DECEMBER 2025

UPM- UNIVERSIDAD POLITECNICA DE MADRID

KNOWLEDGE DISCOVERY PROJECT

MULTIVARIATE TIME SERIES ANOMALY DETECTION

DATA MINING & TIME SERIES

ANOMALY DETECTION ON THE SERVER MACHINE DATASET (SMD)

AUTHORS
EMANUELE ALBERTI — LEANDRO DUARTE — OTTAVIA BIAGI
UPM — DATA MINING & TIME SERIES

KDD Framework

- **Business Understanding**: Detect anomalies in server telemetry to support reliability and early failure detection.
- **Data Understanding**: 28 machines × 38 KPIs (CPU, memory, disk I/O, network).
- **Data Preparation**: Scaling, cleaning, window segmentation.
- **Modelling**: Mean baseline, LSTM AE, Transformer AE + ablation.
- **Evaluation**: Precision, Recall, F1, ROC-AUC, PR-AUC.
- **Deployment Perspective**: Feature-level diagnostics.

Dataset Overview

CONTENT

- Server Machine Dataset (SMD)
- Train = only normal behaviour
- Test = normal + anomalies
- Labels at timestamp level
- 38 KPIs per machine
- Sampling ≈ every few seconds
- Highly multivariate, non-stationary

Machine Selection (Global EDA)

DATA ANALYSIS

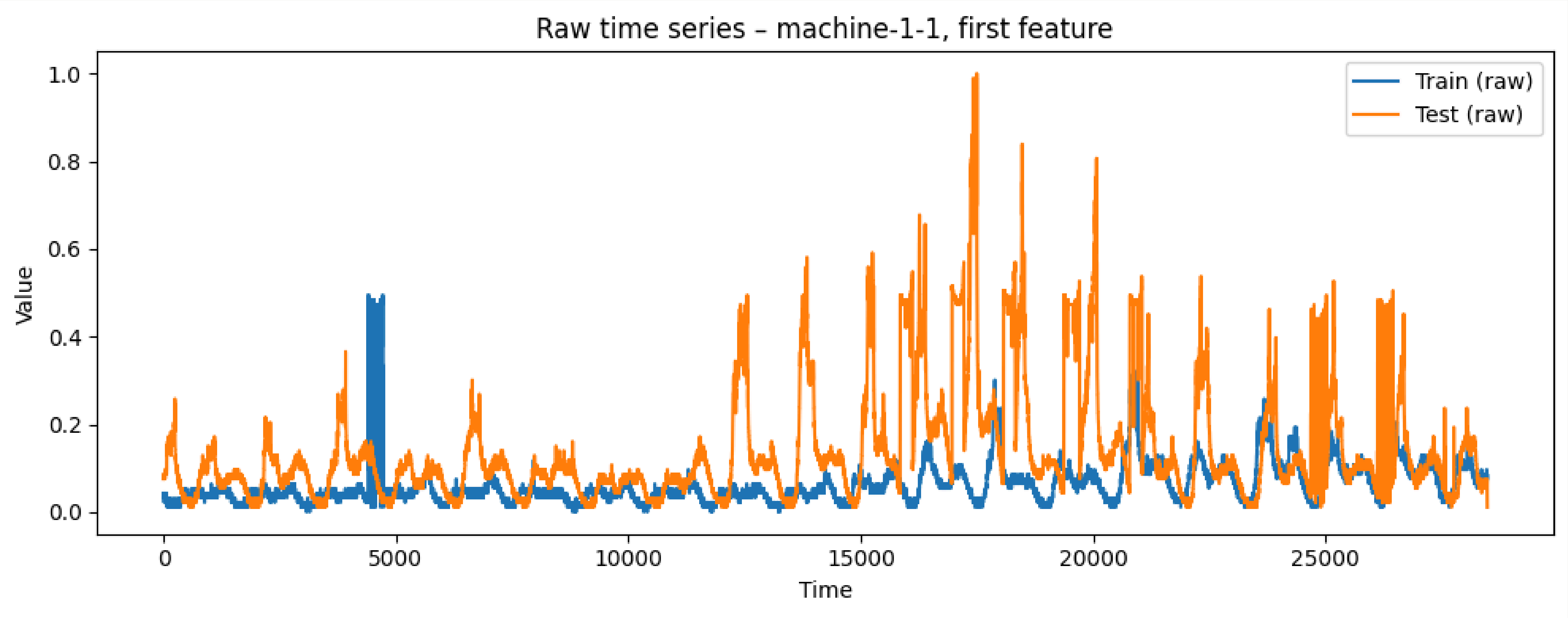
- Variance comparison across 28 machines
- Correlation comparison across 28 machines
- machine-3-7 → highest variance
- machine-1-1 → highest correlation (≈0.458)
- We select machine-1-1 for modelling, because transformers benefit from strong cross-feature dependencies.

	machine	mean_corr
0	machine-1-1	0.457749
1	machine-1-2	0.113721
2	machine-1-3	0.243654
3	machine-1-4	0.266576
4	machine-1-5	0.304304
5	machine-1-6	0.277832
6	machine-1-7	0.250052
7	machine-1-8	0.346967
8	machine-2-1	0.212146
9	machine-2-2	0.252188

Raw Time Series (machine-1-1)

Key observations:

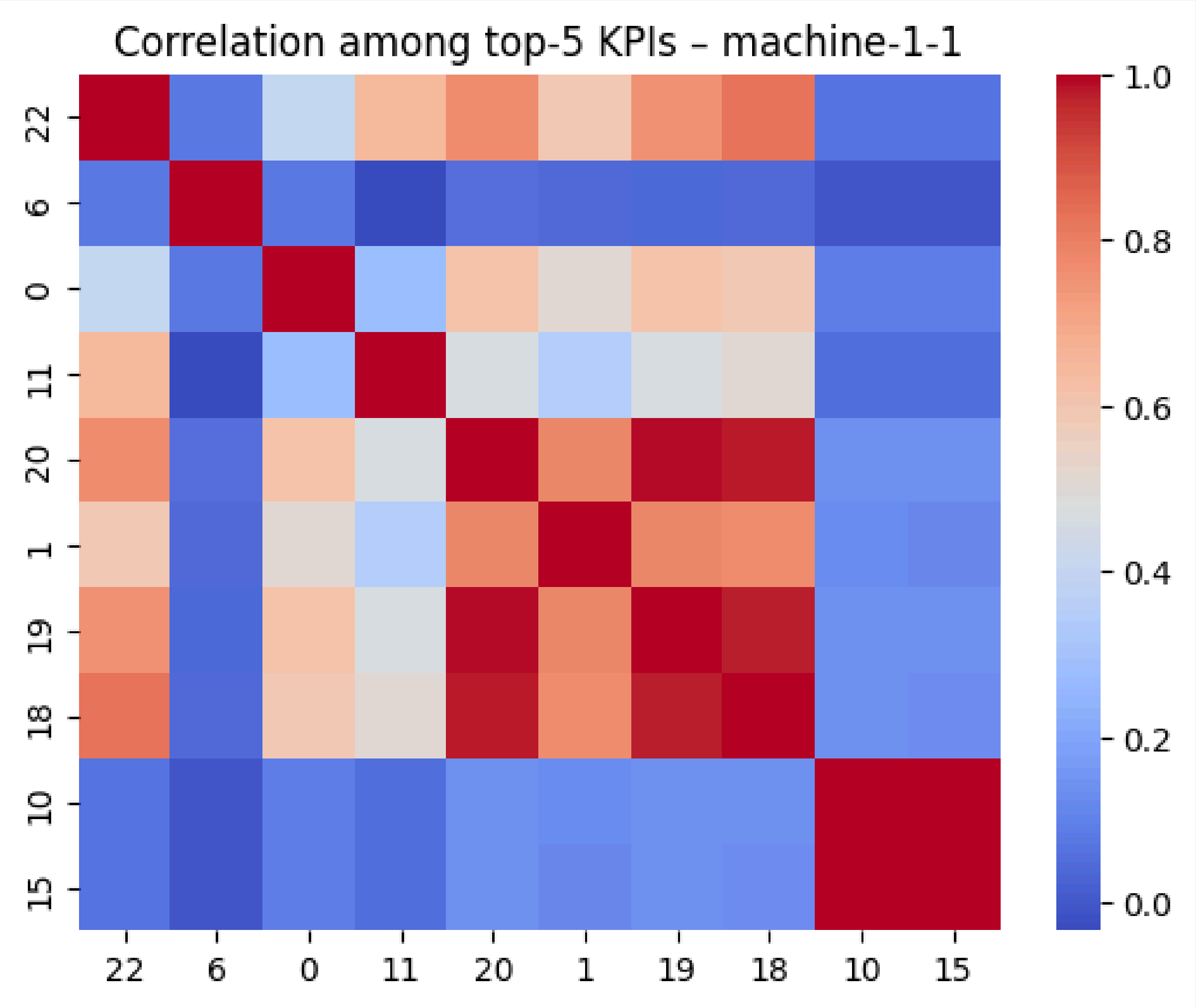
- Train signal: smooth, stable
- Test signal: peaks, shifts → anomalies
- Strong non-stationarity (changing mean + variance)

