Sensitivity-Aware Transformer for Robust Compressor Fault Detection in Natural Gas Pipelines

Khadija Shaheen*, Apoorva Chawla*, Ferdinand Evert Uilhoorn†, Pierluigi Salvo Rossi*,
*Dept. Electronic Systems, Norwegian University of Science and Technology, Trondheim, Norway
†Dept. Gas Engineering, Warsaw University of Technology, Warsaw, Poland
Email: shaheen.khadija@ntnu.no; apoorva.chawla@ntnu.no; ferdinand.uilhoorn@pw.edu.pl; salvorossi@ieee.org

***Abstract*—Compressor faults in natural gas pipelines pose significant risks to operational safety and efficiency. Traditional model-based fault detection methods, relying on residual analysis, often fail to detect system-wide bias anomalies. Although attention-based approaches, such as anomaly transformers, have shown promise for unsupervised anomaly detection, they struggle with noisy and heterogeneous sensor data, resulting in high false alarm rates and unreliable performance in real-world pipelines. To overcome these shortcomings, we propose the sensitivity-aware transformer (SA-Transformer), a robust extension of the anomaly transformer that is insensitive to noise. The framework is designed to detect subtle and abrupt compressor faults. The SA-Transformer features delta-aware attention to model the temporal transitions, effectively suppressing noise-induced variations and minimizing false alarms. Together with a temporal association module that compares attention-based patterns with learned prior using Kullback-Leibler (KL) divergence, the model exhibits better fault sensitivity and robustness under noisy application conditions. Experimental results on simulated pipeline data illustrate that the SA-Transformer reliably captures the fault onset and recovery, offering a reliable and data-efficient solution for anomaly detection in safety-critical pipeline scenarios.**

***Index Terms*—Compressor fault detection, anomaly transformer, time series analysis, natural gas pipelines.**

I. Introduction

Compressor stations play a crucial role in natural gas pipelines by maintaining pressure and ensuring continuous gas flow over long distances. However, unnoticed or late-diagnosed malfunctions can lead to reduced performance efficiency, increased maintenance costs, and potentially hazardous situations [1], [2]. Consequently, timely and reliable fault monitoring is essential to ensure the safety and efficiency of the pipeline network.

Conventional fault detection techniques for pipelines can be generally categorized into two groups: model-based and data-driven methods [3]–[6]. Model-based approaches rely on the physical system model and commonly utilize residual-based methods for detecting sensor faults in natural gas pipelines [7]–[9]. To enhance robustness against uncertainties in estimation-based approaches, hybrid frameworks have also been proposed which incorporate physical priors to alleviate the model sensitivity [10], [11]. However, these approaches often struggle to detect bias-type faults, such as those caused by progressive compressor deterioration, where sensor readings may remain internally consistent but collectively deviate from the actual system condition. While effective for identifying isolated sensor anomalies, these approaches may fail to detect such system-wide biases.

Conversely, the data-driven methods learn directly from the operational data to identify common patterns, without requiring sophisticated physical models, and offering greater sensitivity to system-level faults. Techniques such as autoencoders (AEs) [12], recurrent neural networks (RNNs) [13], and temporal convolutional networks (TCNs) [14], [15] have demonstrated the ability to learn complex temporal evolution for anomaly detection. Attention-based models, including the anomaly transformer, have further pushed the boundary of modeling latent temporal dependencies in an unsupervised manner [16], [17] . However, these models are primarily designed for detecting point anomalies and often struggle with structured or long-duration anomalies. Additionally, they are sensitive to noisy and non-stationary sensor data, prevalent in natural gas pipelines, leading to high false alarm rates and reduced reliability in real-world scenarios [18]–[20].

To address these limitations, we propose the sensitivity-aware transformer (SA-Transformer), a noise-resilient and fault-sensitive extension of the anomaly transformer [16], specifically designed for compressor fault detection in natural gas pipelines. SA-Transformer incorporates a novel delta-aware attention mechanism that effectively captures temporal dynamics by focusing on the relative changes across time steps, while suppressing noise-induced variations and reducing false alarms. To the best of authors' knowledge, the proposed work introduces a novel delta-aware attention for time-series anomaly detection for the first time. Furthermore, SA-Transformer integrates a temporal association module that computes the Kullback-Leibler (KL) divergence between the attention-based patterns and the learned prior, enhancing sensitivity to system-wide anomalies. Experimental results on simulated pipeline datasets demonstrate that the SA-Transformer effectively captures fault onset as well as recovery, offering a robust, data-efficient, and practical solution for anomaly detection in safety-critical pipeline systems.

