Artificial Intelligence in Research and Applications Seminar
GEIST Research Group - Jan.2024

Online Anomaly Explanation

A Case Study in Metro do Porto

Rita P. Ribeiro, João Gama, Bruno Veloso, Miguel Silva
Narges Davari, Ehsan Aminian, Saulo Mastellini, Pedro Pereira

U.PORTO INESCTEC

XPM
eXplainable Predictive Maintenance
CHIST-ERA-19-XAI-012

Motivation

- In data-driven Predictive Maintenance (PdM) problems, deep learning techniques are quite popular
  - good predictive accuracy and capability of modeling complex systems
- A critical issue in PdM applications is the design of a maintenance plan after a fault is detected or predicted.
- It is important to identify the causes of the failure and the component in failure.
- Predictions made by black-box models are difficult for human experts to understand and make the correct decisions.
- Explanations are needed!

Motivation

Overall Goal

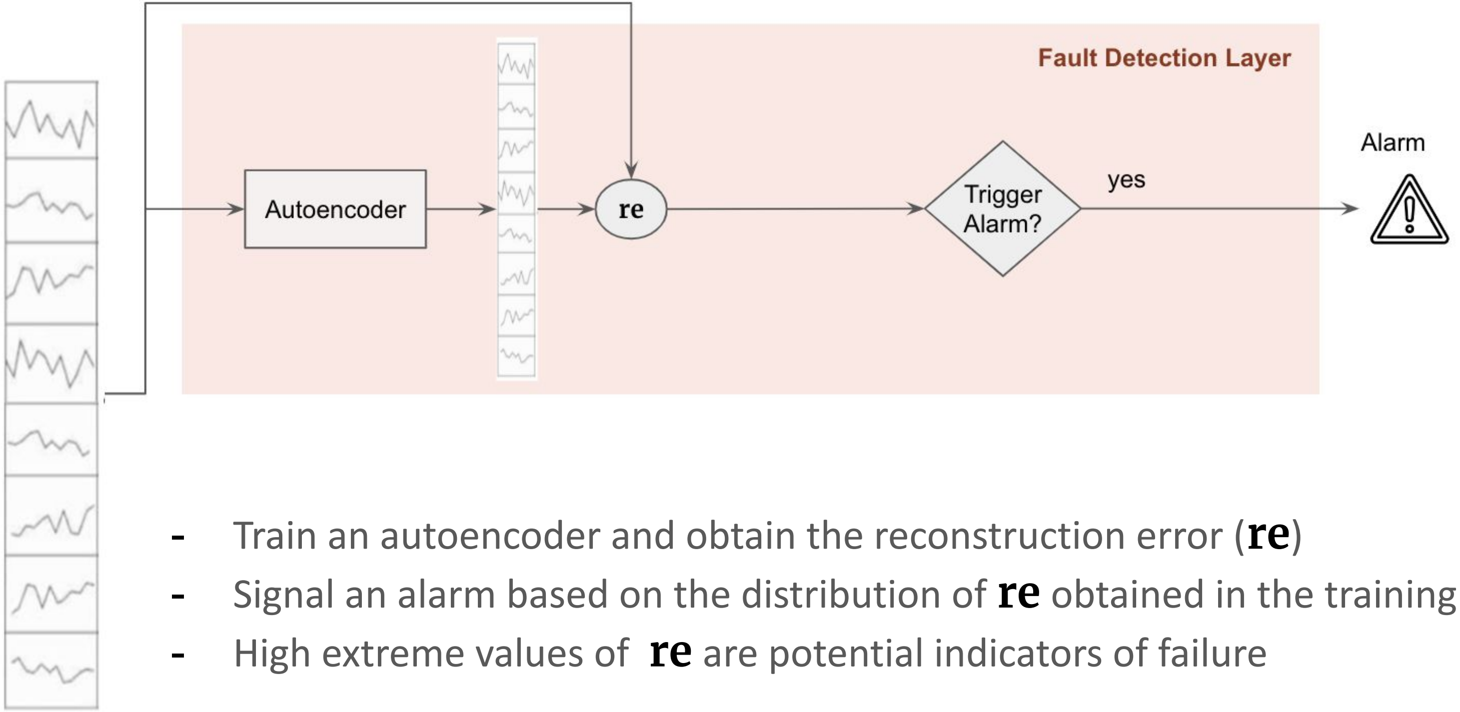
- Predict, identify, and describe the occurrence of defects in the operational units of a system.
- Common XAI methods (e.g. LIME, SHAP) act mostly on offline scenarios.

Contribution

- Online anomaly detection and explanation for faults based on two layers:
  - Fault Detection Layer based on deep learning
  - Anomaly Explanation layer based on rule learning algorithms

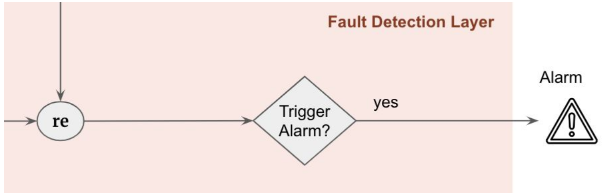
Ribeiro, R.P., Mastelini, S.M., Davari, N., Aminian, E., Veloso, B., Gama, J. (2022):
Online anomaly explanation: A case study on predictive maintenance. In IoTStreams Workshop - ECML PKDD 2022

Fault Detection

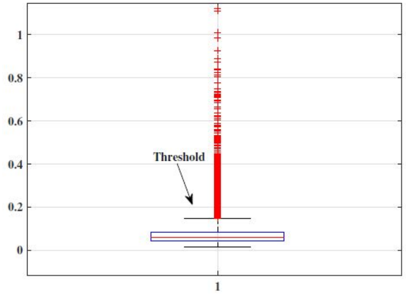


- Train an autoencoder and obtain the reconstruction error (**re**)
- Signal an alarm based on the distribution of **re** obtained in the training set
- High extreme values of **re** are potential indicators of failure