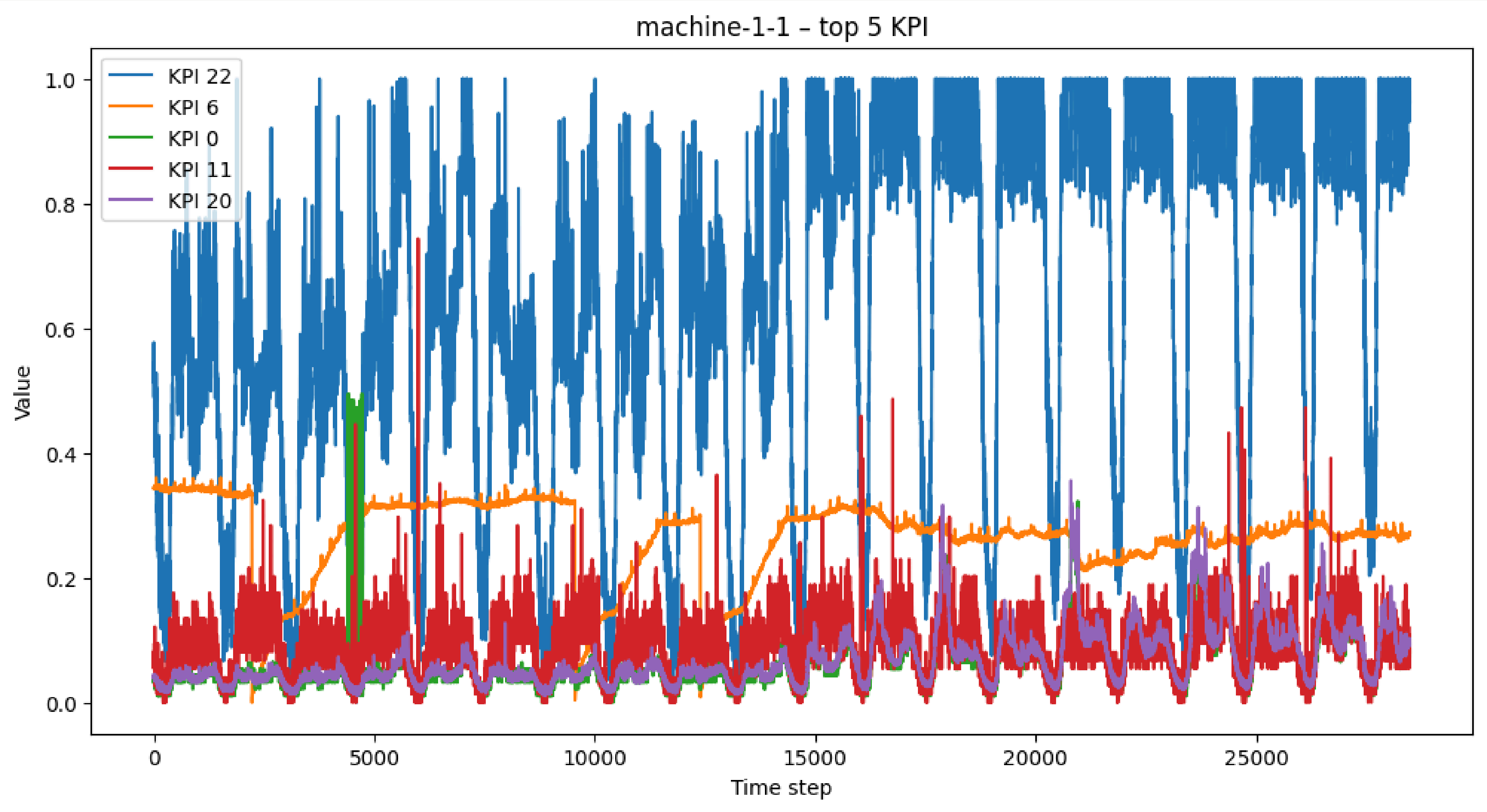


Top KPIs & Multivariate Structure

Key observations:

- KPIs show quasi-periodicity + bursts
- Correlation heatmap reveals 2–3 clusters
- One KPI behaves independently → subsystem behaviour
- Ideal conditions for transformers (self-attention)

