Time Series Anomaly Detection Leveraging MSE Feedback with AutoEncoder and RNN

Ibrahim Delibasoglu
Department of Computer and Information Science (IDA), Linköping University, Sweden
Software Engineering, Sakarya University, Turkey

Fredrik Heintz
Department of Computer and Information Science (IDA), Linköping University, Sweden

Abstract

Anomaly detection in time series data is a critical task in various domains, including finance, healthcare, cybersecurity and industry. Traditional methods, such as time series decomposition, clustering, and density estimation, have provided robust solutions for identifying anomalies that exhibit distinct patterns or significant deviations from normal data distributions. Recent advancements in machine learning and deep learning have further enhanced these capabilities. This paper introduces a novel method for anomaly detection that combines the strengths of autoencoders and recurrent neural networks (RNNs) with an reconstruction error feedback mechanism based on Mean Squared Error. We compare our method against classical techniques and recent approaches like OmniAnomaly, which leverages stochastic recurrent neural networks, and the Anomaly Transformer, which introduces association discrepancy to capture long-range dependencies and DCDetector using contrastive representation learning with multi-scale dual attention. Experimental results demonstrate that our method achieves superior overall performance in terms of precision, recall, and F1 score. The source code is available at `http://github.com/mribrahim/AE-FAR`

2012 ACM Subject Classification Computing methodologies → Machine learning algorithms

Keywords and phrases Time series, Anomaly, Neural networks

Digital Object Identifier 10.4230/LIPIcs.TIME.2024.17

Supplementary Material *Software*: `https://github.com/mribrahim/AE-FAR/`

1 Introduction

Time series data, characterized by its sequential and temporal nature, is extensively utilized across numerous applications, including finance, healthcare, manufacturing, and environmental monitoring. Detecting anomalies within time series data is a critical task to implement an early warning mechanism for unusual patterns that may indicate events such as system failures and frauds [8, 1]. Anomalies, often referred to as outliers or deviants, are data points that deviate markedly from the expected values. In industry, anomalies are often so rare and it is too hard to label them for supervised learning. Hence, most studies in the literature focus on unsupervised methods such as clustering [15] and density estimation [2], or learning representations for only the normal data (supervising only for normal data). Because deep neural networks have the capacity to learn the representation of the normal data, reconstruction from the embedded of that data can be used to determine the anomalies. It means that reconstruction-based models [20] learn how to reconstruct the normal data, and high error in the reconstructed data indicates the anomalies. Similarly, forecasting-based methods [6] are also used to detect anomalies.

Contrastive representation got attention in computer vision tasks [3, 5], and applied for time series problem in a recent study [21]. Contrastive representation learning aims an embedding space emphasizing the distinction between similar and dissimilar data points. Combination of forecasting and reconstruction-based networks are also implemented in the literature [22, 19]. Although prior studies have achieved significant success, they may still suffer from performance degradation, particularly when anomalous points are not uniformly distributed, and anomaly scores or reconstruction errors may vary in the different regions of the data.

In this paper, we propose reconstruction networks based on Mean Squared Error (MSE) Feedback, augmented with Attention and Recurrent Neural Network (RNN) modules. We implement two variations of this architecture: AE-FAR, which employs an Autoencoder with Feedback Attention Reconstruction, and VAE-FAR, which utilizes a Variational Autoencoder (VAE) with Feedback Attention Reconstruction. In VAE-FAR, we integrate an Long Short-Term Memory (LSTM)-VAE with dual attention modules. Specifically, we implement two parallel graph attention mechanisms proposed in MTAD-GAT [22]. These modules are designed to capture temporal dependencies within time series data and relationships between features, enhancing the model's ability to detect anomalies effectively. Previous approaches mostly labels the anomalies if the reconstruction/forecasting error is larger than a prior threshold or dynamically determined threshold. According to the results we observed during our experiments, the reconstruction error does not progress similarly on the entire data set, there are fluctuations. Therefore, it is important to apply different thresholds at different time intervals, i.e. to identify the peaks in the errors. Since the previous and next values must be considered in the peak detection problem, a similar structure is placed in the proposed architecture by the reconstruction error feedback. We integrate a lightweight autoencoder model, a RNN module, and an attention mechanism in our model. Proposed architecture aims to improve reconstruction accuracy by dynamically adjusting the reconstruction of input sequences based on the MSE feedback with an RNN module. Our main motivation is that MSE feedback mechanism further enhances the model's ability to identify anomalies supported by the anomaly criterion. Our anomaly detection method is based on moving average and standard deviation within a sliding window for the reconstruction error.