II. Proposed Method: Sensitivity-Aware Transformer

SA-Transformer serves as an effective method for system-wide compressor fault detection in natural gas pipelines under high noise and stochastic fluctuations. It processes multivariate

[0]This work was partially supported by the Research Council of Norway under Project 311902 (SIGNIFY) within the framework IKTPLUSS.

2025 IEEE SENSORS | 979-8-3315-4467-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/SENSORS59705.2025.11330234

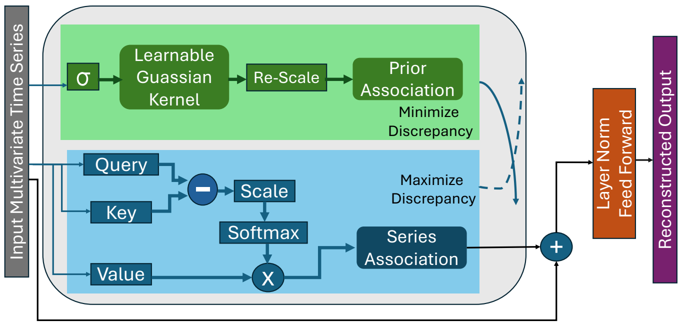


Fig. 1: Proposed SA-Transformer architecture

time series, with four core processing stages: (1) input normalization and standardization; (2) delta-aware attention, which discriminates among temporal transitions; (3) association-guided learning, which jointly trains attention and prior by minimizing a KL divergence to enhance the model adaption to anomalies; (4) input reconstruction and detection, which computes an anomaly score based on the difference from the expected output. A summary of these steps is given in Fig. 1.

Step 1: Input Time Series

Given a multivariate time series $\mathbf{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2 \ldots, \boldsymbol{x}_N] \in \mathbb{R}^{N \times d}$ with N being the number of time steps and d as the number of features (sensors), we standardize the sequence using the z-score transformation as $\tilde{\boldsymbol{x}}_t = \frac{\boldsymbol{x}_t - \mu}{\sigma}, \quad \forall t \in [1, N]$. Here, μ and σ denote the mean and standard deviation of the normalized series, respectively, computed over a sliding window or the full sequence.

Step 2: Delta-Aware Attention

Each normalized input window $\tilde{\mathbf{X}} \in \mathbb{R}^{w \times d}$, where w is the window length and d is the number of features, is linearly projected into query, key, and value matrices as

$$\mathbf{Q} = \tilde{\mathbf{X}}\mathbf{W}_Q, \quad \mathbf{K} = \tilde{\mathbf{X}}\mathbf{W}_K, \quad \mathbf{V} = \tilde{\mathbf{X}}\mathbf{W}_V. \tag{1}$$

Here, $\mathbf{Q} \in \mathbb{R}^{w \times d_a}$, $\mathbf{K} \in \mathbb{R}^{w \times d_a}$, and $\mathbf{V} \in \mathbb{R}^{w \times d_a}$ denote the query, key, and value matrices, respectively and $\mathbf{W}_Q \in \mathbb{R}^{d \times d_a}$, $\mathbf{W}_K \in \mathbb{R}^{d \times d_a}$, and $\mathbf{W}_V \in \mathbb{R}^{d \times d_a}$ are learnable weight matrices with d_a as the attention dimensionality. Further, the delta-aware attention focuses on the temporal difference between the embeddings. The attention scores α_{jk} and output representation $\boldsymbol{z}_j$ for jth query $\boldsymbol{q}_j$, based on kth key $\boldsymbol{k}_k$ and value $\boldsymbol{v}_k$ are calculated as

$$S_{jk} = -\frac{\|\boldsymbol{q}_j - \boldsymbol{k}_k\|^2}{\sqrt{d_a}}, \quad \forall j, k \in [1, w] \tag{2}$$

$$\alpha_{jk} = \frac{\exp(S_{jk})}{\sum_{k'} \exp(S_{jk'})}, \quad \boldsymbol{z}_j = \sum_k \alpha_{jk} \boldsymbol{v}_k \in \mathbb{R}^{d_a}, \tag{3}$$

The resulting output representation $\mathbf{Z} \in \mathbb{R}^{w \times d_a}$ captures the temporally weighted feature representations informed by relative differences between the time steps.