Fault Detection



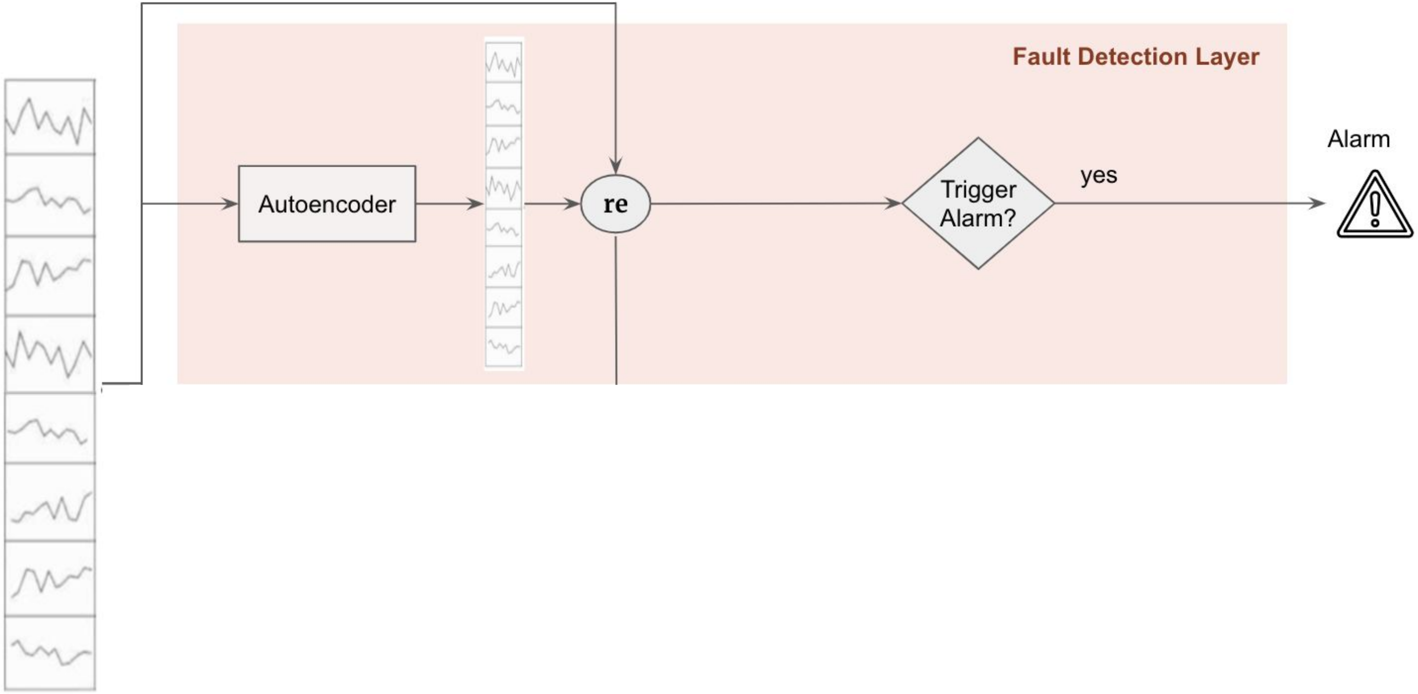
- When to trigger an alarm?
  - High extreme values of re are indicators of fault
  - Use boxplot to identify extreme outliers in the training re distribution
  - If re > Q3+3.IQR then fault (1) else normal (0)



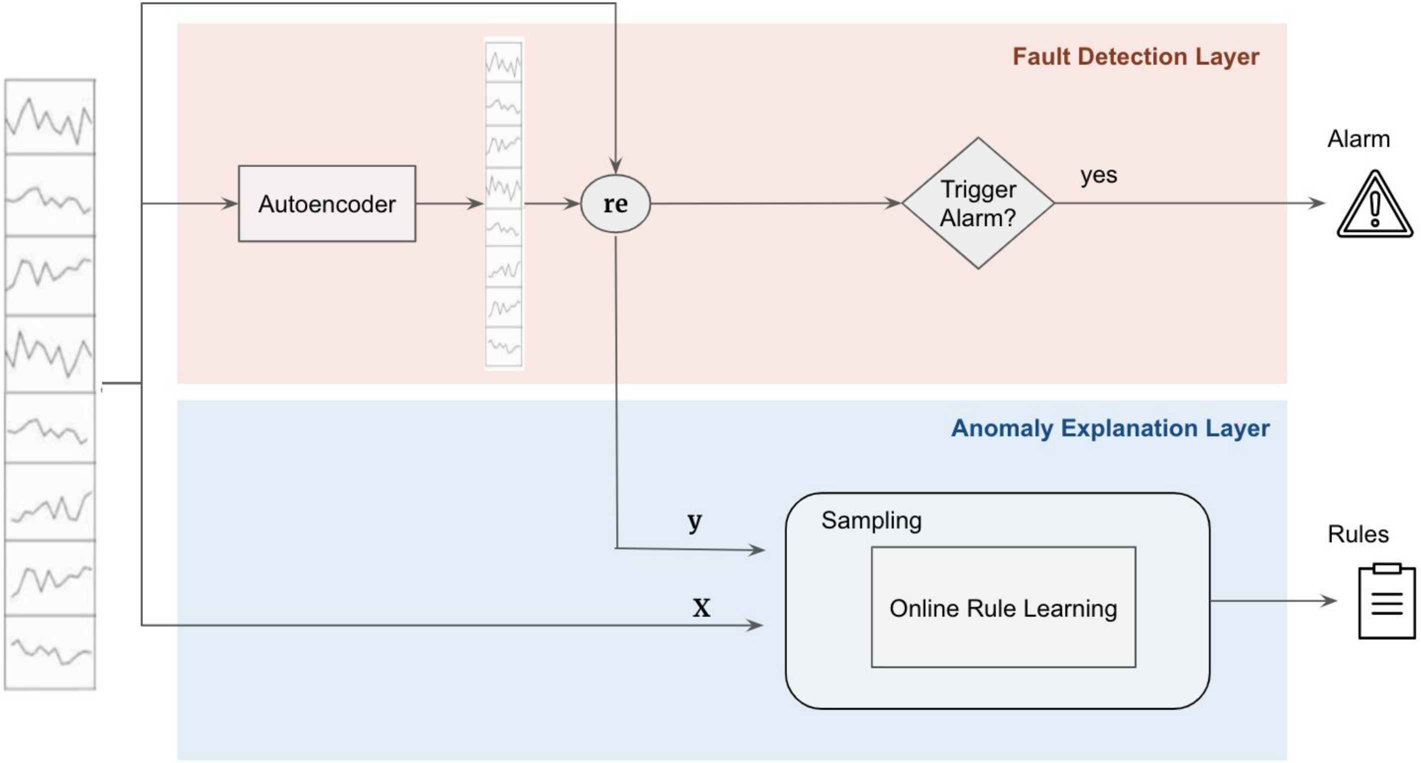
  - Apply a low-pass filter to fault/normal output
    - smooth high frequencies (abrupt changes)
    - reduce false alarms
  - Signal an alarm when subsequent faults makes output > 0.5

