
The Outliers – anomaly detection challenge

AISSAI 2026 – 16 October 2026

Members:

Charly Lassalle

Yuna Dan

Anass Bellachehab

Tudor Stefan

Stéphane Ilic

Lex Avstreikh (joker)

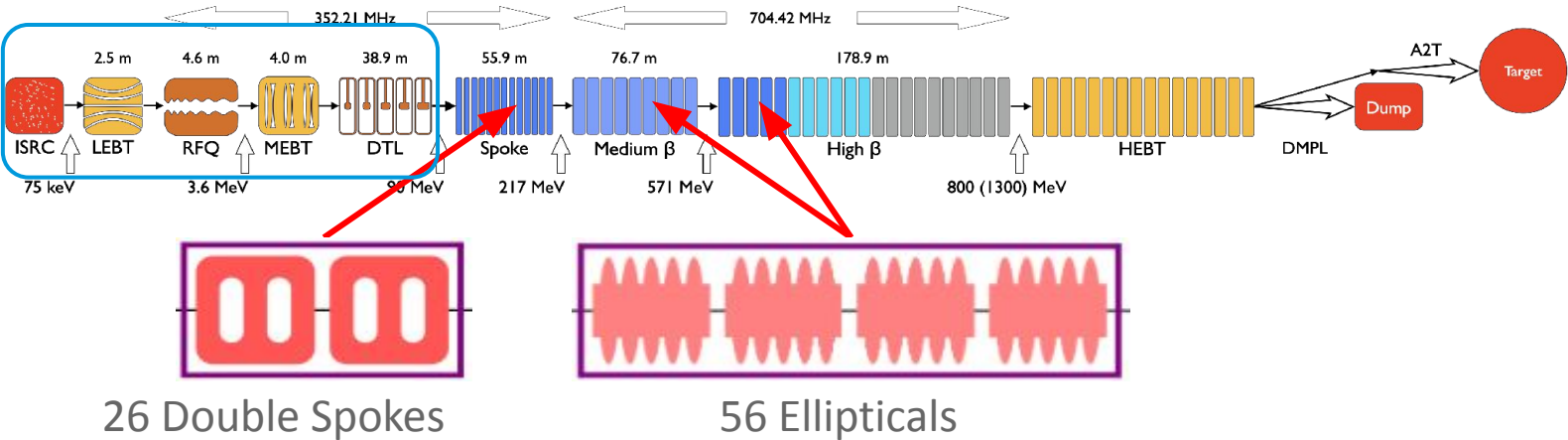
Ishkhan Gorgisyan (case holder)

- 1. Use case**
- 2. Data**
- 3. Methodology**
- 4. Exploratory work**
- 5. Future steps**

Use case – Anomaly Detection on ESS Time-series

European Spallation Source (ESS):

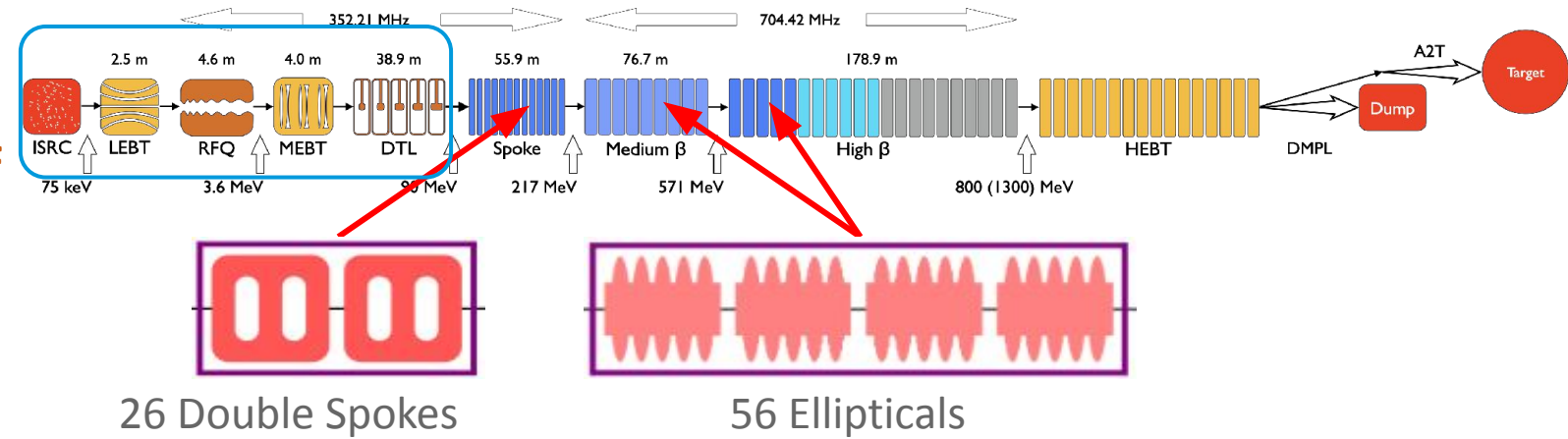
- User facility – beam acceleration with **RF cavities**



Use case – Anomaly Detection on ESS Time-series

European Spallation Source (ESS):

- User facility – beam acceleration with **RF cavities**
- System diagnosis:
 - RF cavity waveforms
 - Scalar readings from entire machine

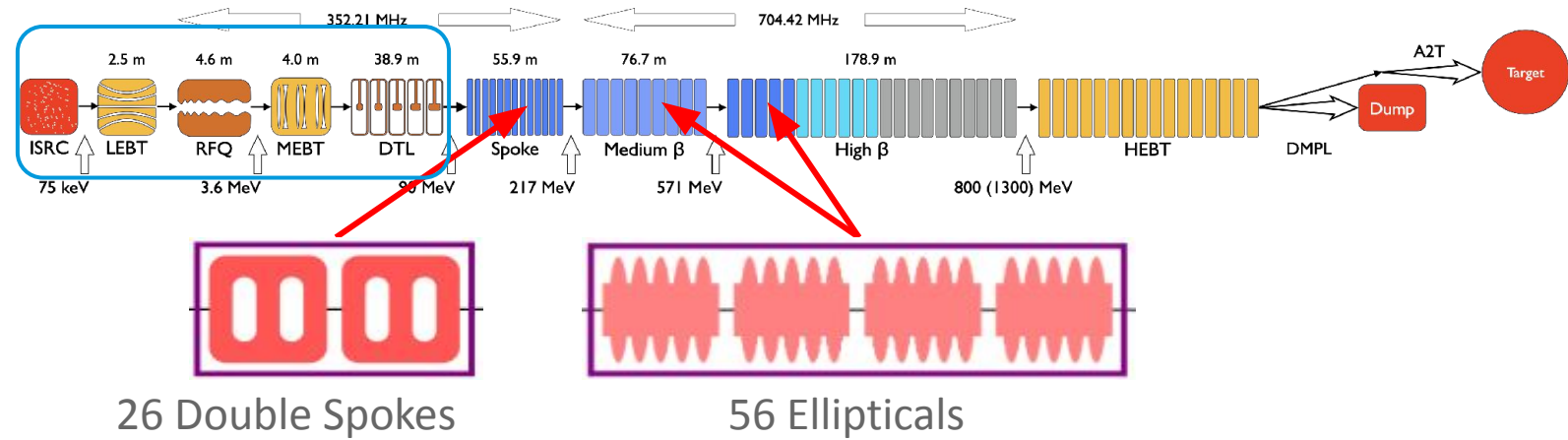


Use case – Anomaly Detection on ESS Time-series

European Spallation Source (ESS):

- User facility – beam acceleration with **RF cavities**

- System diagnosis:
 - RF cavity waveforms
 - Scalar readings from entire machine
- Current state:
 - No model for anomaly detection
 - Post mortem analysis after operation

