Anomaly Detection via Re-encoding in Autoencoder-Based Compression for Time Series Monitoring Applications

Andriy Enttsel *, Aldo Sean Sartor *, Alex Marchioni *, Gianluca Setti ‡†, Riccardo Rovatti *†, Mauro Mangia *†
*DEI, †ARCES, University of Bologna, Italy, ‡CEMSE, KAUST, Saudi Arabia; contact author: andriy.enttsel@unibo.it

***Abstract*—Monitoring systems generate and transmit large volumes of data to processing facilities capable of performing multiple tasks. To reduce transmission and storage costs, data is often compressed, with autoencoders (AEs) emerging as a promising neural network-based approach. This work considers a scenario where the receiver is responsible for both reconstruction and anomaly detection. We propose a novel anomaly detector that operates on the receiver side, approximating the standard anomaly score of conventional AE-based detectors. The proposed approach requires no fine-tuning, as the compression process itself ensures strong detection performance. Moreover, its performance can be further enhanced through a common regularization technique. We validate our method through experiments on two distinct time series datasets: ECG signals and acceleration data.**

***Index Terms*—anomaly detection, dimensionality reduction, autoencoder, information theory, compression**

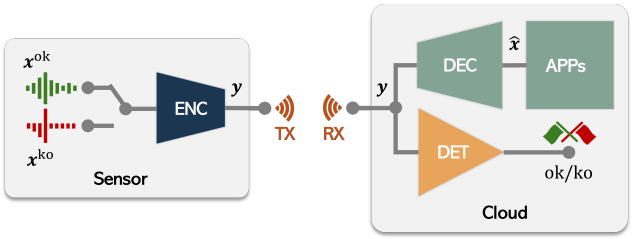

Fig. 1: Block diagram of the system with an encoder (ENC) compressing a signal and a decoder (DEC) recovering it for further processing. The signal, either normal ($\boldsymbol{x}^{\mathrm{ok}}$) or anomalous ($\boldsymbol{x}^{\mathrm{ko}}$), is classified by a detector (DET) at the receiver.

I. Introduction

In many acquisition systems, sensors collect data that is later processed by remote devices to extract essential information. To meet energy and bandwidth limitations, sensors often employ lossy data compression prior to transmission. This approach allows for higher compression ratios compared to lossless methods, though it introduces the risk of information loss, which can negatively impact subsequent data processing.

Among the various lossy compression techniques, autoencoder-based dimensionality reduction (AE) [1] is a widely used method. The AE is a neural network that consists of two sub-networks: an encoder, which reduces the dimensionality of the input, and a decoder, which reconstructs it. These networks are trained together to minimize distortion, which is the difference between the reconstruction and the original input. AEs may be asymmetric, entailing a lightweight encoding to meet the sensor's computational constraints, alongside a potentially costly decoding stage executed in the cloud [2].

However, reconstruction may not be the only critical task. In fields such as Structural Health Monitoring [3], Condition Monitoring [4], and Healthcare Monitoring [5], acquisition systems collect vast amounts of data to assess the health of monitored systems. A common technique for this purpose is anomaly detection (AD) [6], which aims to identify abnormal behaviors and distinguish them from normal conditions.

While the effects of compression on reconstruction have been extensively studied since the advent of rate-distortion theory [7, Chapter 8], the trade-off between compression and AD has only recently gained attention, as discussed in [2]. The authors demonstrated that detection performance rapidly diminishes as distortion increases when dimensionality reduction is optimized for sole reconstruction.

Despite this, AE-based dimensionality reduction remains a popular tool for promoting AD. In [8], a Deep Support Vector Data Description (Deep SVDD) method was proposed, which leverages an AE trained to minimize reconstruction error. The encoder is then fine-tuned to minimize the volume of the latent representation and serves as the detector. Alternatively, [9] integrates latent volume optimization as a regularization term during the AE training process, leading to a Shrink AE (SAE) that does not require further fine-tuning of the encoder. A similar technique is also proposed in [10].