Step 3: Association Modeling

To facilitate the detection of subtle outliers, an association modeling mechanism is proposed in SA-Transformer that compares the learned attention distribution with a predefined prior. The intuition is that the temporal dependencies in time series are typically smooth and localized under normal conditions—meaning attention weights should focus around the current time step. To capture this expectation, we specify *prior* distribution $P_t \in \mathbb{R}^w$at each time step t based on a Gaussian kernel as

$$P_t(i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(i-t)^2}{2\sigma^2}\right), \quad i = 1, 2, \ldots, w. \tag{4}$$

Here, σ is the width of the distribution. This prior serves as a temporal anchor and forces the model to concentrate on adjacent time steps.

The association discrepancy is computed using the KL divergence between the learned attention distribution $\boldsymbol{S}_t$ and the *prior* distribution $\boldsymbol{P}_t$ as

$$\text{AssDis}_t = \text{KL}(\boldsymbol{S}_t \parallel \boldsymbol{P}_t) = \sum_{i=1}^{w} S_t(i) \log\left(\frac{S_t(i)}{P_t(i)}\right). \tag{5}$$

This difference measures the extent by which the model's attention deviates from the normal temporal behavior. The larger difference could be a potential anomaly.

To train the model, we use a *min-max optimization*. Here, we minimize the reconstruction and simultaneously maximize the association discrepancy of the anomalous inputs as

$$\mathcal{L} = \min_{\theta} \max_{P_t} \sum_{t=1}^{N} \left[\|\boldsymbol{x}_t - \hat{\boldsymbol{x}}_t\|^2 + \lambda \cdot \text{AssDis}_t\right]. \tag{6}$$

Here, $\hat{\boldsymbol{x}}_t$ is the reconstructed output, θ denotes the model parameter and λ is a hyper-parameter for trading off the reconstruction fidelity with the temporal difference.

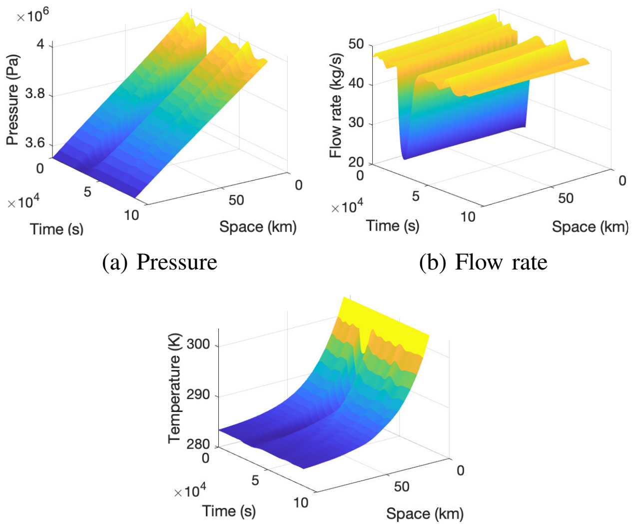
Step 4: Anomaly Scoring

At inference, the model scores the likelihood of anomalies at each work using the combination of reconstruction error and the divergence from the temporal prior. In particular, we calculate the anomaly score as

$$\text{Score}_t = \text{Softmax}(-\text{AssDis}_t) \cdot \|\boldsymbol{x}_t - \hat{\boldsymbol{x}}_t\|^2 \tag{7}$$

III. Experimental Results

To verify the performance of our SA-Transformer, we perform the numerical experiments over a high-pressure natural gas transmission pipeline equipped with distributed pressure, temperature, and flow sensors. The configuration of the simulation is similar to [21]. The simulation produces multivariate time series data over 63 sensor channels for operational data (three days) with an embedded fault event. At the compressor station, two similar reciprocating compressors driven by constant-speed high-voltage motors. A fault case is created where one of the compressors fails, the duration of time inactive is 2 hours. To evaluate robustness, we augment the dataset with different amounts of Gaussian noise. The



(a) Pressure (b) Flow rate

(c) Temperature

Fig. 2: Data with compressor fault.