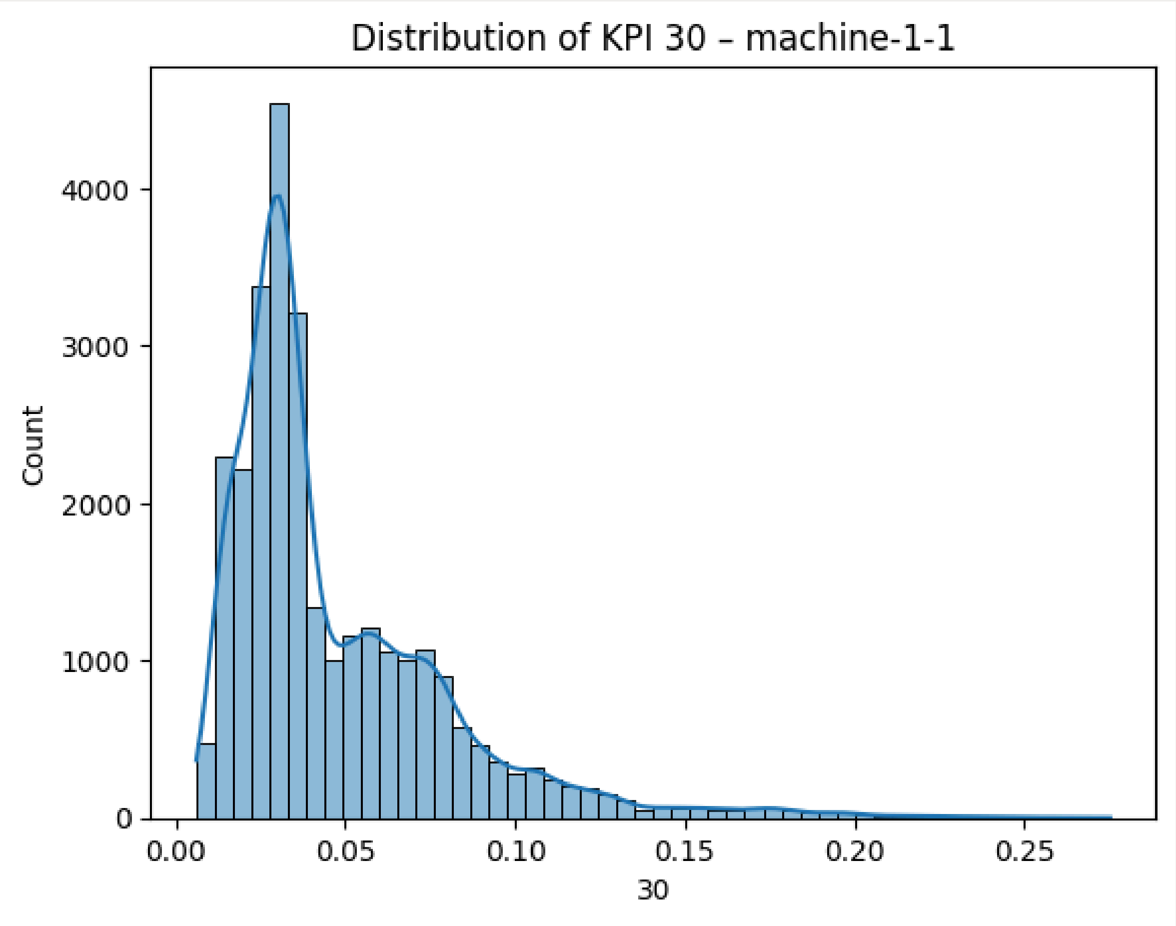


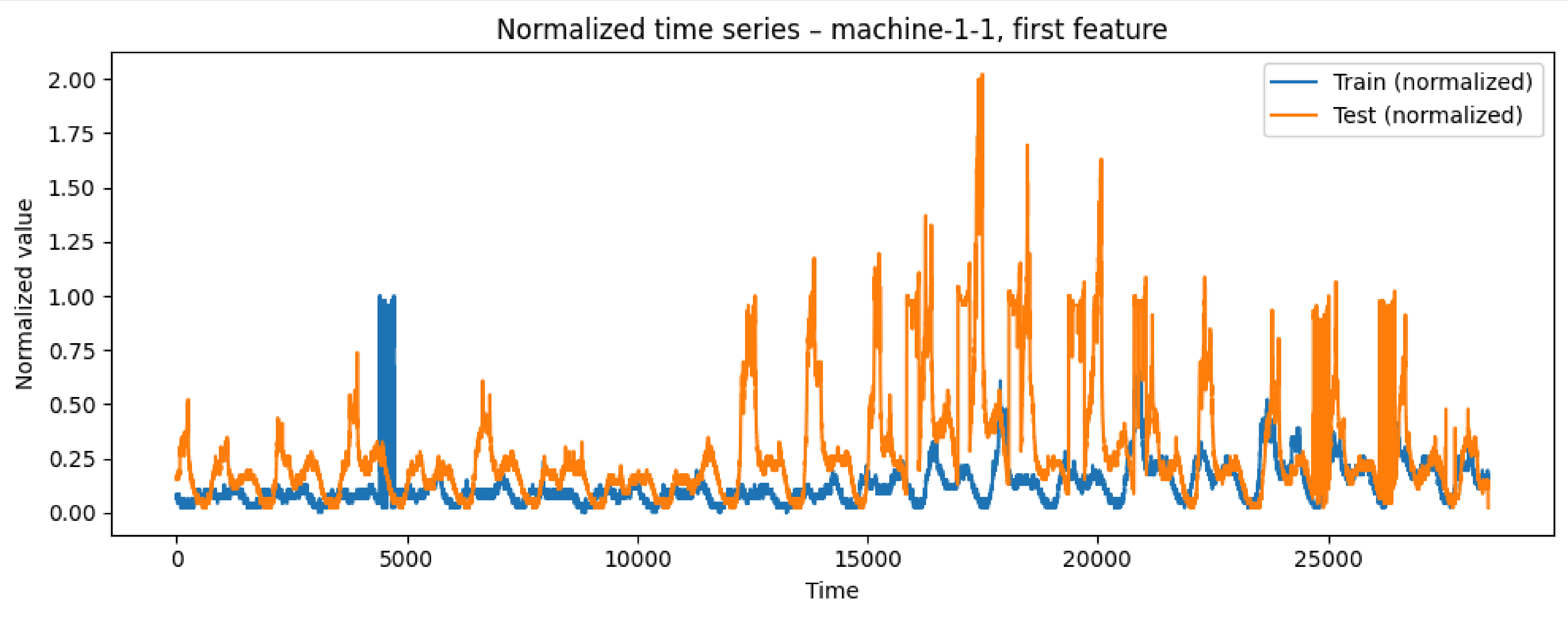


KPI Distributions & Preprocessing

Key observations:

- Heavy-tailed KPI distributions → bursts
- Min–max normalization avoids dominance of large-scale features in MSE
- Sliding-window segmentation with K = 10, 30, 50
- Analogy with ARIMA lag order selection