- Our network is lightweight compared to the state-of-the-art methods and performs well with the proposed MSE feedback module and thresholding for anomaly detection.
- We train a simple autoencoder model and LSTM-VAE improved with graph attention mechanisms. Then we incorporate attention mechanism for the reconstructed values and obtained mean squared error. These pre-trained models are integrated with RNN to further enhance the reconstruction process by remembering the reconstruction MSE errors.
- We use a sliding window-based anomaly detection mechanism that focuses on sudden error changes rather than relying on a predefined general threshold. We show that our model's architecture inherently supports the sliding window-based thresholding approach.
- Through extensive experiments, we show that AE-FAR achieves an overall F1 score of 93.38 on the MSL, SWAT and SMD datasets and 58.48 on the pulp-and-paper industry dataset. VAE-FAR has an overall F1 score of 93.65 on the MSL, SWAT and SMD datasets, and 50.78 on the pulp-and-paper industry dataset.

2 Related works

Classical methods for time series anomaly detection have evolved significantly, adapting to various data characteristics. Techniques such as time series decomposition, clustering, and density estimation have offered robust solutions for identifying anomalies in time series data characterized by distinct patterns or substantial deviations from normal data distributions. Examples of classical anomaly detection methods include the Local Outlier Factor (LOF) [2] and Deep Autoencoding Gaussian Mixture Model (DAGMM) [24], both of which are grounded in density estimation principles. Distance to the cluster center is used as an anomaly score in clustering based methods. ITAD [16] applies a tensor-based decomposition to model normal behavior patterns and uses clustering techniques to group similar patterns. Deep-SVDD [15] trains a neural network to learn a representation of normal data and the objective is to map normal instances close to a central point in the latent space. IForest [11] detects anomalies by isolating observations through a recursive partitioning process. It randomly selects a feature and then chooses a split value between the minimum and maximum values of that feature. Autoregressive models predicting the future values based on past observations are another type of anomaly detection method. Recently, with the rise of deep learning, RNNs and their variants such as LSTM networks have been extensively applied, capable of capturing long-term dependencies and temporal patterns for detecting anomalies in diverse domains. CL-MPPCA [18] an extension of ARIMA, is one such method that compares predicted values with actual values and detects deviations that exceed a certain threshold. It combines the capabilities of LSTM-based neural networks and mixture of several probabilistic PCA models.

Autoencoders are a type of neural networks designed to learn embedded representations of data, typically for the purposes of dimensionality reduction or feature learning. They consist of two main components: an encoder that maps the input data to a lower-dimensional latent space, and a decoder that reconstructs the input data from this latent representation. VAEs extend the autoencoder by encoding inputs into distributions, typically Gaussian. The decoder reconstructs the data from these distributions instead of fixed points in the latent space. Employing VAE, LSTM-VAE [13] and some improved versions [17, 9] of LSTM-VAE are applied for the anomaly problem. OmniAnomaly [17] employs a stochastic RNN framework with GRU, integrating VAEs to model the temporal dependencies. InterFusion [10] proposes a hierarchical VAE to model inter-metric and temporal relationships. MAD-GAN [9] is Generative Adversarial Network(GAN) based method in which LSTM is used in generator and discriminator networks. DGHL [4] proposes a hierarchical latent space representation with convolution networks. BEATGAN [23] is also a reconstruction-based method using generative adversarial networks. MTAD-GAT [22] combines forecasting-based and reconstruction-based networks and anomalies are detected by using both the these outputs. AnomalyTransformer [20] introduces a new Anomaly-Attention mechanism to compute the association discrepancy focusing on the difference between normal and anomalous patterns. DCDetector [21] is a contrastive learning based multi-scale dual attention model.