However, these approaches focus on AEs designed specifically for AD, neglecting reconstruction quality. An AE that balances both tasks is presented in [11], where the AE is trained with a regularization term that controls the differential entropy of the latent representation. This regularization helps the AE prioritize features valuable for AD, even if it slightly reduces reconstruction performance. The trade-off between reconstruction and anomaly detectability in entropy-regularized AE and SAE was further explored in [12].

In this paper, we extend the research in [12] by proposing a novel anomaly detection score, based on the self-assessment technique, which is commonly used in compressed sensing to evaluate reconstruction quality [5], [13]. After reconstructing the received measurement, the self-assessment involves re-projecting the reconstructed instance back into the measurement space and comparing it to the received signal. In

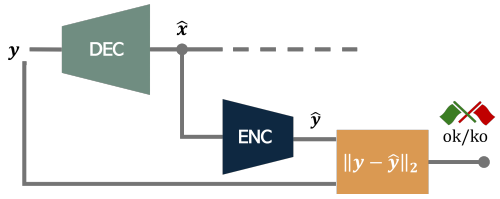

Fig. 2: Block diagram of the detector, where the self-assessment re-encodes the reconstructed signal $\hat{\boldsymbol{x}}$ to obtain $\hat{\boldsymbol{y}}$, used to compute the anomaly score $\|\boldsymbol{y} - \hat{\boldsymbol{y}}\|_2$.

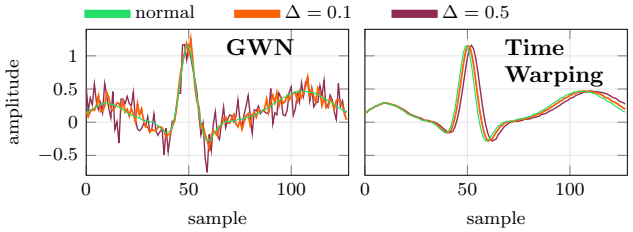

Fig. 3: Examples of GWN and Time Warping anomalies with two values of Δ injected into a window of the ECG signal.

compression scenarios where the original vector is unavailable at the receiver, this approach has been shown to effectively approximate the classical mean squared error between the reconstructed vector and the original. We use this method to approximate the standard anomaly score for an AE-based detector, proposing a detector that requires no fine-tuning, as the compressor alone ensures good detection performance. Additionally, we show that the performance of this score can be further enhanced with a detection-oriented regularization.

We assess the effectiveness of the proposed approach on two different time series: Electrocardiogram signals coming from a Health Monitoring application and accelerometers' waveforms used for Structural Health Monitoring.

The paper is organized as follows: Section II presents the mathematical models; Section III takes two use cases and describes the datasets, model architecture, and numerical evidence. Finally, the conclusions are drawn.

II. Mathematical Models

We consider a system, illustrated in Fig. 1, where a sensor captures an n-dimensional signal $\boldsymbol{x}$ and compresses it into a k-dimensional representation $\boldsymbol{y} = \text{ENC}(\boldsymbol{x})$, where $k < n$. The compressed signal is then transmitted and reconstructed at the receiver using a decoder, yielding $\hat{\boldsymbol{x}} = \text{DEC}(\boldsymbol{y})$. In parallel, a detector (DET) determines whether the signal is normal ($\boldsymbol{x}^{\text{ok}} \sim f_{\boldsymbol{x}}^{\text{ok}}$) or anomalous ($\boldsymbol{x}^{\text{ko}} \sim f_{\boldsymbol{x}}^{\text{ko}}$).

When employing autoencoder (AE)-based compression, the encoder ENC and decoder DEC are neural networks. Since anomalies are rare and typically unknown during training, the AE is trained exclusively on normal signals, i.e., $\boldsymbol{x} = \boldsymbol{x}^{\text{ok}}$. Reconstruction quality is commonly estimated using the Mean Squared Error (MSE).

