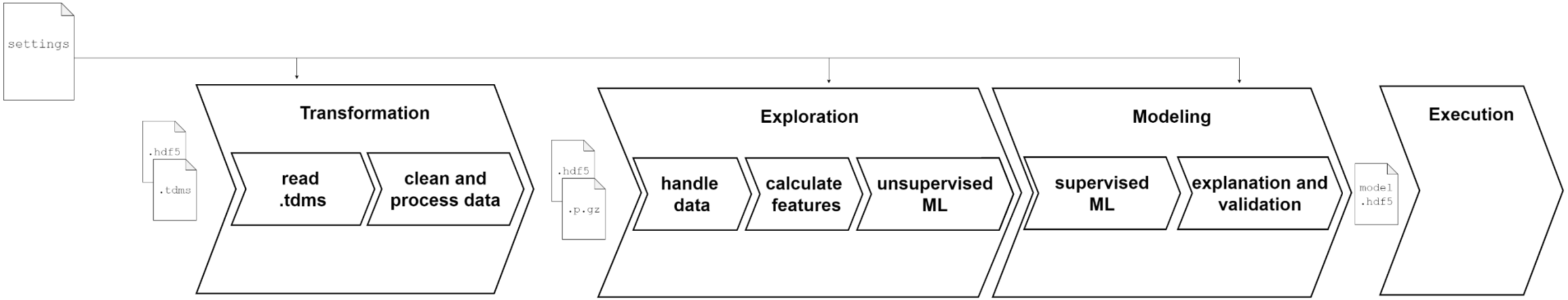
Signals which are taken as reference by network



By eye: Last part of small DC spike seems to be important for breakdown in next pulse

Expert explanation: The signal sporadically changes following breakdowns → **high importance region** often temporally located **at the end** of the compressed RF pulse.

5. ML Framework

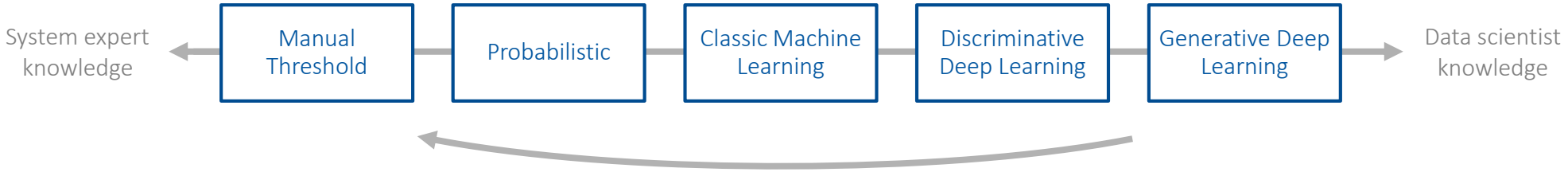


We work on a generic framework (Gitlab CI/CD + code reviews) with the software team, which can be applied to:

- Similar use-cases at CERN
- Similar use-cases from Institutions interested in our work: Consorzio RFX (ITER), Melbox (Melbourne University)
- Similar use-cases from Institutions who work on the same topic: Daresbury Laboratory, DESY

6. Conclusion

1. Show potential
   - A general way for data exploration with **unsupervised ML** has been established
   - Good results for predicting breakdowns in RF cavities with **supervised ML**
2. Build trust
   - The good predicting results could be explained. **Explainable AI**, provided **interface** between ML and experts
     → It was possible to identify **precursors**, which were **not evident** for experts



3. Enable easy application
   - CLIC ML framework for transformation, exploration, modeling is available

7. Outlook

1. Show potential
   - **Further improve** and verify supervised and unsupervised modelling techniques.
   - Under discussion with CLIC team: **Additional sensors** (e.g. vibration)? Increase sampling frequency?
   - Further investigate **physics meaning** of good classification results
   - Explore the possibility to **implement** the used ML **methods** in an operational tool for RF accelerating structure conditioning, to smoothen the conditioning process and avoid as much as possible breakdowns by suitable adjustment of the power in the cavity.
2. Build trust
   - Further improve **explainable AI** techniques, especially suited to show experts what is important
3. Enable easy application
   - ML framework for transformation, exploration, modeling is under development (Gitlab CI/CD + code reviews)

CERN