TABLE I: Performance comparison across different anomaly detection methods

Method	Precision	Recall	F1-score
Isolation Forest [22]	0.1191	0.9800	0.2124
LSTM [27]	0.8909	0.4734	0.6183
AE [24]	0.8569	0.4570	0.5961
VAE [25]	0.8441	0.4502	0.5873
OmniAnomaly [26]	0.8851	0.4703	0.6143
Anomaly Transformer [28]	0.9231	1.00	0.96
SA-Transformer (Proposed)	**0.9636**	**1.00**	**0.9814**

simulation with the fault-injected and noise-corrupted signals is illustrated in Fig. 2.

We measure performance of anomaly detection according to the predefined metrics - precision, recall, and F1-score. We compare our SA-Transformer with a few state-of-the-art baselines: isolation forest [22], long short-term memory (LSTM) based sequence modeling [23], autoencoder (AE) [24], variational autoencoder (VAE) [25], omnianomaly [26], anomaly transformer [16]. As given in Table I, SA-Transformer achieves the best precision and F1-score among all baseline methods, indicating superior capability in correctly identifying the true anomalies while minimizing the false positives. This demonstrates its ability to detect faults in the presence of a large amount of noise and system-wide disturbances and its applicability for safety-critical monitoring of pipeline systems.

IV. Conclusions

We proposed the SA-Transformer specifically for compressor fault detection in high-pressure natural gas pipelines. By combining delta-aware attention with temporal association modeling, SA-Transformer enhanced sensitivity to system-level anomalies while maintaining robustness against high noise levels. Experiments on synthesized multivariate time series data depicted that, compared to the baseline methods, SA-Transformer significantly reduced false positives. This reduction in false alarms is crucial for practical applications, where random noise is ubiquitous and excessive false positives can critically compromise the trust and operational efficiency. Future work will focus on extending the model to support real-time streaming sensor data, scaling to more complex multi-pipeline infrastructures and integration of uncertainty quantification methods to ensure reliable and informed decision making under dynamic and safety critical scenarios.

References

[1] K. Mathioudakis, A. Alexiou, N. Aretakis, and C. Romesis, "Signatures of compressor and turbine faults in gas turbine performance diagnostics: A review," *Energies*, vol. 17, no. 14, p. 3409, 2024.

[2] Q. Chen, L. Zuo, C. Wu, Y. Li, K. Hua, M. Mehrtash, and Y. Cao, "Optimization of compressor standby schemes for gas transmission pipeline systems based on gas delivery reliability," *Reliability Engineering & System Safety*, vol. 221, p. 108351, 2022.

[3] K. Shaheen, A. Chawla, F. E. Uilhoorn, and P. S. Rossi, "Partial-distributed architecture for multi-sensor fault detection, isolation and accommodation in hydrogen-blended natural gas pipelines," *IEEE Internet of Things Journal*, 2024.

[4] A. Chawla, Y. Arellano, M. V. Johansson, H. Darvishi, K. Shaneen, M. Vitali, F. Finotti, and P. Salvo Rossi, "IoT-based monitoring in carbon capture and storage systems," *IEEE Internet of Things Magazine*, vol. 5, no. 4, pp. 106–111, 2022.

[5] K. Shaheen, A. Chawla, F. E. Uilhoorn, and P. S. Rossi, "Partial-distributed filtering for fault detection, isolation and accommodation in natural-gas pipelines," in *2024 27th International Conference on Information Fusion (FUSION)*. IEEE, 2024, pp. 1–8.

[6] F. Mehmood, P. M. Papadopoulos, L. Hadjidemetriou, A. Charalambous, and M. M. Polycarpou, "Model-based fault diagnosis scheme for current and voltage sensors in grid side converters," *IEEE Transactions on Power Electronics*, vol. 38, no. 4, pp. 5360–5375, 2022.