Minimizing MSE as a loss function results in an AE optimized solely for reconstruction, potentially discarding features of $\boldsymbol{x}$ that are crucial for anomaly detection [2]. To mitigate this issue, the authors in [11] proposed incorporating a constraint on the differential entropy of $\boldsymbol{y}^{\text{ok}}$, denoted as $\mathcal{H}(\boldsymbol{y}^{\text{ok}})$. Under the assumption that the latent representation follows an isotropic Gaussian distribution of zero mean, minimizing $\hat{\mathcal{H}}(\boldsymbol{y}^{\text{ok}})$ is equivalent to minimizing the expected squared norm of $\boldsymbol{y}^{\text{ok}}$ [14]. Consequently, to ensure the AE retains information relevant for anomaly detection at the receiver, we adopt the following loss function:

$$\mathcal{L}_{k,\lambda}(\boldsymbol{x}^{\text{ok}}) = \frac{1}{nN}\sum_{i=0}^{N-1}\left\|\boldsymbol{x}_i^{\text{ok}} - \hat{\boldsymbol{x}}_i^{\text{ok}}\right\|^2 + \frac{\lambda}{kN}\sum_{i=0}^{N-1}\|\boldsymbol{y}_i^{\text{ok}}\|^2, \quad (1)$$

where $\boldsymbol{x}_i^{\text{ok}}$, $\hat{\boldsymbol{x}}_i^{\text{ok}}$ and $\boldsymbol{y}_i^{\text{ok}}$ represent one of the N instances of $\boldsymbol{x}^{\text{ok}}$, $\hat{\boldsymbol{x}}^{\text{ok}}$ and $\boldsymbol{y}^{\text{ok}}$, respectively.

It is worth noting that increasing λ improves anomaly detection performance at the cost of reduced reconstruction accuracy. However, enhancing anomaly detection is not as simple as arbitrarily increasing λ; instead, the two terms in (1) must be balanced to achieve an optimal trade-off [9], [11].

When performing AD at the receiver, a straightforward approach is first to reconstruct the signal, $\hat{\boldsymbol{x}}$, and then apply a standard anomaly detection algorithm. From an information-theoretic perspective, this is equivalent to detecting anomalies directly in the compressed domain, i.e., using $\boldsymbol{y}$ [2]. Operating in the compressed domain has a key advantage: the reduced dimensionality of $\boldsymbol{y}$ compared to the original signal mitigates the curse of dimensionality, enhancing the scalability of the detection process. In this work, we propose a hybrid approach that leverages both reconstruction and compressed-domain analysis. Since signal reconstruction is already required, a natural anomaly score would be the reconstruction error, $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2$. However, this requires access to $\boldsymbol{x}$, which is unavailable at the receiver. To address this, we adopt the self-assessment principle and define the anomaly score as:

$$s(\boldsymbol{y}) = \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|_2 = \|\boldsymbol{y} - \text{ENC}\left(\text{DEC}\left(\boldsymbol{y}\right)\right)\|_2 . \quad (2)$$

Fig. 2 summarizes the detection process, which relies on the compressed representation $\hat{\boldsymbol{y}}$ of the reconstructed signal $\hat{\boldsymbol{x}}$ to eliminate the need for the original signal $\boldsymbol{x}$. In the presence of an anomaly, the decoder struggles to accurately reconstruct $\boldsymbol{x}$, leading to a significant discrepancy between $\hat{\boldsymbol{y}}$ and $\boldsymbol{y}$, which serves as a reliable indicator of abnormality.

III. Numerical Setup

We assess the effectiveness of the proposed approach on two different datasets: an Electrocardiogram (ECG) trace coming from a Health Monitoring application and accelerometer (ACC) signal used for Structural Health Monitoring.