Ribeiro, R.P., Pereira, P., Gama, J. (2016): Sequential anomalies: a study in the railway industry. Machine Learning 105, 127–153

Online Anomaly Explanation System

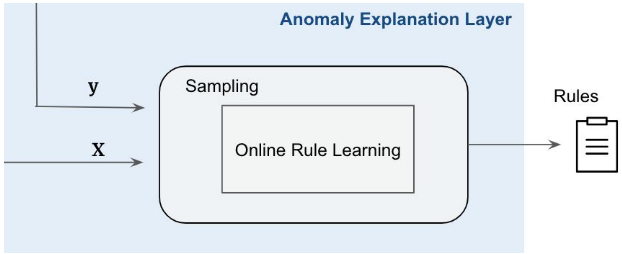


Online Anomaly Explanation System



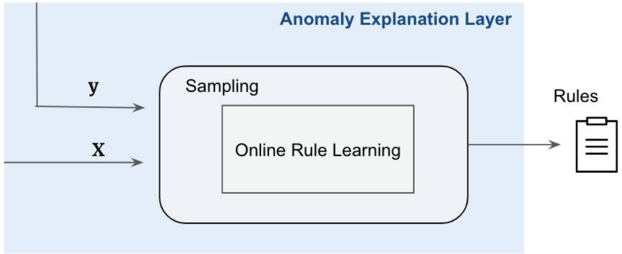
Fault Explanation

- How to derive the set of rules?
  - Input features of the AE ($\mathbf{X}$) and $\mathbf{re}$ as target ($\mathbf{y}$)
  - Adaptive Model Rules (AMRules)
    - Rule-based algorithm for incremental regression tasks
    - The consequent of the rule contains the sufficient statistics to:
      - expand a rule
      - make predictions
      - detect changes
  - Output:
    - Ordered set of rules: decision list that outputs the first rule that covers an example



Duarte, J, Gama, J., Bifet, A. (2016): Adaptive Model Rules From High-Speed Data Streams. ACM Trans. Knowl. Discov. Data

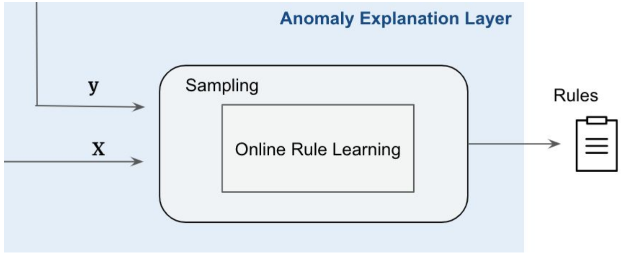
Fault Explanation



- How to derive the set of rules?
  - Both layers run online and in parallel
  - For each example
    - classifies it as fault/normal and inputs it to the rule learning algorithm
  - But, it is an imbalanced regression data stream scenario
  - The goal is to be accurate at high extreme values of re
  - One approach is to resort to data-level strategies to tackle imbalance
    - Sampling strategy based on Chebyshev's inequality

Duarte, J, Gama, J., Bifet, A. (2016): Adaptive Model Rules From High-Speed Data Streams. ACM Trans. Knowl. Discov. Data

Fault Explanation



- How to derive the set of rules?
  - Chebyshev's Inequality
    - Y random variable
    - finite expected value $\overline{y}$ and finite non-zero variance σ^2
    - for any real number t > 0, we have

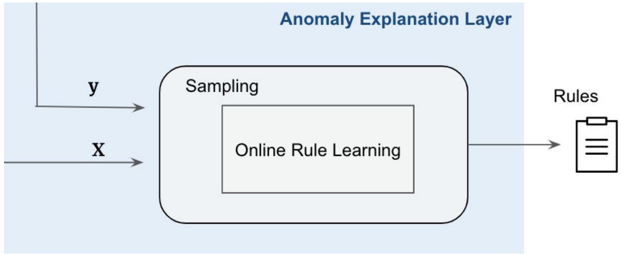
$$\Pr(|y - \overline{y}| \geq t\sigma) \leq \frac{1}{t^2}$$

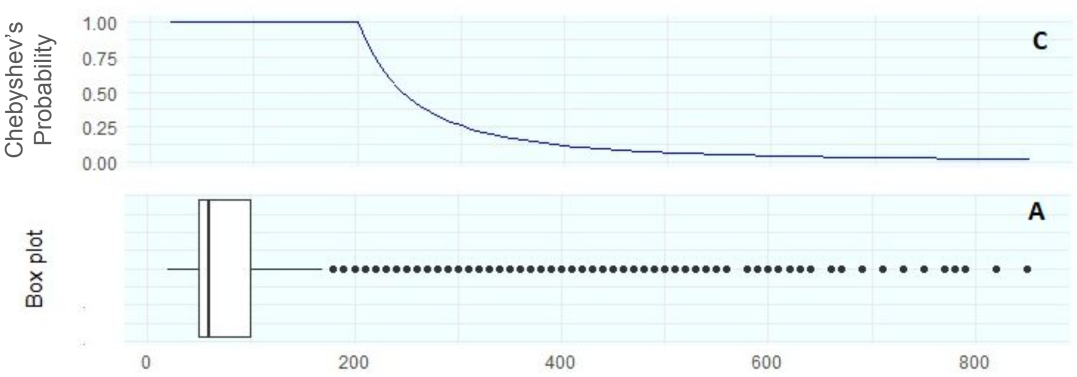
  - Based on the updated mean and variance of y
  - Use it as a heuristic to disclose the type incoming examples (average or extreme value)

Aminian, E., Ribeiro, R.P., Gama, J. (2021): Chebyshev approaches for imbalanced data streams regression models. Data Mining and Knowledge Discovery 35, 2389–2466