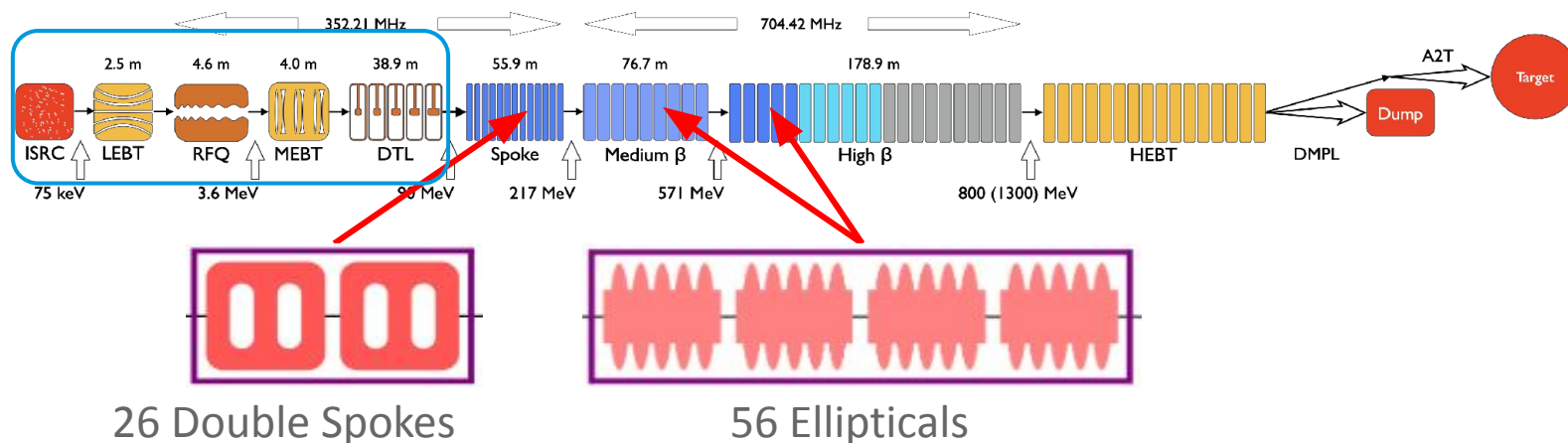


Use case – Anomaly Detection on ESS Time-series

European Spallation Source (ESS):

- User facility – beam acceleration with **RF cavities**

- System diagnosis:
 - RF cavity waveforms
 - Scalar readings from entire machine
- Current state:
 - No model for anomaly detection
 - Post mortem analysis after operation



Challenge:

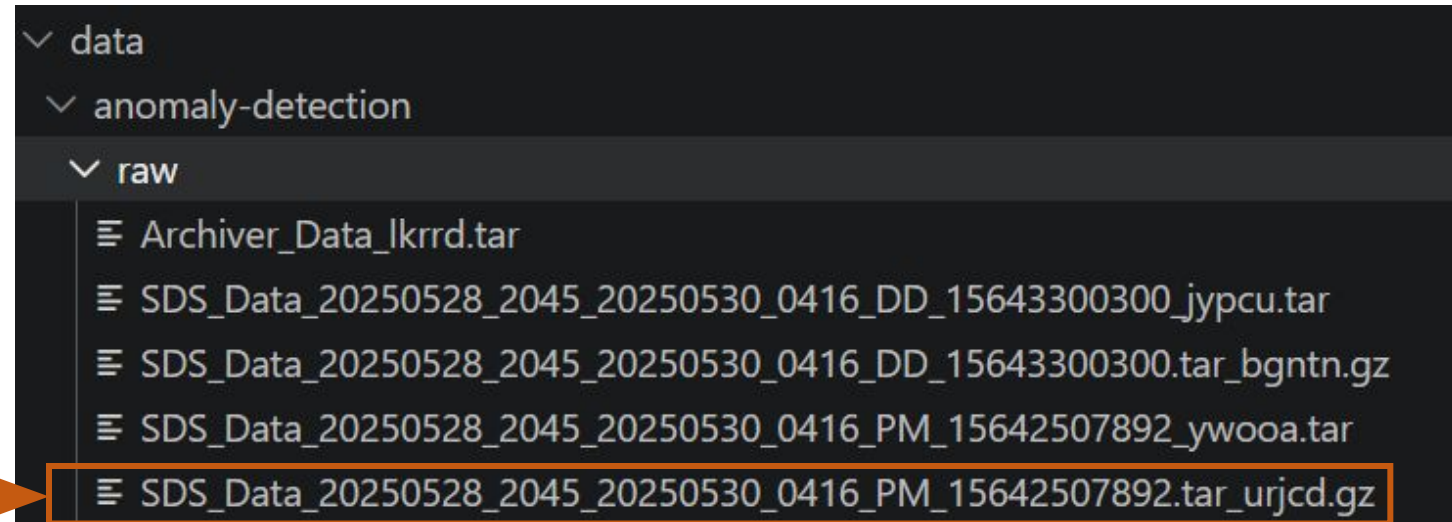
- Anomaly detection
- Fault characterization and classification
- Fault precursor identification and prediction

Data – Initial look for better understanding

Data origin:

- Commissioning campaigns at ESS
 - 2025 March-June
 - 2026 January-July

E.g. campaign on 28 May 2025 04:16 PM

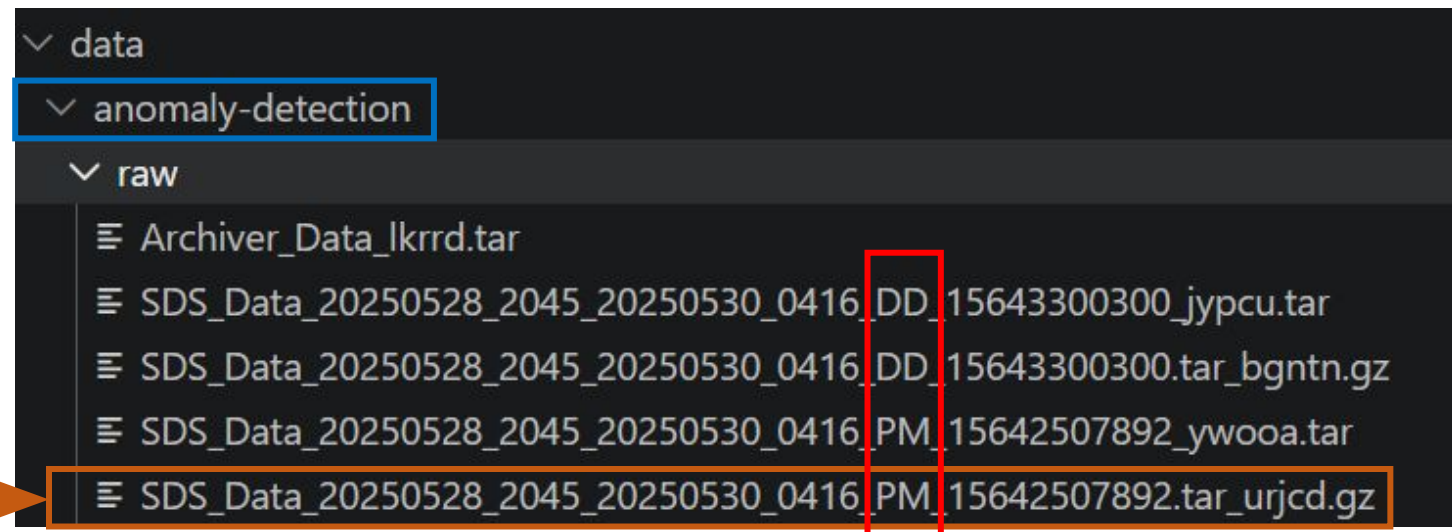


Data – Initial look for better understanding

Data origin:

- Commissioning campaigns at ESS
 - 2025 March-June
 - 2026 January-July

E.g. campaign on 28 May 2025 04:16 PM



Data structure:

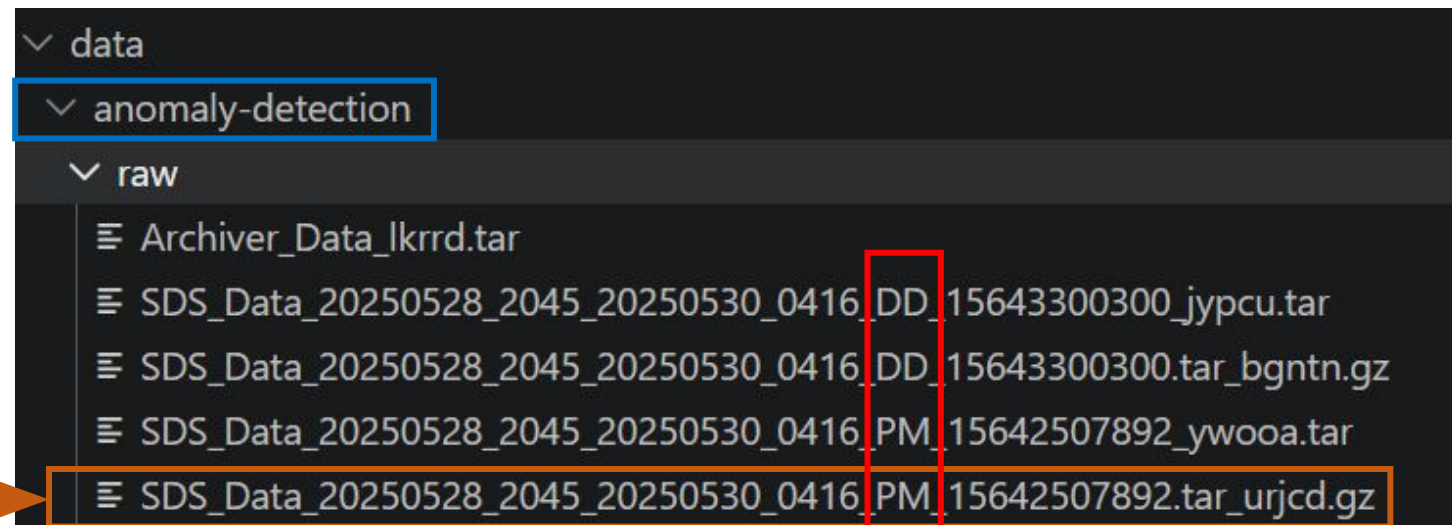
- Scalar data (from entire machine)
- For each campaign – DD and PM data:
 - DD (data on demand)
 - PM (post mortem)

Data – Initial look for better understanding

Data origin:

- Commissioning campaigns at ESS
 - 2025 March-June
 - 2026 January-July

E.g. campaign on 28 May 2025 04:16 PM



Data structure:

- Scalar data (from entire machine)
- For each campaign – DD and PM data:
 - DD (data on demand)
 - PM (post mortem)

Campaign data:

Parquet files

Data origin:

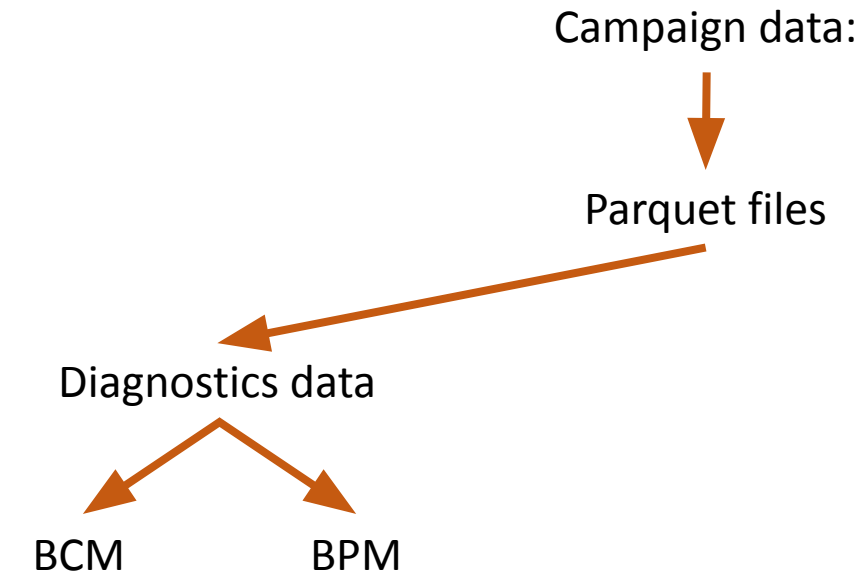
- Commissioning campaigns at ESS
 - 2025 March-June
 - 2026 January-July

E.g. campaign on 28 May 2025 04:16 PM

```
data
├── anomaly-detection
└── raw
    ├── Archiver_Data_lkrrd.tar
    ├── SDS_Data_20250528_2045_20250530_0416_DD_15643300300_jypcu.tar
    ├── SDS_Data_20250528_2045_20250530_0416_DD_15643300300.tar_bgntn.gz
    ├── SDS_Data_20250528_2045_20250530_0416_PM_15642507892_ywooa.tar
    └── SDS_Data_20250528_2045_20250530_0416_PM_15642507892.tar_urjcd.gz
```

Data structure:

- Scalar data (from entire machine)
- For each campaign – DD and PM data:
 - DD (data on demand)
 - PM (post mortem)

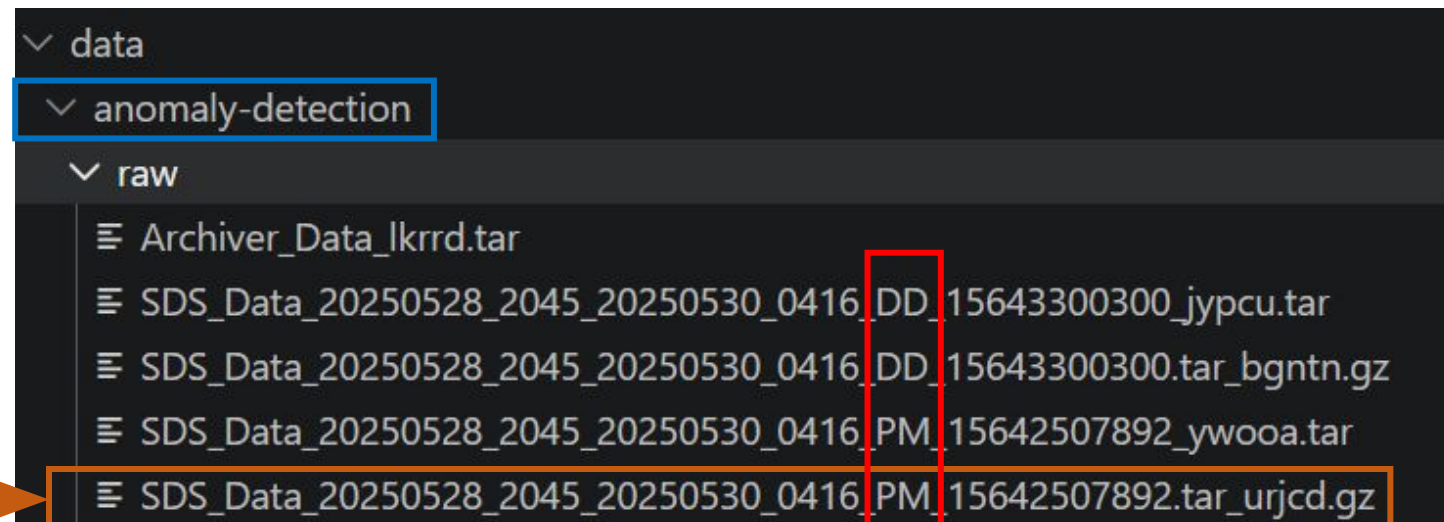


Data – Initial look for better understanding

Data origin:

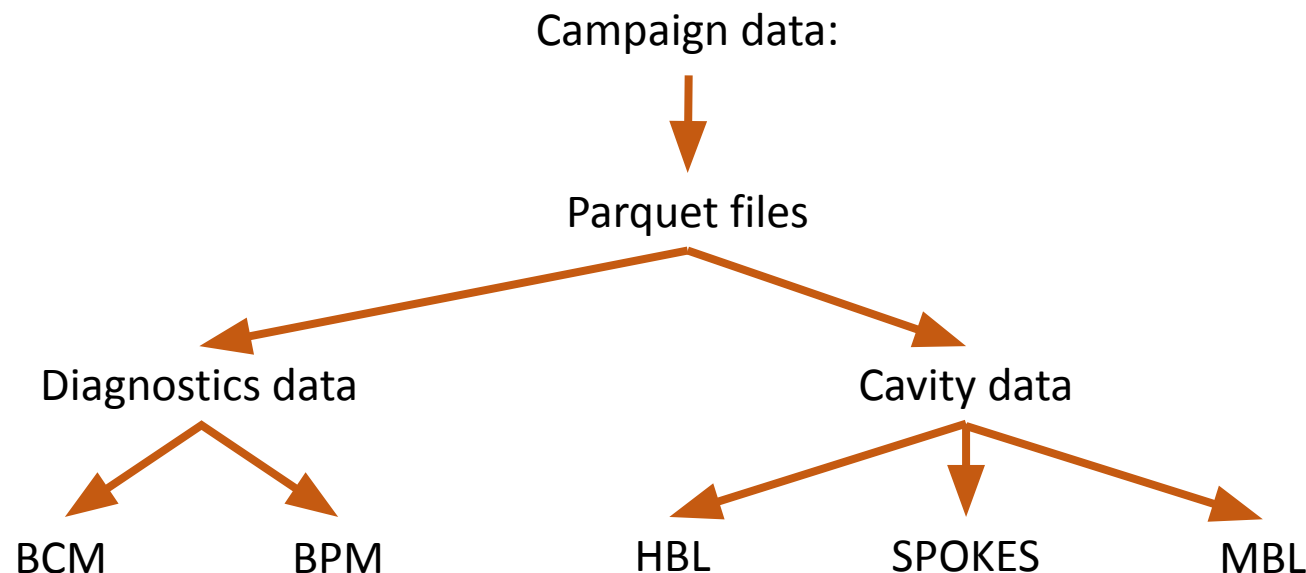
- Commissioning campaigns at ESS
 - 2025 March-June
 - 2026 January-July

E.g. campaign on 28 May 2025 04:16 PM



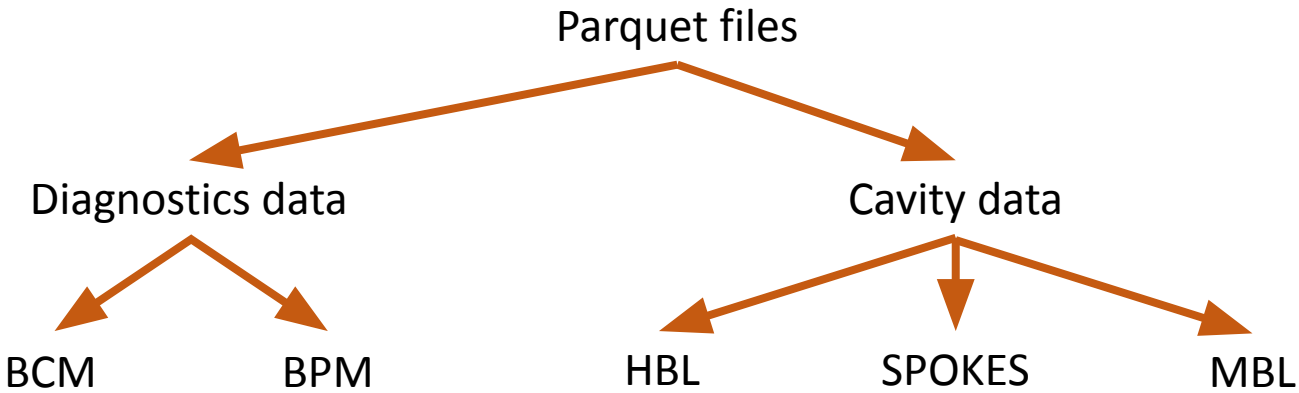
Data structure:

- Scalar data (from entire machine)
- For each campaign – DD and PM data:
 - DD (data on demand)
 - PM (post mortem)



Parquet file content:

- 6 waveforms / cavity – RF waveform data
- **PV** - cavity identifier
- beam_state/**beam_present** - beam present/not



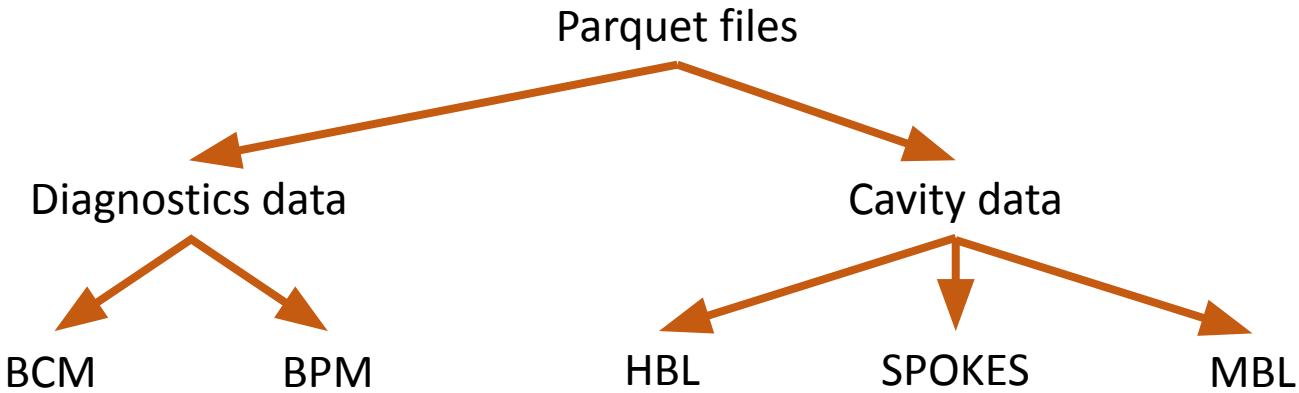
cycle_id	pv_base	trig_code	trig_user_tag	beam_mode	beam_state	beam_present
16096681837	MBL-090RFC:RFS-DIG-101:Dwn0-CmpSDS-PMData	14	0	Probe	BEAM_INHIBIT	False
16096681838	MBL-090RFC:RFS-DIG-101:Dwn0-CmpSDS-PMData	14	0	Probe	BEAM_INHIBIT	False
16096681839	MBL-090RFC:RFS-DIG-101:Dwn0-CmpSDS-PMData	14	0	Probe	BEAM_INHIBIT	False

Parquet file content:

- 6 waveforms / cavity – RF waveform data
- **PV** - cavity identifier
- beam_state/**beam_present** - beam present/not

Cleaning data:

1. Divide into DD and PM
 - DD assumed nominal - training
 - PM may contain anomalies – testing
2. Keep only beam_present = true
3. Group all data by cavity name



cycle_id	pv_base	trig_code	trig_user_tag	beam_mode	beam_state	beam_present
16096681837	MBL-090RFC:RFS-DIG-101:Dwn0-CmpSDS-PMData	14	0	Probe	BEAM_INHIBIT	False
16096681838	MBL-090RFC:RFS-DIG-101:Dwn0-CmpSDS-PMData	14	0	Probe	BEAM_INHIBIT	False
16096681839	MBL-090RFC:RFS-DIG-101:Dwn0-CmpSDS-PMData	14	0	Probe	BEAM_INHIBIT	False

Step 1 - data selection:

- DD data considered for training, PM for model testing
- Grouping data for each RF cavity

Step 1 - data selection:

- DD data considered for training, PM for model testing
- Grouping data for each RF cavity

Step 2 - autoencoder

- Consider only one RF cavity
- Split DD data:
 - 80% training
 - 10% validation
 - 10% testing
- Testing on PM data for same cavity

Step 1 - data selection:

- DD data considered for training, PM for model testing
- Grouping data for each RF cavity

Step 3 - characterization and labelling

- Threshold anomaly score for fault characterization
- Labelling anomalies based on clusters