A. ECG

Synthetic ECG signals were generated following [12] with a sampling rate of 256 sps, and 35 dB of injected white noise. For training and validation, we generate 2×10^6 windows, each containing $n = 128$ samples. An additional 10^4 windows are created to evaluate performance.

Regarding anomalies, prior research [15], [16] has shown that the lack of real-world anomalies can be mitigated by injecting synthetic perturbations into normal signals. Following

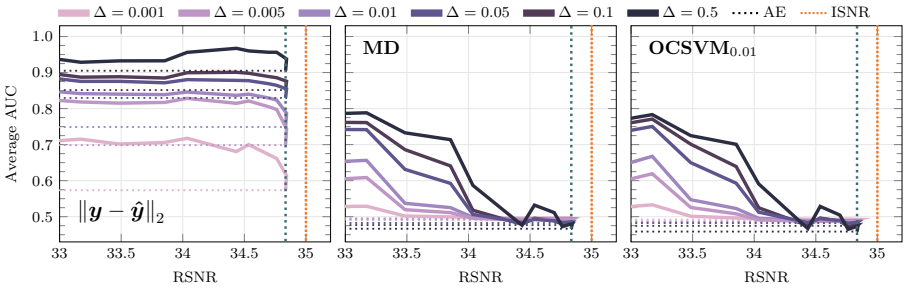

Fig. 4: Anomaly detection-reconstruction trade-off for ECG signals represented in terms of AUC (averaged over 12 anomalies) vs. RSNR curves for different levels of anomaly intensity Δ.

[16], we consider 12 types of perturbations corresponding either to system faults, such as Constant, Impulse, and Gaussian white noise (GWN), or to changes in the sensed phenomenon, such as Time Warping and Principal Subspace Alteration.

Anomalies are injected into each window of the normal test set with an intensity measured by the *deviation* metric [17]:

$$\Delta = \frac{1}{n}\mathbf{E}\left[\left\|\boldsymbol{x}^{\mathrm{ok}} - \boldsymbol{x}^{\mathrm{ko}}\right\|^2\right]. \tag{3}$$

Examples of two types of anomalies are shown in Fig. 3, comparing the original signal with anomalous signals at two different deviation levels Δ. As expected, higher values of Δ result in more pronounced anomalies.

The AE architecture is inherited from [12] and follows an asymmetric design. Specifically, ENC applies a simple linear transformation, mapping an $n = 128$ dimensional input to a $k = 16$ dimensional output, achieving a compression ratio (CR) of 8. The decoder DEC processes the encoder's output with three hidden layers with $2n$, $2n$, and n units, respectively, using ReLU activations, followed by a linear output layer.

AEs are trained using the Adam optimizer [18] with a batch size of 128 and an initial learning rate of 0.001, which decreases whenever the loss reaches a plateau, with a patience of 20 epochs. Reconstruction performance is evaluated using the Reconstruction Signal-to-Noise Ratio (RSNR):

$$\mathrm{RSNR} = \mathbf{E}\left[\frac{\|\boldsymbol{x}^{\mathrm{ok}}\|_2}{\|\boldsymbol{x}^{\mathrm{ok}} - \hat{\boldsymbol{x}}^{\mathrm{ok}}\|_2}\right]_{\mathrm{dB}}. \tag{4}$$

We compare the performance of our detector with two commonly used unsupervised detectors applied to $\boldsymbol{y}$, the same as those analyzed in [12]:

- Mahalanobis distance (MD) [19]: Computes the anomaly score as the squared distance of $\boldsymbol{y}$ from the mean vector, with each component weighted by its variance.
- OCSVM$_\nu$ [20]: Uses a kernel function to map $\boldsymbol{y}$ into a higher-dimensional feature space, where a hyperplane separates normal and anomalous samples. The score is given by the distance of $\boldsymbol{y}$ from the origin in this feature space. The parameter ν controls the fraction of training vectors selected as support vectors.

Detection performance is ultimately evaluated using the Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC) [21].