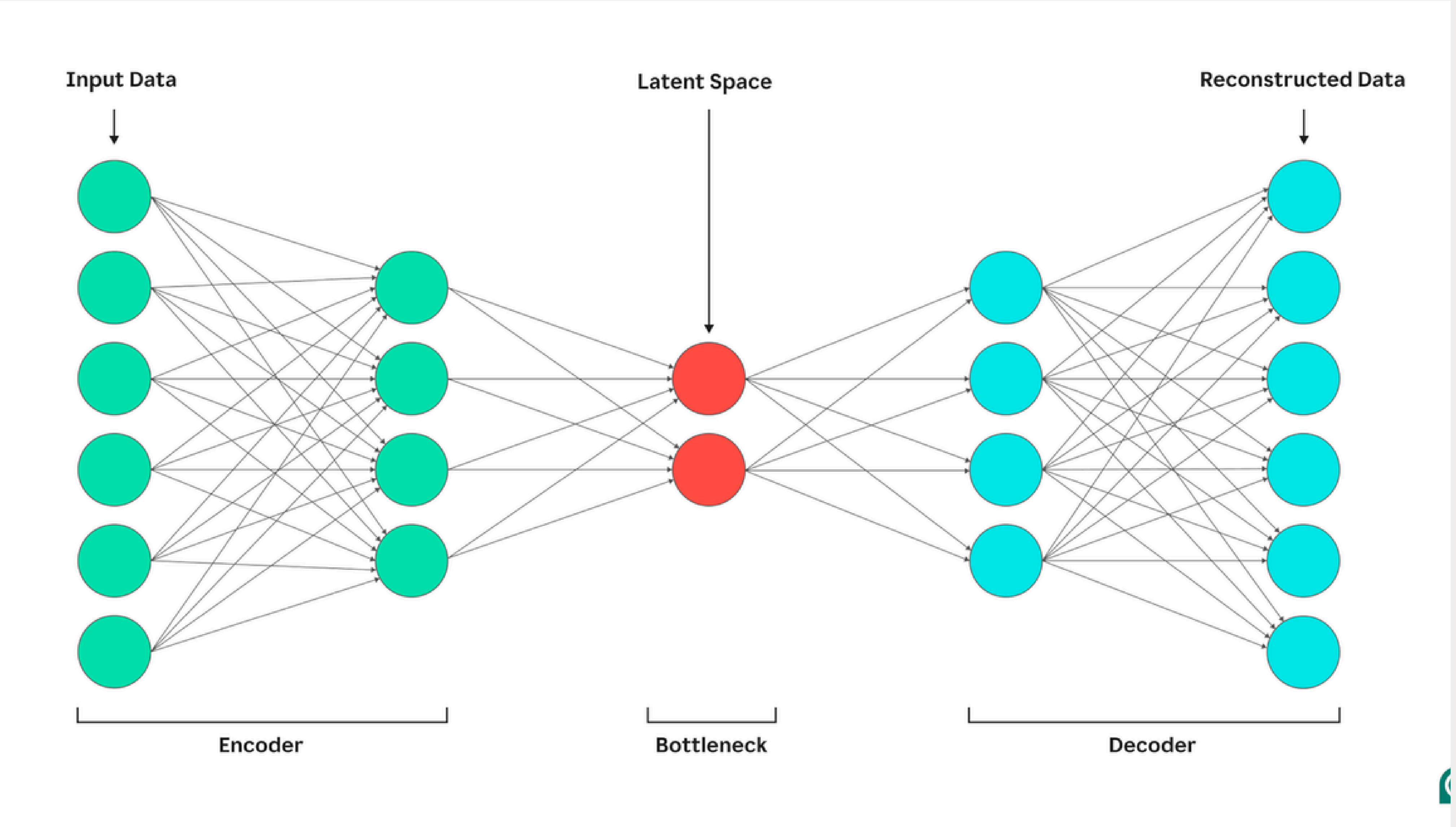




Model Overview

MODELS TESTED:

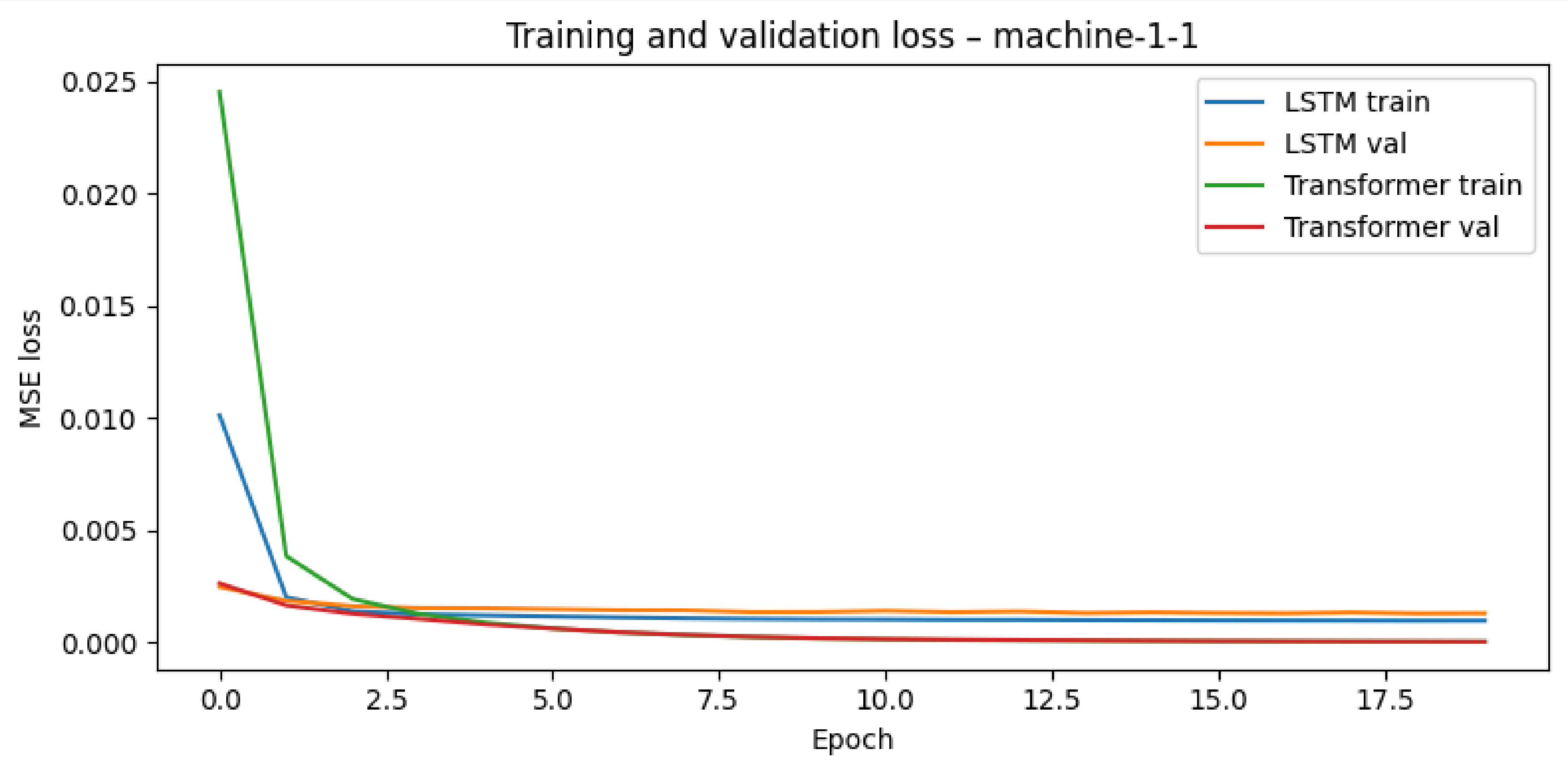
- **Mean reconstruction baseline**
- **LSTM Autoencoder**
- **Transformer Autoencoder**
  - Positional embeddings
  - 4-head attention
  - LN pre-norm
  - Lightweight architecture