[7] K. Shaheen, A. Chawla, F. E. Uilhoorn, and P. S. Rossi, "Model-based sensor-fault detection and isolation in natural-gas pipelines for transient flow," in *2023 IEEE SENSORS*. IEEE, 2023, pp. 1–4.

[8] F. E. Uilhoorn, "State-space estimation with a bayesian filter in a coupled pde system for transient gas flows," *Applied Mathematical Modelling*, vol. 39, no. 2, pp. 682–692, 2015.

[9] K. Shaheen, A. Chawla, F. E. Uilhoorn, and P. S. Rossi, "Sensor-fault detection, isolation and accommodation for natural-gas pipelines under transient flow," *IEEE Transactions on Signal and Information Processing over Networks*, 2024.

[10] H. Nordal and I. El-Thalji, "Assessing the technical specifications of predictive maintenance: A case study of centrifugal compressor," *Applied Sciences*, vol. 11, no. 4, p. 1527, 2021.

[11] K. Shaheen, A. Chawla, F. E. Uilhoorn, and P. S. Rossi, "Partial-distributed particle filter for multi-sensor fault diagnosis in carbon dioxide pipelines," *IEEE Internet of Things Journal*, 2025.

[12] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Lstm-based encoder-decoder for multivariate time series anomaly detection," in *ICML Workshop on Anomaly Detection*, 2016.

[13] K. Hundman, V. Constantinou, I. Laporte, J. Colwell, and T. Soderstrom, "Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 387–395.

[14] Y. Zhao and et al., "Multivariate time-series anomaly detection via dynamic graph forecasting," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[15] W. K. F. Al-Tekreeti, K. R. Kashyzadeh, and S. Ghorbani, "Advancements in gas turbine fault detection: A machine learning approach based on the temporal convolutional network–autoencoder model," *Applied Sciences*, vol. 14, no. 11, p. 4551, 2024.

[16] H. Xu, Y. Chen, D. Wang, and G. Long, "Anomaly transformer: Time series anomaly detection with association discrepancy," in *International Conference on Learning Representations (ICLR)*, 2022.

[17] A. Z. Sellam, I. Benaissa, A. Taleb-Ahmed, L. Patrono, and C. Distante, "Maat: Mamba adaptive anomaly transformer with association discrepancy for time series," *Engineering Applications of Artificial Intelligence (preprint on arXiv/SSRN)*, 2025, arXiv:2502.07858 / SSRN 5177260.

[18] X. Zhang, M. Li, and L. Chen, "Mfdna: A multiscale feature-dilated network with adversarial training for noisy anomaly detection," *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

[19] K. Shaheen, A. Chawla, F. E. Uilhoorn, and P. S. Rossi, "Hybrid technique for sensor fault diagnosis in natural-gas pipelines," in *2024 IEEE SENSORS*. IEEE, 2024, pp. 1–4.

[20] C. Shimillas, K. Malialis, K. Fokianos, and M. M. Polycarpou, "Transformer-based multivariate time series anomaly localization," *arXiv preprint*, 2025, arXiv:2501.08628.

[21] K. Shaheen, A. Chawla, F. E. Uilhoorn, and P. S. Rossi, "Model-based architecture for multi-sensor fault detection, isolation and accommodation in natural-gas pipelines," *IEEE Sensors Journal*, 2023.

[22] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 Eighth IEEE International Conference on Data Mining*. IEEE, 2008, pp. 413–422.

[23] S. Lin, R. Clark, R. Birke, S. Schönborn, N. Trigoni, and S. Roberts, "Anomaly detection for time series using vae-lstm hybrid model," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Ieee, 2020, pp. 4322–4326.

[24] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[25] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[26] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, J. Pei, and B. Zhang, "Omnianomaly: Detecting anomalies in multivariate time series with variational autoencoders," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, 2019, pp. 4880–4886.

[27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[28] H. Xu, H. Zhao, Z. Li, L. Yuan, H. Zha, and S. Zhang, "Anomaly transformer: Time series anomaly detection with association discrepancy," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.