Figs. 4 and 5 illustrate the detection-reconstruction trade-off for the ECG signal by varying λ in (1). Fig. 4 shows the AUC, averaged over 12 anomalies, as a function of reconstruction performance. Each subplot corresponds to a different detector, with colors indicating anomaly intensities. The orange vertical dotted line represents the intrinsic SNR due to noise, while the dark green line highlights that the standard AE ($\lambda = 0$) achieves $34.8\,\mathrm{dB}$, closely matching the intrinsic SNR. Horizontal dotted lines indicate the detection performance when $\lambda = 0$.

In the two rightmost subplots, both MD and OCSVM fail to exploit latent space information to distinguish normal from anomalous signals at $\lambda = 0$. Regularization improves their performance but reduces reconstruction accuracy, with the best detection achieved at $33.17\,\mathrm{dB}$ for both detectors. Notably, at $\Delta = 0.5$, the AUC increases from the initial 0.46 for both MD and OCSVM to 0.79 and 0.78, respectively. Conversely, in the leftmost plot, the proposed detector already exhibits strong performance at $\lambda = 0$, which can be further improved at the expense of a minor $0.5\,\mathrm{dB}$ reduction in RSNR. For instance, at $\Delta = 0.5$, the AUC increases from 0.91 to 0.97.

Fig. 5 provides a more detailed analysis by displaying the trade-off for each anomaly separately. The red line represents the performance of a random detector, while the other two colors correspond to two different anomaly intensities. The solid line denotes the performance of the proposed detector, whereas the dashed line represents OCSVM. The densely dotted and dotted lines indicate the performance of OCSVM and the proposed detector, respectively, at $\lambda = 0$.

The results demonstrate that the proposed detector consistently outperforms OCSVM across all anomalies, both with and without regularization. For instance, in the case of particularly challenging Mixing with Constant and Time Warping anomalies, the proposed detector benefits from regularization, whereas OCSVM performs worse than a random detector.

B. ACC

The proposed method is also applied to the structural assessment of Bridge S101 in Vienna [22], [23]. The bridge has been structurally characterized by identifying a set of natural oscillation frequencies as key indicators of its integrity. The structure was equipped with accelerometer sensors distributed along its length, capturing vibration signals that reveal modal

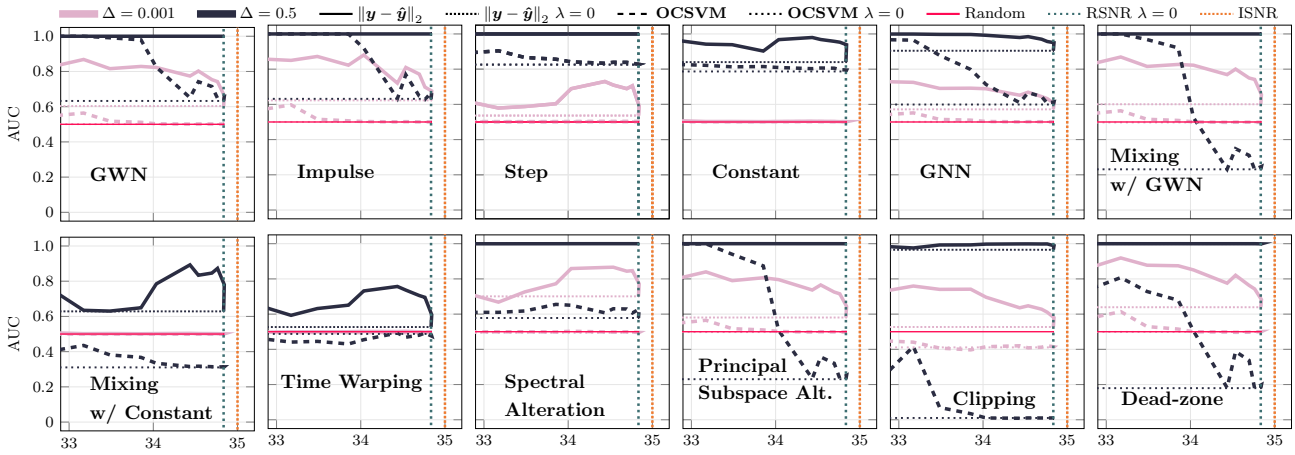

Fig. 5: Anomaly detection-reconstruction trade-off for each ECG data anomaly in terms of AUC vs. RSNR curves for two levels of anomaly intensity Δ.

shapes. We consider a scenario where this data must be compressed before transmission.