Step 2 - autoencoder

- Consider only one RF cavity
- Split DD data:
 - 80% training
 - 10% validation
 - 10% testing
- Testing on PM data for same cavity

Step 1 - data selection:

- DD data considered for training, PM for model testing
- Grouping data for each RF cavity

Step 3 - characterization and labelling

- Threshold anomaly score for fault characterization
- Labelling anomalies based on clusters

Step 2 - autoencoder

- Consider only one RF cavity
- Split DD data:
 - 80% training
 - 10% validation
 - 10% testing
- Testing on PM data for same cavity

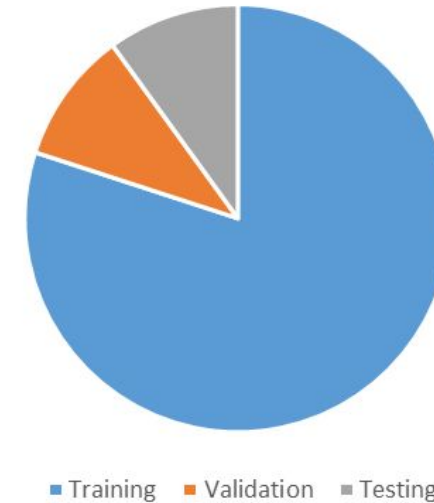
Step 4 - generalization and prediction

- Repetition for all cavities
- Add extra information with scalars data
- Searching -30 minutes before each anomaly for signs of impending fault - precursors

Step 1 - data selection:

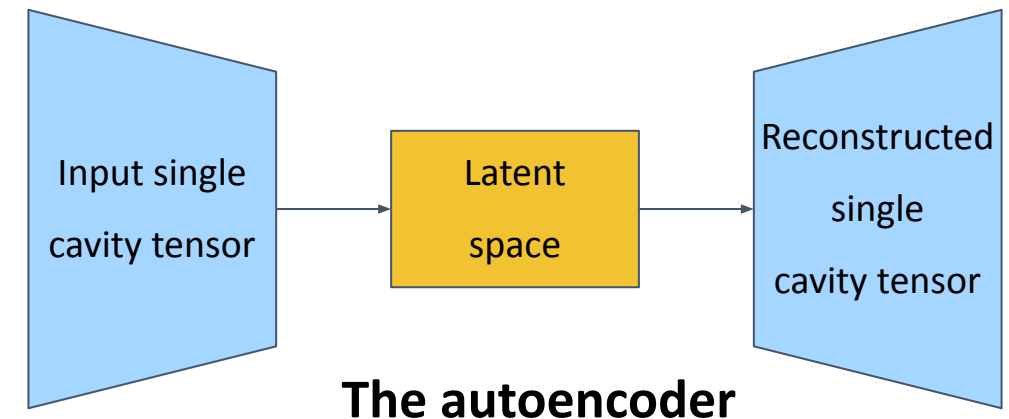
- Cavity MBL-090RFC_RFS-EVR-101 chosen as example
- Parquet files:
 - DD: 80% training, 10% validation, 10% test
 - PM: fully used for testing

Proportion of DD data



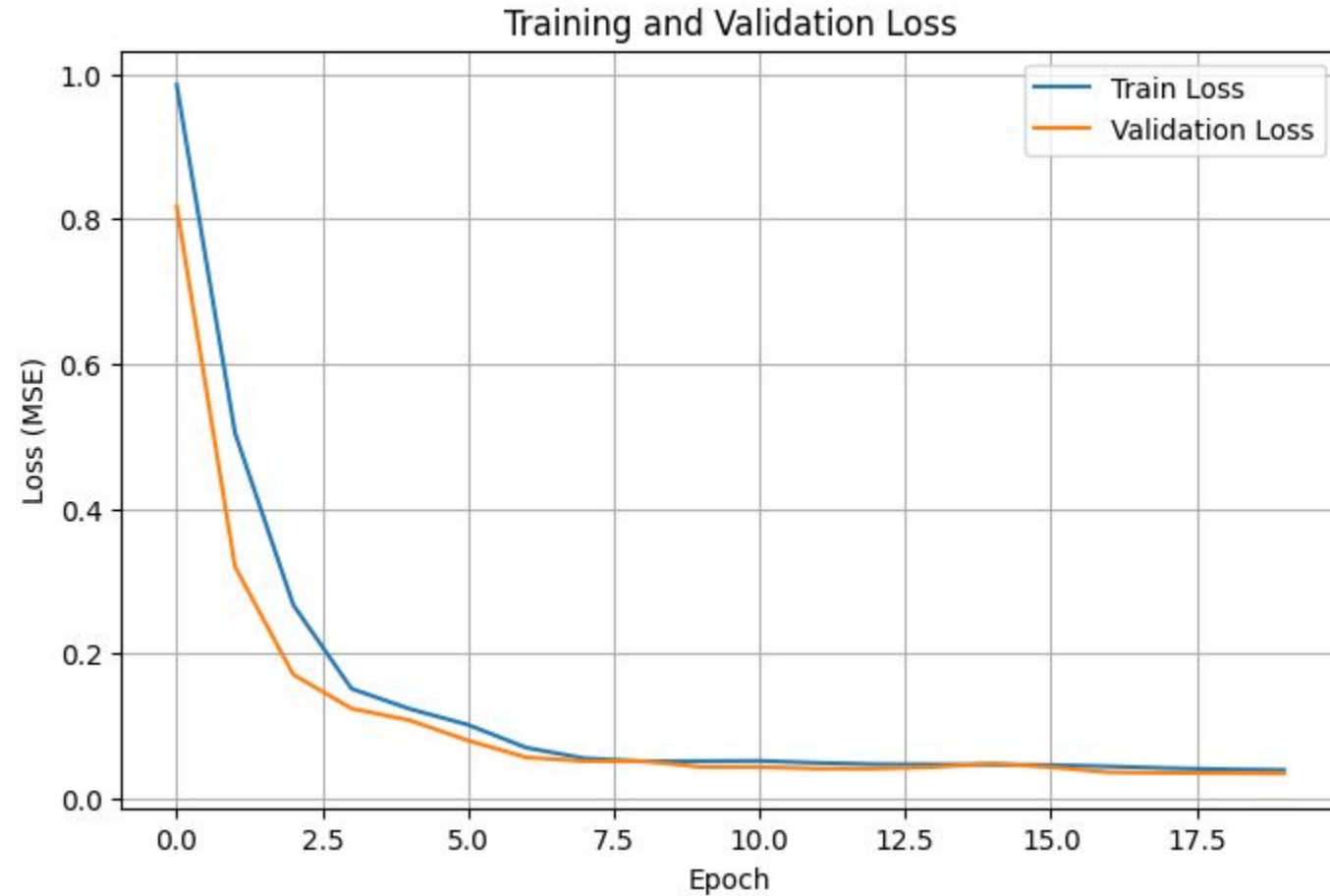
Step 2 - autoencoder:

- Parquet files converted to tensors
- Tensor contains 6 waveforms, all data points/waveform and cycle ID
- Autoencoder layer type: Conv1D
- Ground truth depending on cycle ID



Hyperparameters used:

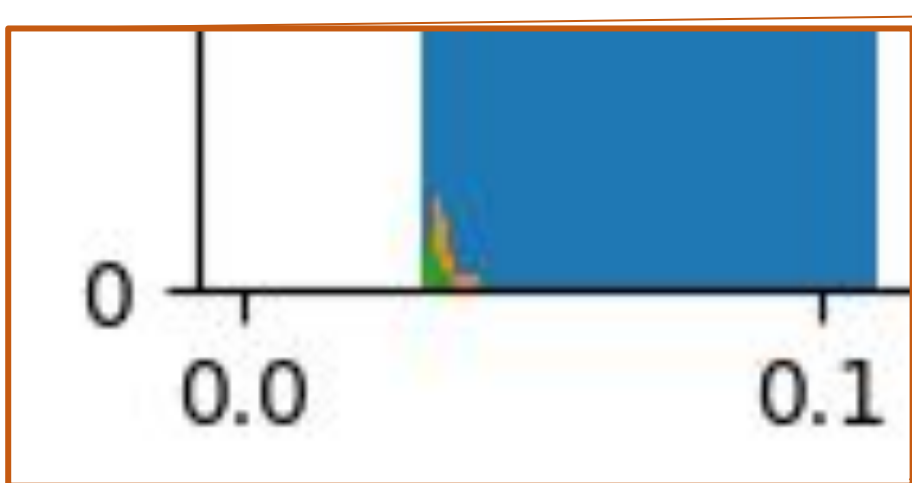
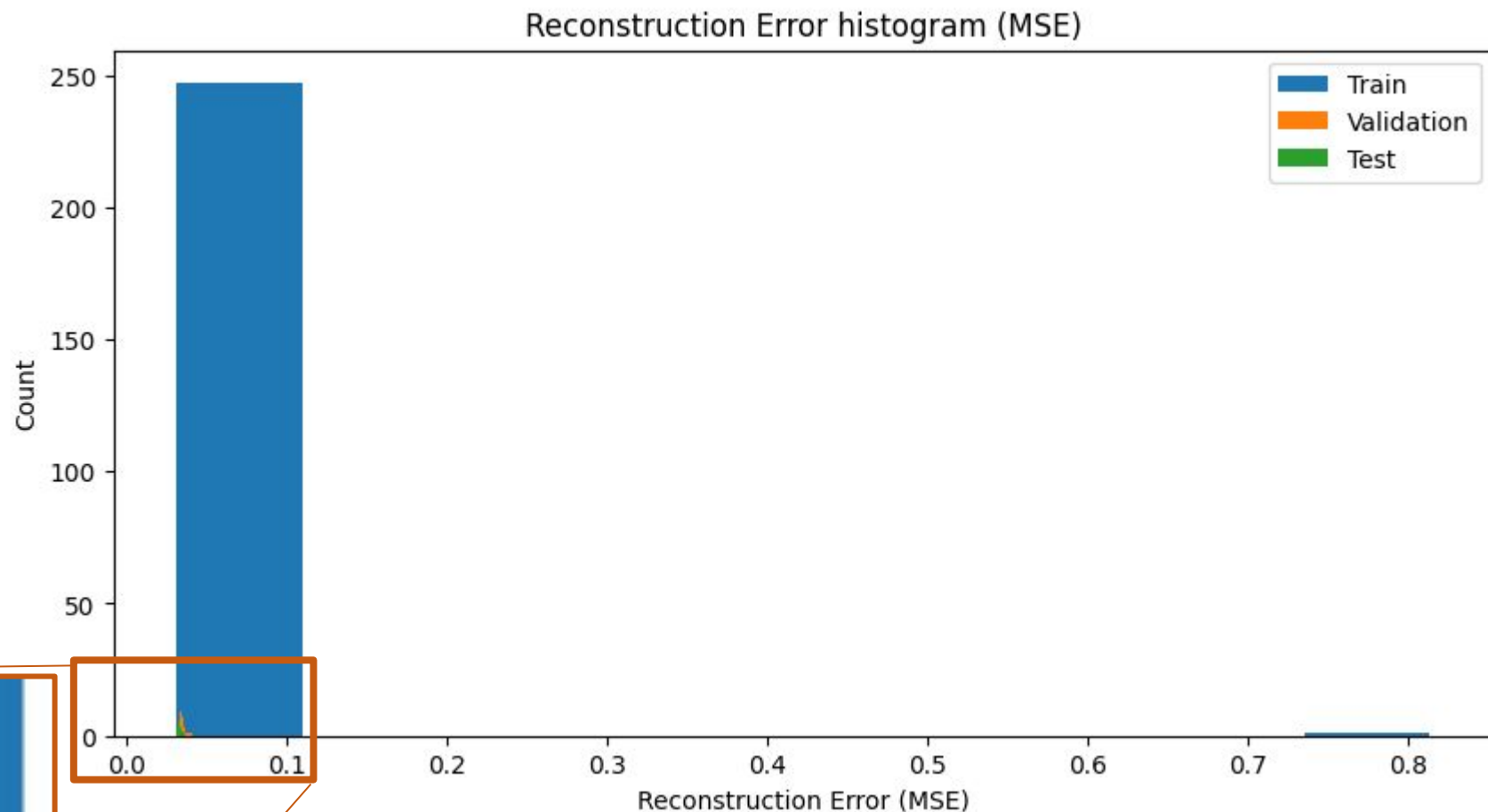
- batch_size = 4
- Optimizer = Adam
- Learning rate = 10^{-3}
- Loss = mean square error
- Max epochs = 20



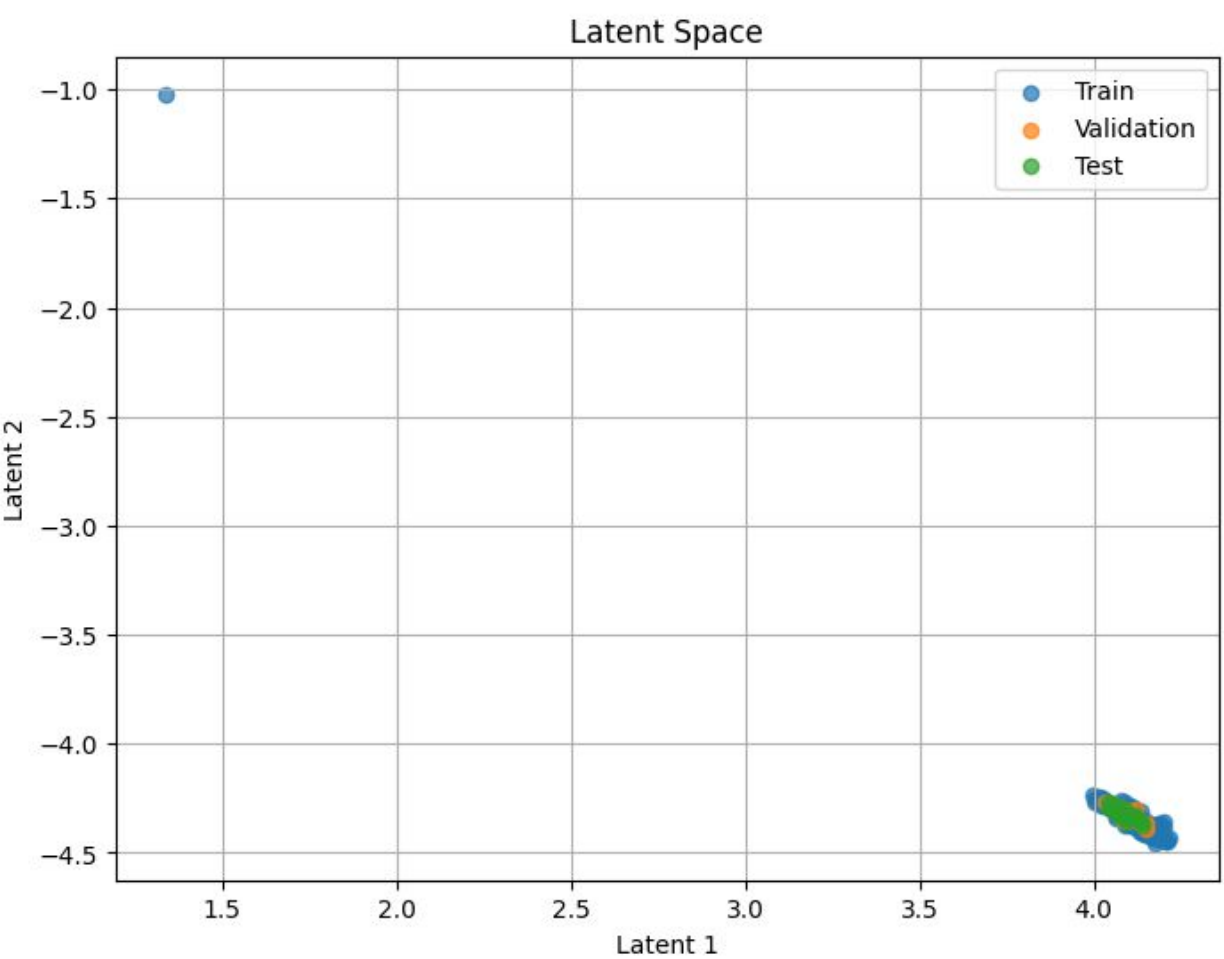
Exploratory work – Autoencoder training

Reconstruction error:

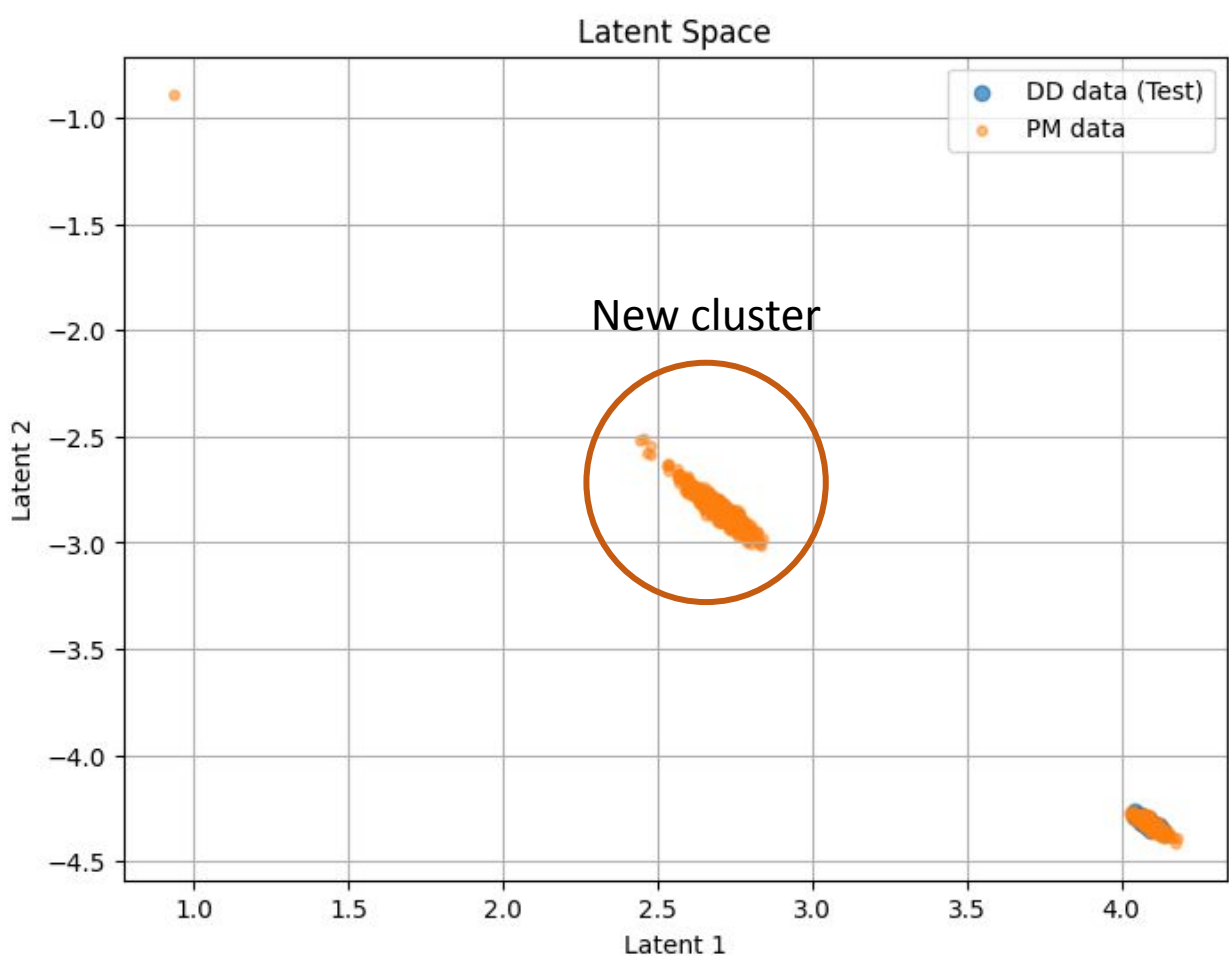
- ~250 point very low MSE
- 1 point with MSE ~0.8
- Threshold:
 - manually selected
 - set at 0.15 MSE



Exploratory work – Latent space visualization



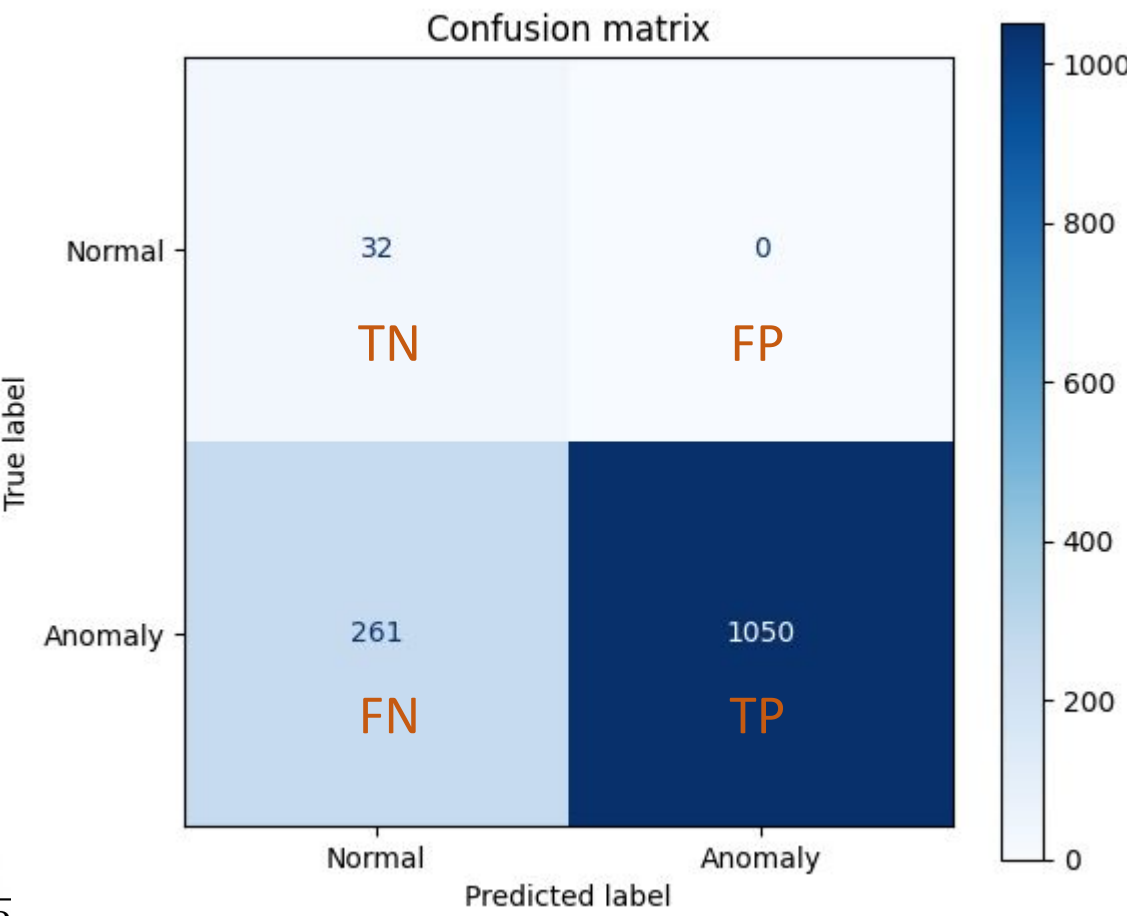
DD data splitted: training, validation, testing



Test data: DD and PM

Confusion matrix results:

- Accuracy : 0.8057
- Precision : 1.0000
- Recall : 0.8009
- Specificity : 1.0000
- F1-score : 0.8895