Fault Explanation

- How to derive the set of rules?
  - ChebyUS
    - under-sampling strategy
    - the example selection is inversely proportional to the Chebyshev's probability

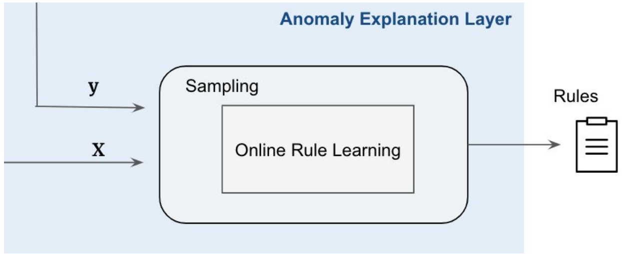


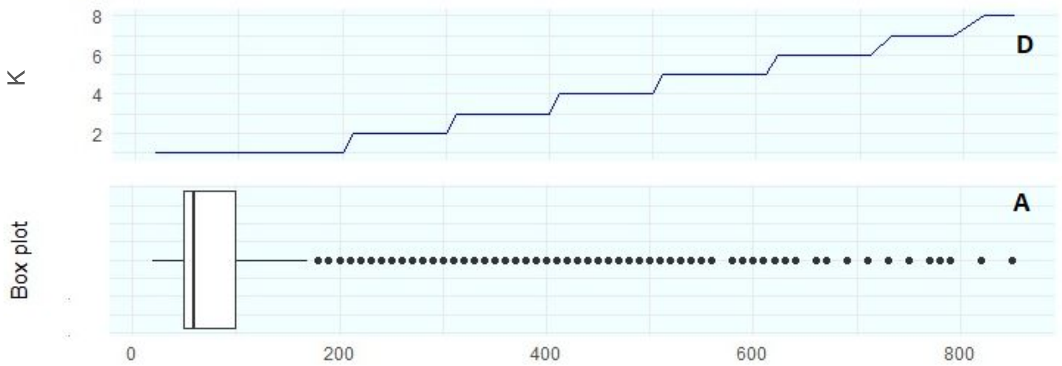


Aminian, E., Ribeiro, R.P., Gama, J. (2021): Chebyshev approaches for imbalanced data streams regression models. Data Mining and Knowledge Discovery 35, 2389–2466

Fault Explanation

- How to derive the set of rules?
  - ChebyOS
    - over-sampling strategy
    - the example is replicated as many times (K) as far it is from the mean, given that it is farther than the variance





Aminian, E., Ribeiro, R.P., Gama, J. (2021): Chebyshev approaches for imbalanced data streams regression models. Data Mining and Knowledge Discovery 35, 2389–2466

Online Anomaly Explanation System

This architecture allows two levels of explanations

- Global Explanations
  - These rules reproduce the AE network behaviour
  - They explain the conditions when and why AE predicts high re values.
- Local Explanations
  - Rules that are triggered by a given an example.
  - If re exceeds a threshold, the system outputs the rules triggered by that example.

Case Study on Metro do Porto

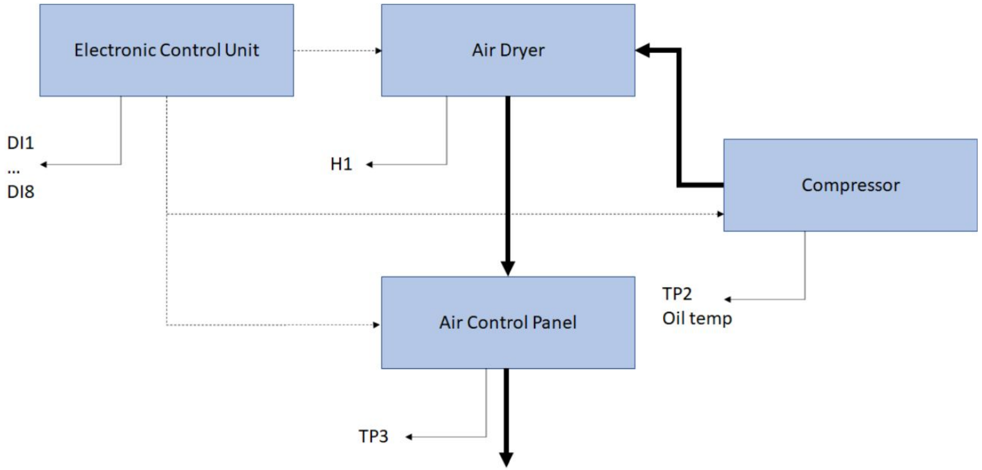
Case Study on Metro do Porto

- The Compressed-Air Production Unit (APU) in Metro do Porto fleet is a vital system without redundancy.
- Responsible for alignment of the vehicle with the platform at the stations depending on the number of passengers.
- Some of its failures are undetectable according to traditional maintenance criteria.
- It is one of the equipments that most contribute to the cancellation of trips.
- Goal:
  identify normal/abnormal behaviours in the data stream obtained from sensors installed in the APU system while the train is in operation.

Case Study on Metro do Porto

MetroPT-2 data set

- Air Production Unit (APU)
  - 16 sensors installed in 4 modules
- 1 Hz sampling frequency
- 5 minute data packets
- 3 month data sample
- 2022-04-28 until 2022-07-28
- $\approx 7 \times 10^6$ examples.



Veloso, B., Gama, J., Ribeiro, R., & Pereira, P. (2022). MetroPT-2: A Benchmark dataset for predictive maintenance (Version V2) [Data set]. Zenodo

Case Study on Metro do Porto

MetroPT-2 data set: 8 analogue sensors

nr.	Module	Description
Analogue		
1	Compressor	TP2 - Compressor Pressure
2	Air Control Panel	TP3 - Pneumatic panel Pressure
3	Air Control Panel	H1 - Pressure above 10.2 Bar
4	Air Dryer	DV - Air Dryer Tower Pressure
5	Air Control Panel	Reservoirs - Pressure
6	Compressor	Oil Temperature
7	Air Control Panel	Flow meter
8	Compressor	Motor Current