We focus on a single vertical acceleration channel with both normal (7.6×10^6 samples at $f_s = 500$ Hz) and anomalous (2.64×10^6 samples) data. The data were decimated to $f_s = 50$ Hz and split into batches of 20 consecutive $n = 256$-dimensional windows, randomly and equally divided for training and testing.

The compressor follows the same architecture (with $n = 256$ and $k = 32$) and training strategy (with $\lambda = 0$) as the one in the previous subsection. The training set is augmented with 144×10^3 overlapping windows of length n.

The primary objective is to determine whether the spectral characteristics of the data have changed since a common indicator of structural degradation is a variation in modal frequencies. In our case the modal frequencies $\{4.05, 6.3, 9.69, 13.29, 15.93\}$ are expected to undergo a maximum $\pm 7.5\%$ variation [22].

In this scenario, anomaly detection can be based on the Itakura–Saito spectral distance (ISSD) [24], [3]. Specifically, we employ the symmetric ISSD (SISSD) anomaly score [4] defined as:

$$\text{SISSD} = \frac{1}{M}\sum_{k=0}^{M-1}\left(\frac{\hat{S}^R(f_k)}{\hat{S}(f_k)} - \log\frac{\hat{S}^R(f_k)}{\hat{S}(f_k)} + \frac{\hat{S}(f_k)}{\hat{S}^R(f_k)} - \log\frac{\hat{S}(f_k)}{\hat{S}^R(f_k)} - 2\right), \tag{5}$$

where M is the number of frequency points, $\hat{S}^R(f)$ is the reference PSD computed offline under normal operating conditions, while $\hat{S}(f)$ is computed online and compared with $\hat{S}^R(f)$ to detect deviations from the healthy state.

Following [4], since anomalies are expected to affect specific frequency bands rather than the entire spectrum, we also employ the multivariate SISSD (MSISSD) that computes

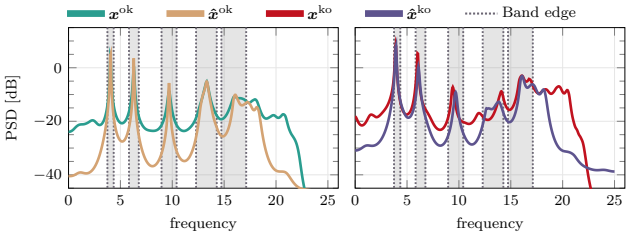

Fig. 6: Power spectral densities for original and reconstructed normal and anomalous signals. The shaded areas highlight the bands of interest.

SISSD over B spectral bands $S_b(f)$, $b = 0, \ldots, B-1$ of the power spectral density (PSD) $S(f)$, and is defined as:

$$\text{MSISSD} = \begin{bmatrix}\text{SISSD}_0 & \text{SISSD}_2 & \ldots & \text{SISSD}_{B-1}\end{bmatrix}. \tag{6}$$

Additionally, we compute the average SISSD across the same bands:

$$\overline{\text{MSISSD}} = \frac{1}{B}\sum_{b=0}^{B-1}\text{SISSD}_b. \tag{7}$$

The PSD is estimated using the Burg method [25] over K consecutive n-dimensional windows of the reconstructed signal $\hat{\boldsymbol{x}}$. $\hat{S}^R(f)$ is estimated from the training set.