Precision: $\frac{TP}{TP + FP}$

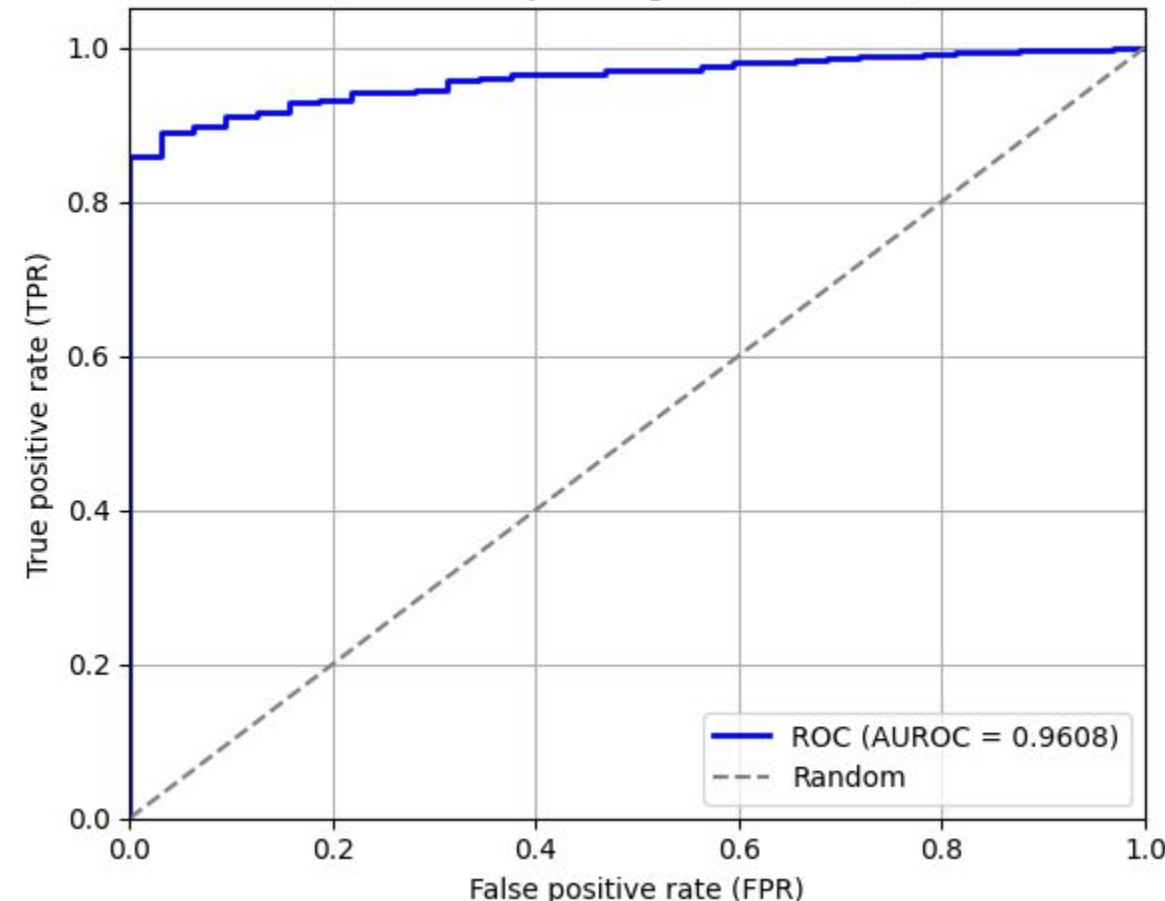
Specificity: $\frac{TN}{TN + FP}$

Recall: $\frac{TP}{TP + FN}$

F1-score: $\frac{2TP}{2TP + FP + FN}$

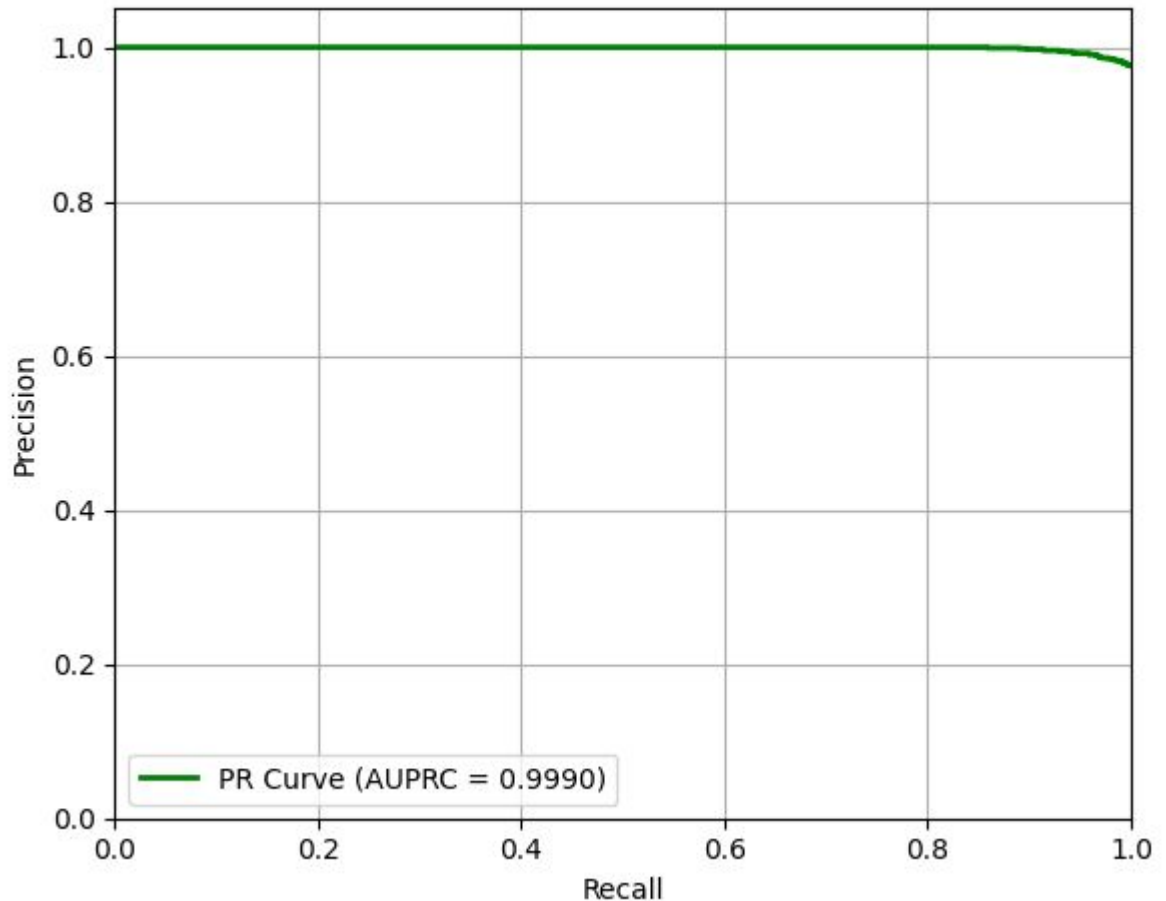
Accuracy: $\frac{TP + TN}{TP + TN + FP + FN}$

ROC (Receiver Operating Characteristic) curve



Area under curve: 96.0%

Precision-Recall curve



Precision-recall: 99.9%

Remaining steps

Step 3 - characterization and labelling

- Threshold anomaly score for fault characterization
- Labelling anomalies based on clusters

Step 4 - generalization and prediction

- Repetition for all cavities
- Add extra information with scalars data
- Searching -30 minutes before each anomaly for signs of impending fault - precursors

Challenges

1. Limitation of using 3D tensors (cycle IDs, signals, time stamps) due to PM data not having common cycle IDs for all cavities, but OK for DD data
2. Inclusion of BCM, BPM and scalars data required for further information on anomalies, can also act as precursors
3. Use of features versus waveforms as inputs for the models

Thank you, it was fun :)