3 Method

Let X be a multivariate time-series sequence of length N:

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$$

The unsupervised time series anomaly detection problem aims to ascertain the anomalous nature of X without the availability of labeled data. This work proposes a composite model that combines an AE or VAE, an attention mechanism, and RNN to address this problem. The general overall of proposed architecture is shown in Figure 1. AE-FAR integrates multiple neural network components to enhance time series prediction accuracy through error correction. It consists of three main parts: an autoencoder, an attention layer, and RNN.

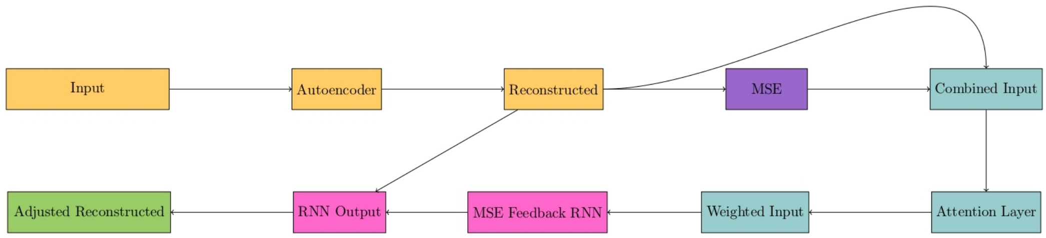

Figure 1 General overview of the proposed AE-FAR architecture.

3.1 Overall architecture

We train a simple AE to reconstruct the input **X**. $\hat{\text{X}}$ represents the reconstructed input values with AE network. In the proposed composite model we use the MSE error of that AE.

$$\begin{aligned} \hat{X} &= \text{Autoencoder}(X) \\ \text{MSE} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2 \end{aligned} \tag{1}$$

The combination of $\hat{\text{X}}$ and MSE is passed to the attention mechanism. The attention layer is responsible for assigning different importance levels to various parts of the input sequence. That helps the model focus on the most relevant features during the prediction process. The attention layer uses a sequential model consisting of linear transformations, a Tanh activation function, and a Softmax function to compute attention weights. In the following equation, α represents the attention weights. These weights are then applied to the input features to produce a weighted input.

$$\begin{aligned} z &= [\hat{X}, \text{MSE}] \\ \alpha &= \text{Softmax}(\text{Linear}(\tanh(\text{Linear}(z)))) \\ z' &= \alpha \odot z. \end{aligned} \tag{2}$$

The MSE Feedback RNN module is a recurrent neural network that processes sequences of inputs and provides feedback based on the mean squared error between the predicted and actual values. This module includes an RNN layer followed by a fully connected layer. The RNN processes the input sequence and outputs hidden states, which are then transformed by the fully connected layer to produce the final output. Finally, the adjusted reconstruction, incorporating RNN predictions, forms the model's output, aiming to refine and improve the accuracy of predictions over time.

$$\begin{aligned} h &= \text{RNN}(z') \\ \hat{y} &= \hat{X} + h \end{aligned} \tag{3}$$

AE-FAR integrates the autoencoder, attention mechanism, and MSE Feedback RNN to enhance the forecasting capabilities. Combining all parts, the model output is:

$$\hat{y} = \hat{X} + \text{RNN}(\alpha \odot [\hat{X}, \text{MSE}]) \tag{4}$$

This approach aims to enhance the reconstructed output ŷ by incorporating error-driven adjustments from the RNN, thereby improving predictive accuracy in time-series analysis. Sample reconstruction errors and anomaly points are shown in Figure 2 comparing the results of the AE with the proposed AE-FAR model, highlighting the influence of the MSE feedback module. The figure illustrates a region containing 14 actual anomalies, represented by vertical red dotted lines. The AE method detects 39 anomaly points, whereas the AE-FAR method identifies 30 anomaly points. It is observed that the AE-FAR method has a lower mean error value compared to the AE method, and its anomaly detections are closer to the actual anomaly points. The mean distances between the predicted anomaly indexes and the ground truth anomaly indexes are 87.5 and 72.8 for the AE method and the AE-FAR method, respectively.