This methodology serves as a baseline for our approach, where the reconstructed signal $\hat{\boldsymbol{y}}$ is first computed for each n-dimensional window out of K. The anomaly score is then obtained by evaluating the discrepancy between the concatenated signal $\boldsymbol{Y} = [\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{K-1}]$ and its reconstruction $\hat{\boldsymbol{Y}} = [\hat{\boldsymbol{y}}_0, \ldots, \hat{\boldsymbol{y}}_{K-1}]$.

Fig. 6 presents the PSD averaged over all windows for both normal and anomalous conditions, comparing the original and reconstructed signals. The spectral bands of interest are highlighted in gray. Despite a compression ratio of CR = 8, the autoencoder effectively preserves most relevant spectral information while reducing noise.

TABLE I: Detection performance in terms of AUC for different detectors across different values K of processed instances.

K	$\left\|\boldsymbol{Y}-\hat{\boldsymbol{Y}}\right\|_2$	SSISD_0	SISSD_1	SISSD_2	SISSD_3	SISSD_4	$\overline{\text{MSISSD}}$	SISSD
1	0.79	0.34	0.69	0.47	0.66	0.45	0.51	0.75
2	0.81	0.49	0.65	0.44	0.65	0.51	0.57	0.67
3	0.83	0.60	0.69	0.48	0.68	0.58	0.65	0.76
4	0.85	0.65	0.79	0.53	0.70	0.66	0.72	0.81
5	0.87	0.68	0.82	0.56	0.70	0.69	0.76	0.85
10	0.92	0.81	0.86	0.69	0.78	0.82	0.88	0.93
15	0.96	0.88	0.88	0.77	0.82	0.90	0.94	0.96
20	0.96	0.91	0.93	0.83	0.84	0.93	0.97	0.97

Detection performance results are summarized in Table I. Our detector is compared with the detectors based on SISSD for different values of K. In all cases, increasing K leads to improved detection performance, except for SSISD at $K = 2$. Even for $K = 1$, the proposed approach achieves strong performance and consistently outperforms the other methods up to $K = 5$. For larger delays, the performance is comparable to SISSD. Overall, these results suggest that the proposed approach is more sensitive to anomalies and well-suited for prompt detection.

IV. Conclusion

This work considers the task of anomaly detection downstream of signal compression. We propose a novel anomaly detection method based on re-encoding within an autoencoder compression framework, where self-assessment enables anomaly detection without requiring access to the original data. Experimental results confirm its superiority over other conventional methods across different applications. For ECG signals, our detector achieves robust anomaly detection while preserving reconstruction quality, unlike Mahalanobis Distance and One-Class SVM, which require greater compromises in fidelity to improve detection. In structural health monitoring, our approach outperforms SSISD and $\overline{\text{MSISSD}}$ while eliminating the need for spectrum computation or frequency band selection, greatly simplifying deployment.

V. Acknowledgment

This study was carried out within the FAIR - Future Artificial Intelligence Research and received funding from the European Union Next-Generation EU (Piano Nazionale di Ripresa e Resilienza (PNRR) –Missione 4 Componente 2, Investimento 1.3 – D.D. 1555 11/10/2022, PE00000013). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

References

[1] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," *Aiche Journal*, vol. 37, pp. 233–243, 1991.

[2] A. Marchioni, A. Enttsel, M. Mangia, R. Rovatti, and G. Setti, "Anomaly detection based on compressed data: An information theoretic characterization," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 54, no. 1, pp. 23–38, 2024.

[3] F. Zonzini, V. Dertimanis, E. Chatzi, and L. D. Marchi, "System identification at the extreme edge for network load reduction in vibration-based monitoring," *IEEE Internet of Things Journal*, vol. 9, no. 20, pp. 20 467–20 478, 2022.

[4] R. Diversi and N. Speciale, "A multidimensional health indicator based on autoregressive power spectral density for machine condition monitoring," *Sensors*, vol. 24, no. 15, 2024.