Case Study on Metro do Porto

MetroPT-2 data set: 8 digital sensors

Digital		
9	Electronic Control Unit	COMP - Compressor on/off
10	Electronic Control Unit	DV electric - Compressor outlet valve
11	Electronic Control Unit	Towers - Active tower number
12	Electronic Control Unit	MPG - Pressure below 8.2 Bar
13	Electronic Control Unit	LPS - Pressure is lower than 7 bars
14	Electronic Control Unit	Towers Pressure
15	Compressor	Oil Level - Level below min
16	Air Control Panel	Caudal impulses

Case Study on Metro do Porto

Problem Definition

- Detect an upcoming catastrophic failure
  - train breaking down and having to be replaced
- Warning must be given 2 hours before the LPS signal is active
- Ground truth: failures indicated in the Maintenance Report

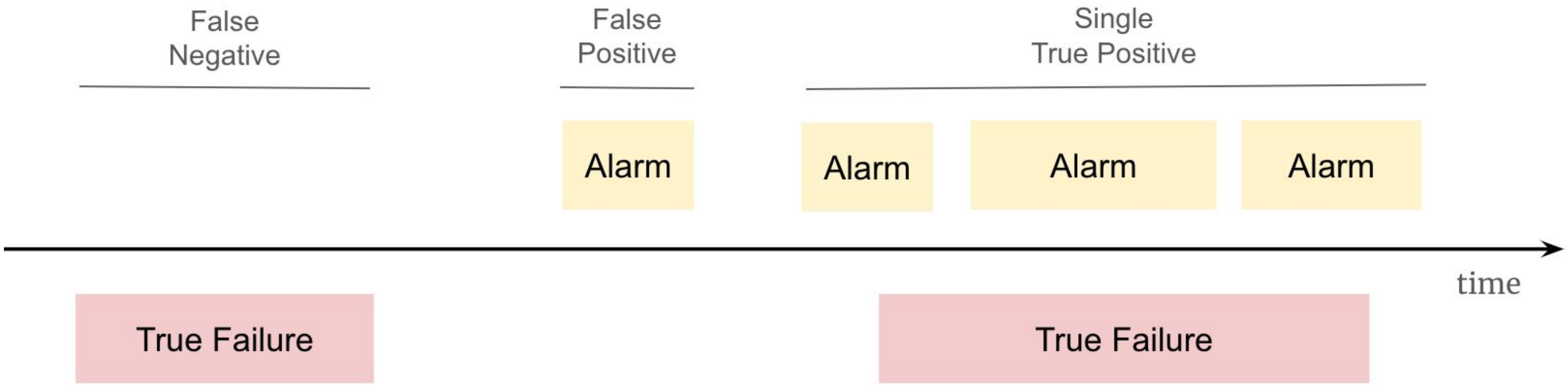
#	Start Time	End Time	Failure	LPS Time
1	2022-06-04 10:19:24	2022-06-04 14:22:39	Air Leak	2022-06-04 11:26:01
2	2022-07-11 10:10:18	2022-07-14 10:22:08	Oil Leak	2022-07-13 19:43:52

Silva, M., Veloso, B., Gama, J. (2023): Predictive Maintenance, Adversarial Autoencoders and Explainability. ECML PKDD 2023

Case Study on Metro do Porto

Evaluation

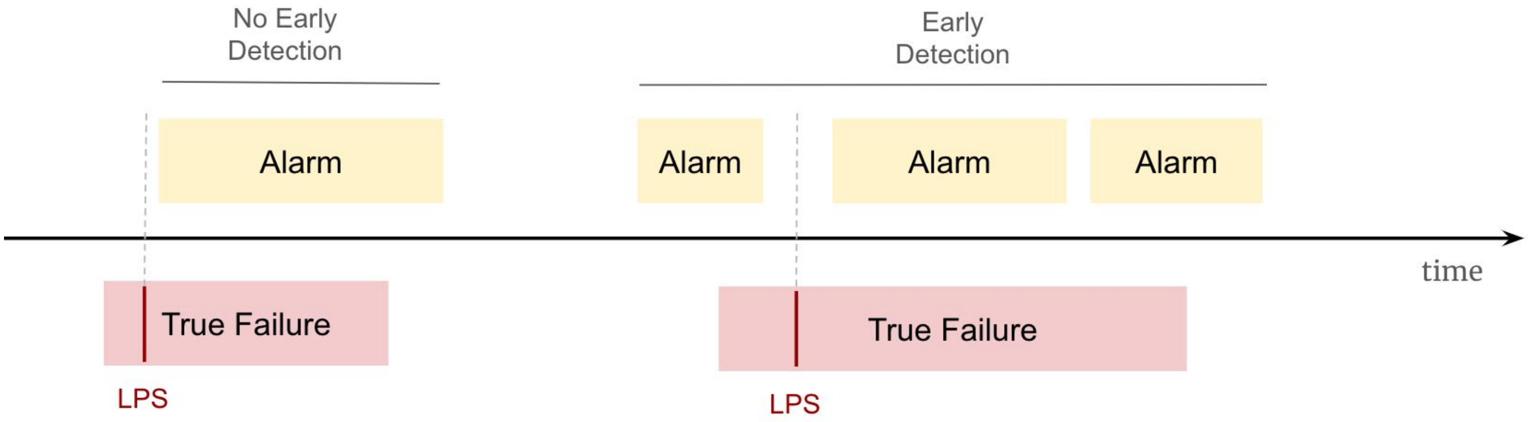
- True Positive: model outputs a failure that overlaps with the observed failures.
- Predicted failures that are less than 1 day apart are merged