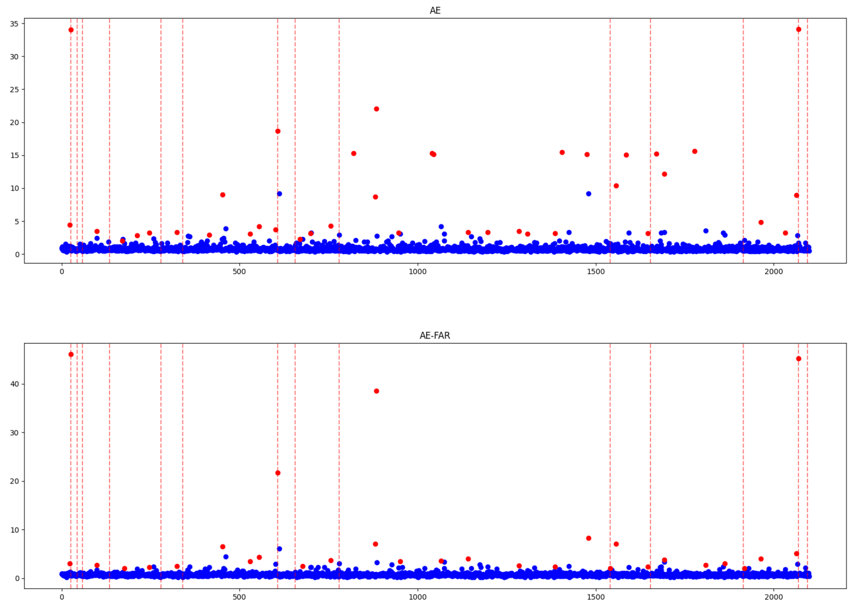

Figure 2 Reconstruction error and detected anomaly points. Vertical red dotted lines indicate the real anomalies.

3.2 LSTM-VAE with attention layers

We employ a different variant of that proposed architecture by using VAE instead of AE. In the VAE network, we use graph-based feature and attention layers proposed in MTAD-GAT [22]. The general overview of the VAE improved with the attention layers is shown in the Figure 3. Feature attention layer aims to emphasize the most relevant features for each time step while temporal attention layer applies attention across the temporal dimension, focusing on the most significant time steps for each feature. The encoded representation is formed by concatenating the original input X with the outputs of the feature and temporal attention layers. Reparameterize function implements the reparameterization trick to sample latent variables from a Gaussian distribution inferred by the encoder's output. Decoder module reconstructs the input sequence from the sampled latent variables.

3.3 Anomaly criterion

We use a sliding window-based approach to dynamically compute thresholds based on statistical properties of the reconstruction error. The proposed anomaly detection strategy is based on the MSE values using a moving average and standard deviation within a sliding