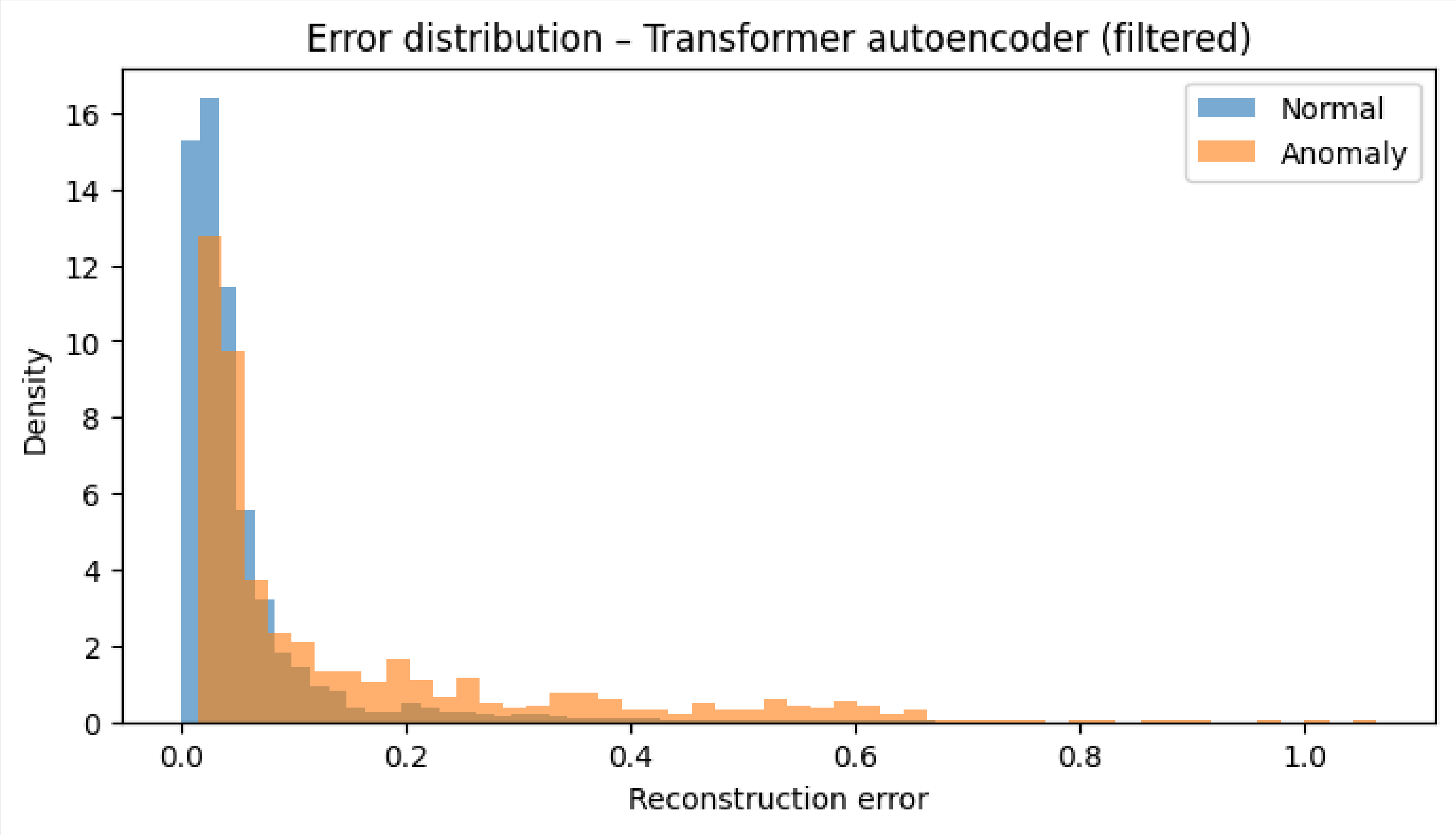


Training Behaviour & Error Distribution

TRASFORMER AUTOENCODER

- Transformer trains slower but reaches lower validation loss
- Overlap between normal and abnormal scores → limited max F1 achievable
- Outliers indicate numerical instabilities → motivate Huber loss





Quantitative Results

MAIN MODEL RESULTS – MACHINE-1-1 (K=10, SMOOTHED THRESHOLD)

- Mean baseline: **F1 = 0.52** (surprisingly strong)
- LSTM AE: high recall, low precision
- Transformer (LN): balanced
- Transformer + Huber: **best learned model** → F1 = 0.49, ROC-AUC = 0.904
- Mean baseline sets a very strong reference: simple global mean → F1 0.52