Case Study on Metro do Porto

Evaluation

– **Early detection:** 2 hours before LPS signal is active

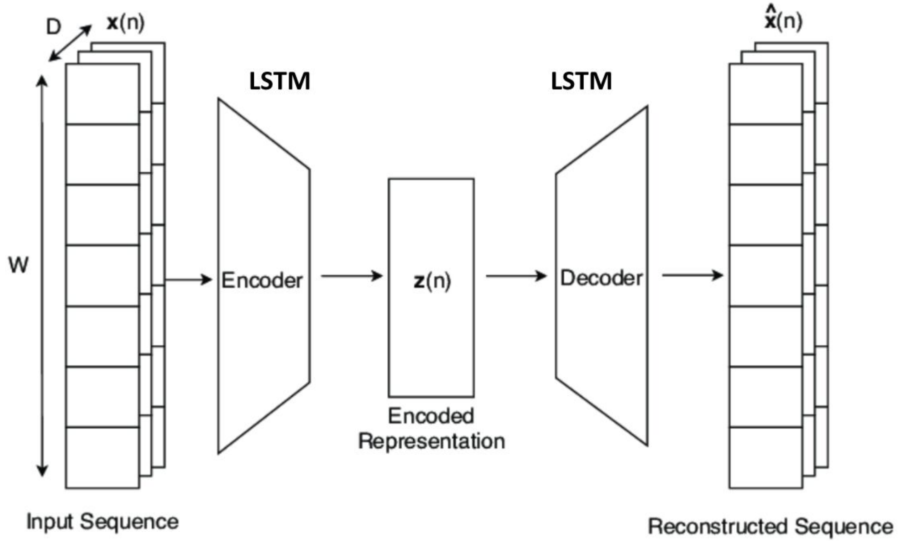


Experimental Setup

- Data Chunks
  - Compressor cycles-based approach
    - failures may contain few compressor cycles - no alarm after low pass filter.
    - compressor cycles may last until train breaks down - no opportunity for early detection.
  - Fixed time-based approach
    - data chunks of 30 min
- Prequential Evaluation
  - train: 1 month = 8674 sequences / test: 2 months = 14591 sequences
- Methods
  - AE + AMRules with ChebyOS