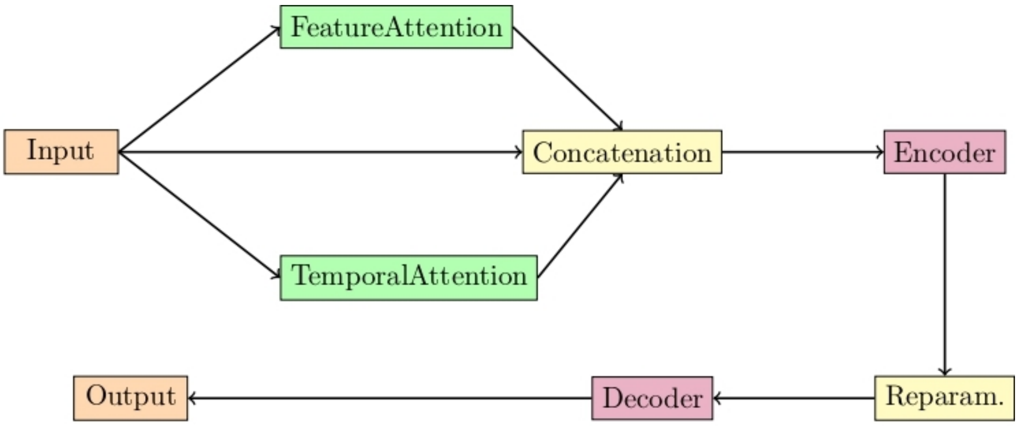

Figure 3 General overview of the VAE with graph attention.

window. This approach is particularly useful in anomaly detection where the MSE distribution may change over time. Our model's architecture inherently supports the sliding window-based thresholding mechanism, making it more effective in dynamically adjusting to varying anomaly patterns within the time series data. Window size is a significant hyper-parameter in time series analysis to split time series into instances instead of using only a single point as input. We use the same window size for the the sliding window used to compute the moving average and moving standard deviation. The adaptive nature of the sliding window approach ensures that the detection mechanism remains sensitive to new patterns, providing a robust tool for maintaining robust performance in anomaly detection.

Given a sequence of MSE values mse_list $= [mse_1, mse_2, \ldots, mse_t]$, dynamic threshold is calculated based on average and standard deviation for a window size w as follows:

$$\begin{aligned} avg[t] &= \frac{1}{w}\sum_{i=0}^{w} mse_{t-i} \\ std[t] &= \sqrt{\frac{1}{w}\sum_{i=0}^{w}(mse_{t-i} - avg[t])^2} \end{aligned} \tag{5}$$

The dynamic threshold $T[t]$ for each input X in a time step is then calculated using a user-defined threshold factor β, and anomalies(A) are identified at each time step t if the MSE mse_t exceeds the computed dynamic threshold $T[t]$:

$$T[t] = avg[t] + \beta \times std[t] \tag{6}$$

$$A_t = \begin{cases} 1 & \text{if } mse_t > T[t] \\ 0 & \text{otherwise} \end{cases}$$

4 Experiments

4.1 Benchmark datasets

We perform experiments using four datasets: the Mars Science Laboratory (MSL) [7] rover dataset, the Server Machine Dataset (SMD) [17], Secure Water Treatment (SWAT) [12], and pulp-and-paper manufacturing industry dataset [14]. MSL is collected by NASA and it reflects the rover's operational status and environment. SMD contains data collected from various server machines in a data center. It includes metrics such as CPU usage, memory usage, and network traffic, aimed at detecting anomalies that may indicate hardware failures or network issues. SWAT includes sensor data from critical infrastructure systems. The pulp-and-paper industry dataset used in our experiments has a total of 59 input features, after removing the categorical features x28 and x61. This dataset consists of 18,398 rows, of which only 124 rows are labeled as anomalies. A significant characteristic that distinguishes this dataset from others is the presence of non-consecutive anomalies, which poses a unique challenge for anomaly detection methods. We create two subset datasets (called Paper-1 and Paper-2) for testing purposes, containing 29 and 58 anomalies respectively.

We adopt a widely recognized adjustment technique to ensure a fair comparison with existing methods in the literature. That adjustment approach refines predicted anomaly labels in time series data by ensuring that if any single point in an anomalous segment is detected, the entire segment is marked as anomalous. This strategy is justified by the observation that detecting a single anomalous point triggers an alert for the entire segment in real-world applications. While this adjustment technique significantly impacts datasets like MSL, SMD, and SWAT, it has no effect on Paper-1 and Paper-2 datasets. We employ a neighborhood-based matching strategy to assess the performance on the paper industry dataset. This approach was chosen due to the difficulty of precisely determining the exact timing of anomalies in this particular industry dataset. We apply a slight temporal misalignments between true and predicted anomalies by defining a window size k. True positives are correctly predicted anomalies within k indices of actual anomalies, while false positives are predictions without a corresponding true anomaly in the window. This approach enhances robustness in detecting anomalies in time series data by accounting for small deviations, thus providing a more accurate assessment of model performance in practical scenarios. It offers a comprehensive evaluation of precision, recall, and F1 score, rounded for clarity. Table 4 represents the performance comparison with this approach.