Model	Precision	Recall	F1	ROC-AUC	PR-AUC
Mean baseline	0.568	0.472	0.516	0.911	0.577
LSTM Autoencoder	0.186	0.99	0.314	0.878	0.516
Transformer (LN, 20 epochs)	0.352	0.448	0.394	0.844	0.42
Transformer (LN, 40 epochs)	0.356	0.542	0.43	0.877	0.462
Transformer (Huber loss)	0.35	0.677	**0.486**	**0.904**	0.437
Transformer (High Dropout)	0.36	0.475	0.41	0.882	0.457
Transformer (Mixed PosEnc)	0.354	0.474	0.406	0.872	0.448
Transformer (NormOut)	0.322	0.437	0.371	0.754	0.375

Effect of Window Size

EFFECT OF WINDOW SIZE ON PERFORMANCE – MACHINE-1-1

- Transformer improves significantly from K=10 → K=30
- K=50 degrades (noise + overfitting)
- Transformers require sufficient context but degrade without meta-learning

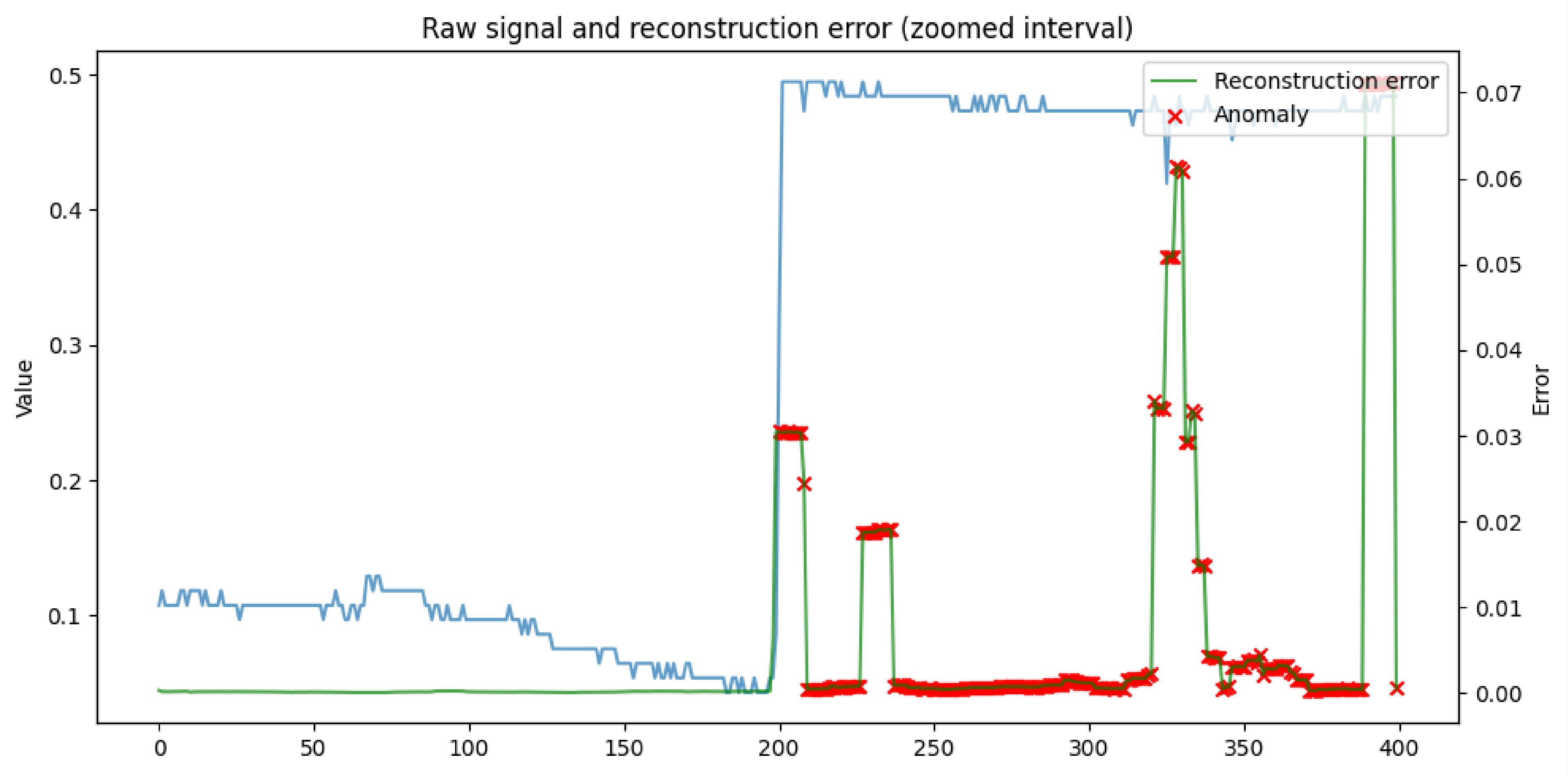
K	Model	Precision	Recall	F1	ROC-AUC
10	LSTM AE	0.185	0.984	0.312	0.872
10	Transformer (LN)	0.352	0.448	0.394	0.844
30	LSTM AE	0.184	1	0.31	0.893
30	Transformer (LN)	0.374	0.536	0.44	0.896
50	LSTM AE	0.212	0.957	0.347	0.896
50	Transformer (LN)	0.347	0.501	0.41	0.881