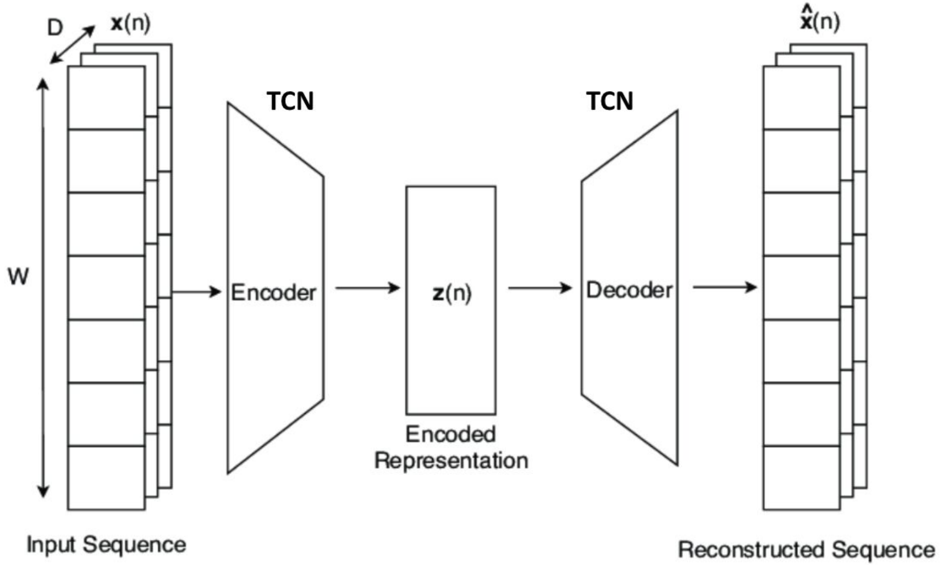
Experimental Setup: Autoencoders

- **Long-Short Term Memory Autoenconder (LSTM-AE)**



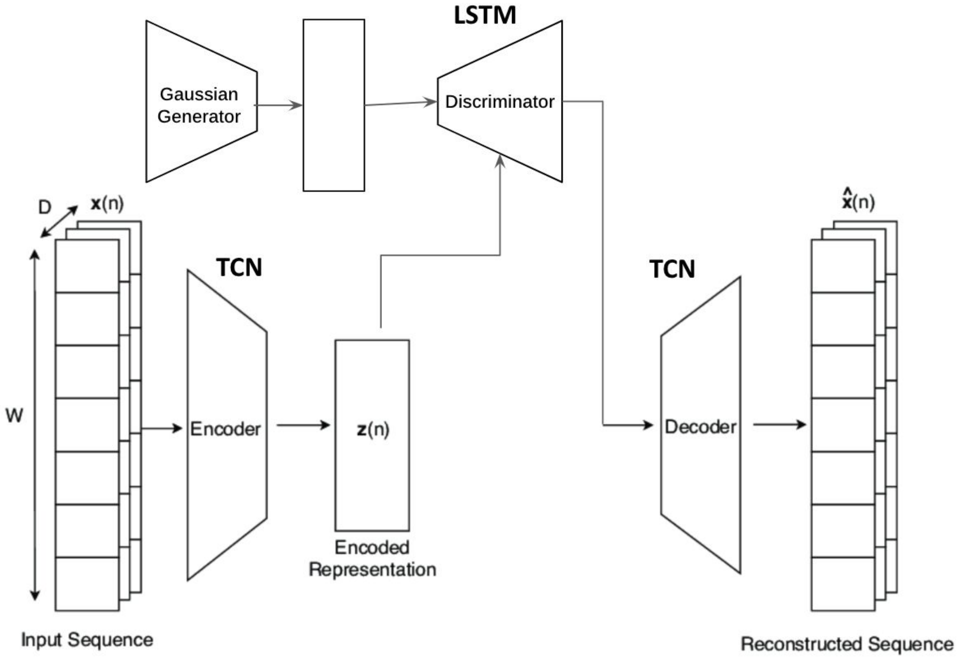
Experimental Setup: Autoencoders

- **Temporal Convolution Network Autoencoder (TCN-AE)**

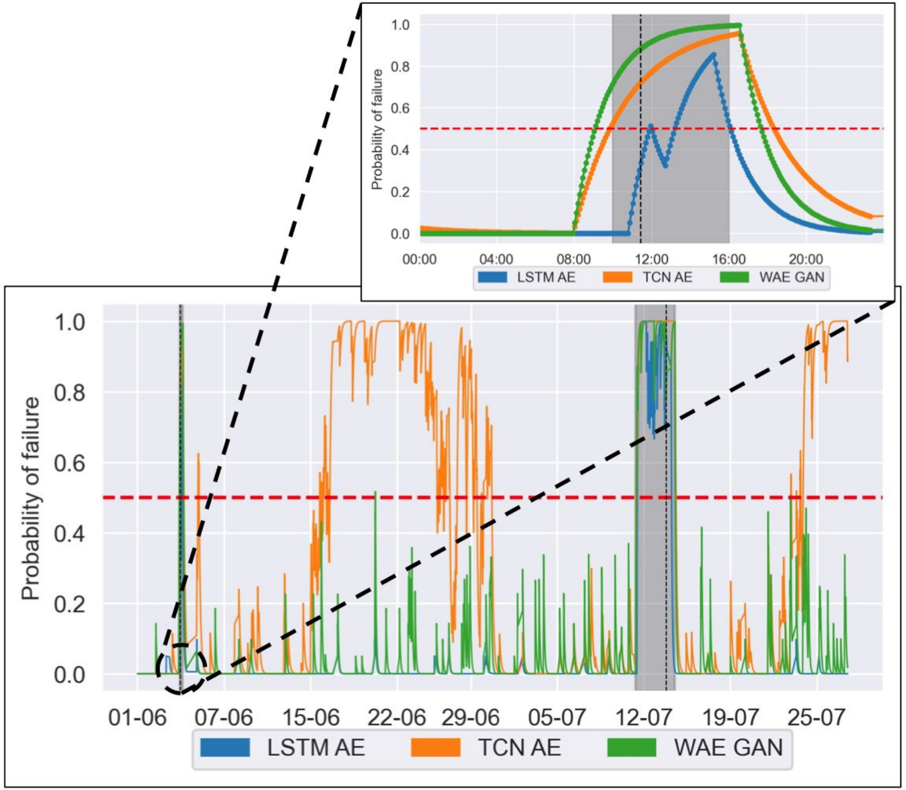


Experimental Setup: Autoencoders

- **Wasserstein Autoencoders with Generative Adversarial Networks (WAE-GAN)**



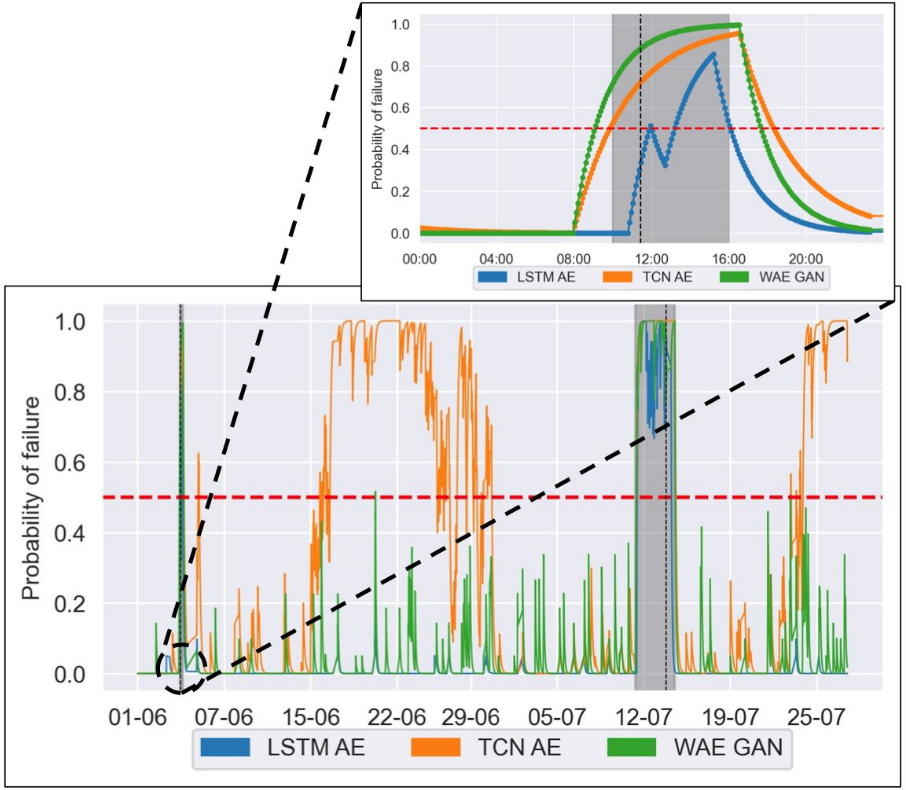
Experimental Results: Online Failure Detection



LSTM AE

- detects both failures
- does not generate false alarms
- is unable to detect the first failure before the LPS signal.

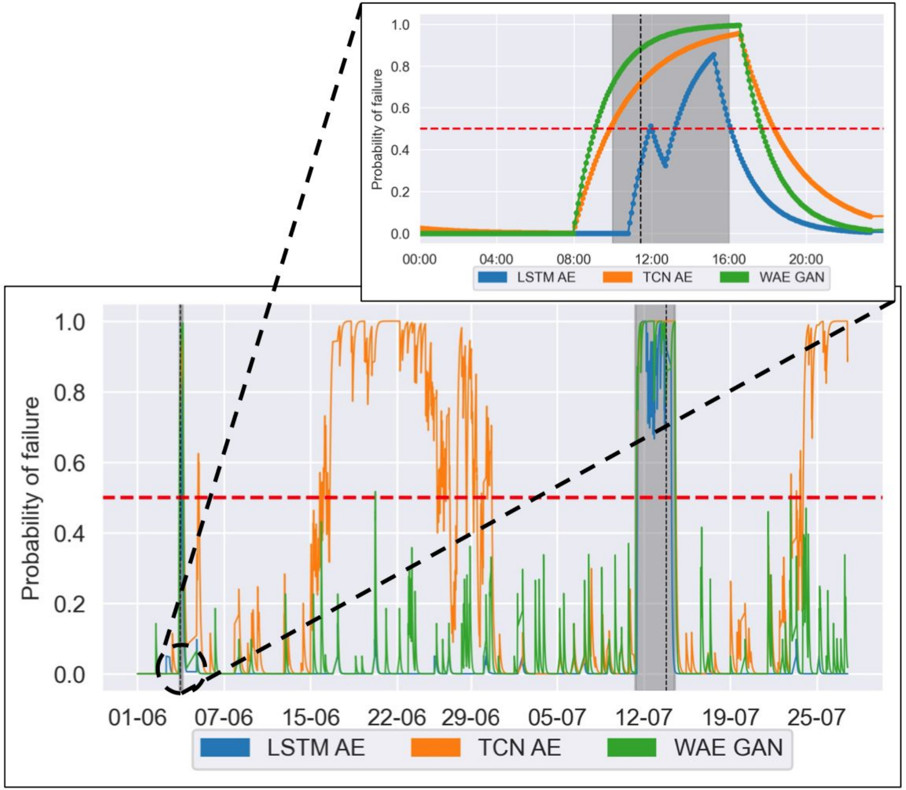
Experimental Results: Online Failure Detection



TCN-AE

- detects both failures early
- generates two false alarms
- F1 of 0.67

Experimental Results: Online Failure Detection



WAE-GAN

- detects the two failures at least 2h before the LPS signal is active
- does not any false alarm
- achieves a perfect F1 score

Experimental Results: Online Explainability

Rules for the 1st failure - Air Leak

R1: $(H1 \leq 8.8\ bar) \wedge (\text{Oil Temperature} > 58.5°\text{C})$

[Active time: 5.6% prior, 68% during failure.]