4.2 Implementation details

We use a fixed window size *50* for all datasets to split time series into instances. The same value of 50 is also used for k to apply a slight temporal misalignments for pulp-and-paper industry dataset. The AE model consists of four linear layers: the first layer compresses the input from a flattened dimension of $window * inputsize$ to 32 neurons, the second layer further reduces the size to 16 neurons, and then gradually expanded to the original input size. For the VAE network, the dimension of *hidden size* and *latent size* are 64 and 32, respectively. RNN component is an 8-layer RNN that processes the combined input with dimensions $inputsize + 1$. It uses a hidden size of $inputdim/2$ for the hidden state and maps the output to the original input dimension using a fully connected layer. The attention mechanism takes an input dimension of $inputsize + 1$ and maps it to an attention dimension *64*. The experiments with these hyperparameter selection are implemented in PyTorch with NVIDIA GeForce RTX 3060 graphic card. We use Adam optimizer with an initial learning rate of 10^{-4} and set the batch size to 128. We used early stopping during training by monitoring the validation error, with an early stop value set to 10, to prevent overfitting and ensure optimal model performance. We choose a fixed threshold value as the default comparison, and for the proposed anomaly criterion, we define different threshold factors β 4, 5 and 6.5 for SMD, SWAT and MSL, respectively.

Table 1 Ablation study on MSL, SMD and SWAT. * indicates the proposed anomaly criterion instead of fixed threshold.

	MSL			SMD			SWAT		
Methods	***P***	**R**	***F1***	***P***	**R**	***F1***	***P***	**R**	***F1***
AE	**95.70**	20.91	34.32	38,10	68,09	48,86	91.62	76.41	83.33
AE-FAR	83.57	90.45	**86.87**	62.72	52.01	56.86	92.12	76.41	83.53
VAE-FAR	52.87	**96.87**	68.41	**81.28**	**76.81**	**78.98**	**97.43**	**77.09**	**86.07**
AE^*	85.23	**99.54**	91.83	91.89	87.78	89.79	94.87	96.97	95.91
$AE-FAR^*$	**92.59**	95.13	**93.85**	**92.01**	88.20	90.06	**97.40**	96.36	**96.87**
$VAE-FAR^*$	90.23	94.89	92.50	89.95	**94.59**	**92.21**	94.49	**98.07**	96.24

4.3 Results and analysis

We use commonly-used evaluation measures: precision, recall, F1 score for performance comparison. The ablation study represented in Table 1 evaluates the performance on MSL, SMD and SWAT benchmark datasets. The first three rows represent the performance with fixed general threshold, while the other rows represent the performance with the proposed anomaly criterion. The best results obtained with both approaches are shown separately in **bold**. The results clearly demonstrate the effectiveness of the anomaly criterion method improving the AE, AE-FAR and VAE-FAR models. The AE with anomaly criterion shows substantial gains in recall and F1 scores, especially for the MSL and SWAT datasets. VAE-FAR has better recall compared to the AE-FAR with the general fixed threshold. Both of the AE-FAR and VAE-FAR models, when combined with the proposed anomaly criterion, consistently outperforms other configurations, achieving the highest F1 scores across all datasets. This highlights the robustness and accuracy of the proposed methods in multivariate time-series anomaly detection.