[5] M. Mangia, L. Prono, A. Marchioni, F. Pareschi, R. Rovatti, and G. Setti, "Deep neural oracles for short-window optimized compressed sensing of biosignals," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 14, no. 3, pp. 545–557, 2020.

[6] C. C. Aggarwal, *Outlier Analysis*. Springer Cham, 2017.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.

[8] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, "Deep one-class classification," in *Proceedings of the 35th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. PMLR, 10–15 Jul 2018, pp. 4393–4402.

[9] V. L. Cao, M. Nicolau, and J. McDermott, "Learning neural representations for network anomaly detection," *IEEE Transactions on Cybernetics*, vol. 49, no. 8, pp. 3074–3087, 2019.

[10] A. Ghafourian, H. Shui, D. Upadhyay, R. Gupta, D. Filev, and I. Soltani, "Targeted collapse regularized autoencoder for anomaly detection: Black hole at the center," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–11, 2024.

[11] A. Enttsel, A. Marchioni, G. Setti, M. Mangia, and R. Rovatti, "Enhancing anomaly detection with entropy regularization in autoencoder-based lightweight compression," in *2024 IEEE 6yh International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, 2024, *accepted*.

[12] A. Enttsel, A. Marchioni, L. Manovi, R. Nikpali, G. Ravaglia, G. Setti, R. Rovatti, and M. Mangia, "Anomaly detection-reconstruction trade-off in autoencoder-based compression," in *2024 32nd European Signal Processing Conference (EUSIPCO)*, 2024, pp. 1957–1961.

[13] F. Martinini, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "Compressed sensing inspired neural decoder for undersampled mri with self-assessment," in *2021 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2021, pp. 01–06.

[14] L. Ruff, R. A. Vandermeulen, N. Görnitz, A. Binder, E. Müller, K.-R. Müller, and M. Kloft, "Deep semi-supervised anomaly detection," in *International Conference on Learning Representations*, 2020.

[15] L. Ruff, J. R. Kauffmann, R. A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T. G. Dietterich, and K.-R. Müller, "A unifying review of deep and shallow anomaly detection," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 756–795, 2021.

[16] A. Enttsel, S. Onofri, A. Marchioni, M. Mangia, G. Setti, and R. Rovatti, "A general framework for the assessment of detectors of anomalies in time series," *IEEE Transactions on Industrial Informatics*, 2024.

[17] A. Enttsel, F. Martinini, A. Marchioni, M. Mangia, R. Rovatti, and G. Setti, "Second-order statistic deviation to model anomalies in the design of unsupervised detectors," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015.

[19] R. De Maesschalck, D. Jouan-Rimbaud, and D. Massart, "The mahalanobis distance," *Chemometrics and Intelligent Laboratory Systems*, vol. 50, no. 1, pp. 1–18, 2000.

[20] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," in *Proceedings of the 12th International Conference on Neural Information Processing Systems*, ser. NIPS'99. Cambridge, MA, USA: MIT Press, 1999, p. 582–588.

[21] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006, rOC Analysis in Pattern Recognition.

[22] V. C. Engineers, "Progressive damage test s101 flyover reibesdorf," *Rep. No. 08*, vol. 2308, 2009.

[23] D. M. Siringoringo, Y. Fujino, and T. Nagayama, "Dynamic characteristics of an overpass bridge in a full-scale destructive test," *Journal of Engineering Mechanics*, vol. 139, no. 6, pp. 691–701, 2013.

[24] C. Magnant, E. Grivel, A. Giremus, L. Ratton, and B. Joseph, "Classifying autoregressive models using dissimilarity measures: A comparative study," in *2015 23rd European Signal Processing Conference (EUSIPCO)*, 2015, pp. 998–1002.

[25] S. Kay, *Modern Spectral Estimation: Theory and Application*, ser. Prentice-Hall Signal Processing Series: Advanced monographs. Prentice Hall, 1988.