R2: $(\text{Oil Temperature} > 60.8°\text{C}) \wedge (TP2 > 9.2\ bar) \wedge (Reservoirs > 9.8\ bar)$

[Active time: 0.3% prior, 0.8% during failure.]

R3: $(\text{Motor Current} > 3.8\ A) \wedge (7.0\ bar \leq TP2 \leq 7.2\ bar) \wedge (\text{Oil Temperature} > 58.5°\text{C})$

[Active time: 0.01% prior, 7.3% during failure.]

Experimental Results: Online Explainability

Rules for 2nd failure - Oil Leak

R1: $(65.1^\circ\mathrm{C} \leq \text{Oil Temperature} \leq 71.5^\circ\mathrm{C}) \wedge (H1 \leq 9.6\ bar) \wedge$
$(Reservoirs > 8.8\ bar) \wedge (Flowmeter > 0.2 m^3/h)$
[Active time: 0.3% prior, 37% during failure.]

R2: $(\text{Oil Temperature} > 65.1^\circ\mathrm{C}) \wedge (H1 > 0\ bar)$
[Active time: 0.8% prior, 48% during failure.]

R3: $(\text{Oil Temperature} > 54.6^\circ\mathrm{C}) \wedge (TP2 > 9.2\ bar)$
[Active time: 2.6% prior, 6.5% during failure.]

R4: $(Flowmeter > 25 m^3/h) \wedge (\text{Oil Temperature} < 95.8^\circ\mathrm{C})$
[Active time: 0.01% prior, 9.1% during failure.]

Case Study on Metro do Porto

Wrap-up

- Diagnosing failures by modelling time series of sensor data.
- Deep autoencoder architectures with different regularization mechanisms.
- Autoencoder architecture with adversarial regularization achieves requirements of early detection and no false alarms.
- Explainability rules indicate failures are explained by sensors related to problems discovered by maintenance teams.

Online Anomaly Explanation in Predictive Maintenance

Two-layer architecture

- The two learning systems, the deep learning and the rule learning system, are complementary.
- AE works in unsupervised mode using data from the normal behavior.
- The rule learner works in supervised mode, where the target is the reconstruction error of the AE computed in real-time.
- The methodology is general enough to be applied to other online imbalanced streaming scenarios that use black-box models to predict peaks or bursts in events.

References

Aminian, E., Ribeiro, R.P., Gama, J. (2021): Chebyshev approaches for imbalanced data streams regression models. Data Mining and Knowledge Discovery 35, 2389–2466 (2021) https://doi.org/10.1007/s10618-021-00793-1

Davari, N., Veloso, B., Ribeiro, R.P., Pereira, P.M., Gama, J. (2021): Predictive maintenance based on anomaly detection using deep learning for air production unit in the railway industry. In: DSAA 2021 https://doi.org/10.1109/DSAA53316.2021.9564181

Davari, N., Veloso, B., Ribeiro, R.P., Gama, J. (2022): Fault forecasting using data-driven modeling: A case study for Metro do Porto data set. In IoTStreams Workshop - ECML PKDD 2022 https://doi.org/10.1007/978-3-031-23633-4_26

Davari, N., Veloso, B., Ribeiro, R.P., Gama, J. (2023). MetroPT-3 Dataset. UCI Machine Learning Repository. https://doi.org/10.24432/C5VW3R

Duarte, J, Gama, J., Bifet, A. (2016): Adaptive Model Rules From High-Speed Data Streams. ACM Trans. Knowl. Discov. Data; 2016 https://doi.org/10.1145/2829955

Pashami, S., et al. (2023): Explainable Predictive Maintenance. CoRR abs/2306.05120 (2023) https://doi.org/10.48550/arXiv.2306.05120

References

Ribeiro, R.P., Mastelini, S.M., Davari, N., Aminian, E., Veloso, B., Gama, J. (2022): Online anomaly explanation: A case study on predictive maintenance. In IoTStreams Workshop - ECML PKDD 2022 https://doi.org/10.1007/978-3-031-23633-4_25

Ribeiro, R.P., Moniz, N. (2020) Imbalanced regression and extreme value prediction. Mach Learn 109, 1803–1835 (2020). https://doi.org/10.1007/s10994-020-05900-9

Ribeiro, R.P., Pereira, P., Gama, J. (2016): Sequential anomalies: a study in the railway industry. Machine Learning 105, 127–153 (2016) https://doi.org/10.1007/s10994-016-5584-6

Silva, A., Ribeiro, R.P., Moniz, N. (2022). Model Optimization in Imbalanced Regression. In: Pascal, P., Ienco, D. (eds) Discovery Science. DS 2022. Lecture Notes in Computer Science(), vol 13601. Springer, Cham. https://doi.org/10.1007/978-3-031-18840-4_1

Silva, M., Veloso, B., Gama, J. (2023): Predictive Maintenance, Adversarial Autoencoders and Explainability. ECMLPKDD 2023. https://doi.org/10.1007/978-3-031-43430-3_16

Veloso, B., Gama, J., Ribeiro, R., & Pereira, P. (2022). MetroPT2: A Benchmark dataset for predictive maintenance (Version V2) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7766691

Online Anomaly Explanation

A Case Study in Metro do Porto

Thank you for your attention

Rita P. Ribeiro - rpribeiro@fc.up.pt

U.PORTO INESCTEC

eXplainable Predictive Maintenance
CHIST-ERA-19-XAI-012