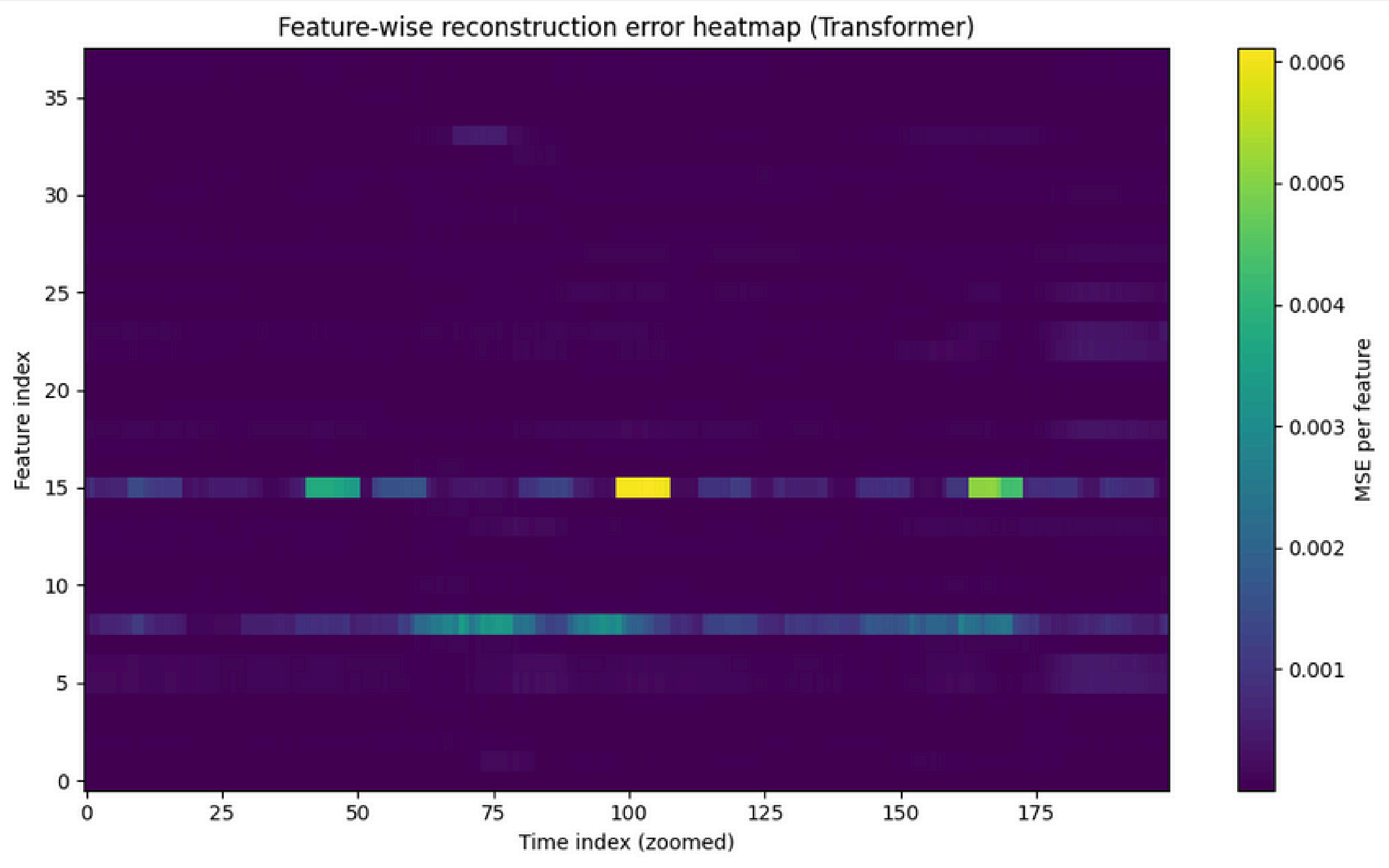
Increasing window size improves attention models up to K = 30.
Larger windows degrade performance without meta-learning, consistent with findings in TranAD

Qualitative Diagnostics

MODELS TESTED:

- Spikes align with anomalies
- False positives from high volatility
- Feature-level heatmaps show which KPIs drive anomalies (indices 20–33)
- These indices correspond mainly to disk/network throughput metrics





Multi-Machine Generalisation

PERFORMANCE WHEN TRAINING JOINTLY ON 4 MACHINES:

- machine-1-1: F1 = 0.405
- machine-1-2: F1 = 0.361
- machine-1-3: F1 = 0.062
- machine-2-1: F1 = 0.083

This confirms that the latent space is machine-specific; shared training collapses

Family	Best Model	F1 Score	Notes
Baseline	Mean Reconstruction	**0.52**	Surprisingly strong
LSTM	LSTM AE (K=10)	0.31	Very high recall, low prec.
Transformer	Transformer + Huber Loss	**0.49**	Best learned model
Transformer MSE	Transformer LN (K=30)	0.44	Best MSE-based transformer

Conclusion: strong cross-machine heterogeneity → per-machine conditioning needed

Comparison with State-of-the-Art & Final Takeaways

PERFORMANCE WHEN TRAINING JOINTLY ON 4 MACHINES:

- Strong EDA → solid choice of machine-1-1
- Pipeline fully aligned with KDD
- Huber transformer is best robust model
- Mean baseline sets high reference
- Feature-level diagnosis is a strength

Model	Window Size	Meta-Learning	Adversarial	Self-Conditioning	Training Scope	F1 Score
OmniAnomaly	100	No	No	No	28 machines	0.94
TranAD	100–500	Yes	Yes	Yes	28 machines	0.96
Transformer AE (Huber)	10	No	No	No	1 machine	0.49
Mean Baseline (ours)	10	No	No	No	1 machine	0.52

Our model is lightweight (1 machine, K=10, no meta-learning), suitable for real-time costrains. While SOTA methods rely on long context windows and architectural stabilisation.

Q&A

THANK YOU

AUTHORS:

EMANUELE ALBERTI
LEANDRO DUARTE
OTTAVIA BIAGI

UPM- UNIVERSIDAD POLITECNICA DE MADRID