Table 2 represents the ablation study for subsets of pulp-and-paper industry datasets. The AE-FAR model alone shows high precision for the Paper-1 dataset at 66.67%, but recall is very low at 14.81%, resulting in an F1 score of 24.24%. This suggests that the model is very conservative in identifying anomalies, leading to fewer false positives but many missed anomalies with the fixed threshold. On the other hand, VAE-FAR has more stabil and better results compared to the AE-FAR with the fixed threshold. Using the proposed anomaly criterion significantly improves performance as shown in last three rows. For the Paper-1 dataset, it achieves a balanced precision of 56.41% and a high recall of 81.48%, resulting in the highest F1 score of 66.67% among all methods. VAE-FAR* has the second best value with an F1 score of 58.46 with the anomaly criterion. For the Paper-2 dataset, AE-FAR* achieves a precision of 37.29% and a recall of 77.19%, resulting in the highest F1 score of 50.29% . This demonstrates the effectiveness of the anomaly criterion in improving the model's ability to detect anomalies accurately. The AE model benefits significantly from the anomaly criterion, especially in terms of recall, but at the cost of precision. The AE-FAR model alone shows high precision but struggles with recall, indicating a conservative anomaly detection approach. The combination of AE-FAR with the anomaly criterion achieves the best overall performance, striking a balance between precision and recall and resulting in the highest F1 scores for both datasets.

Table 3 shows the performance comparison of different anomaly detection methods in the literature. For three real world datasets MSL, SMD and SWAT: AE-FAR avd VAE-FAR achieve an overall F1-Score of 93.59 and 93.65, while DCdetector and AnomalyTransformer have 93.37 and 93.32, respectively. Table 4 presents the performance comparison with the

Table 2 Ablation study on pulp-and-paper industry dataset. * indicates the proposed anomaly criterion instead of fixed threshold.

Methods	Paper-1 *P*	Paper-1 R	Paper-1 *F1*	Paper-2 *P*	Paper-2 R	Paper-2 *F1*
AE	44.74	**62.96**	**52.31**	23.58	**43.86**	30.67
AE-FAR	**66.67**	14.81	24.24	**40.0**	14.04	20.78
VAE-FAR	45.83	40.74	43.14	31.34	36.84	**33.87**
AE^*	38.46	**92.59**	54.35	29.49	**80.70**	43.19
$AE-FAR^*$	**56.41**	81.48	**66.67**	**37.29**	77.19	**50.29**
$VAE-FAR^*$	50.0	70.37	58.46	32.73	63.16	43.11

Table 3 Results on multivariate benchmark datasets. All results are presented as percentages; the best values are in **bold**, and the second-best are underlined.

Dataset	*MSL*			*SWAT*			*SMD*		
Metric	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
LOF	47.72	85.25	61.18	72.15	65.43	68.62	56.34	39.86	46.68
IForest	53.94	86.54	66.45	49.29	44.95	47.02	42.31	73.29	53.64
DAGMM	89.60	63.93	74.62	89.92	57.84	70.40	67.30	49.89	57.30
ITAD	69.44	84.09	76.07	63.13	52.08	57.08	86.22	73.71	79.48
CL-MPPCA	73.71	88.54	80.44	76.78	81.50	79.07	82.36	76.07	79.09
Deep-SVDD	91.92	76.63	83.58	80.42	84.45	82.39	78.54	79.67	79.10
BeatGAN	89.75	85.42	87.53	64.01	87.46	73.92	72.90	84.09	78.10
OmniAnomaly	89.02	86.37	87.67	81.42	84.30	82.83	83.68	86.82	85.22
InterFusion	81.28	92.70	86.62	80.59	85.58	83.01	87.02	85.43	86.22
AnomalyTransformer	92.09	<u>95.15</u>	93.59	91.55	96.73	94.07	89.40	**95.45**	**92.33**
DCdetector	**93.69**	**99.69**	**96.60**	93.11	**99.77**	<u>96.33</u>	83.59	91.10	87.18
AE-FAR	<u>92.59</u>	95.13	<u>93.85</u>	**97.40**	96.36	**96.87**	**92.01**	88.20	90.06
VAE-FAR	90.23	94.89	92.50	<u>94.49</u>	<u>98.07</u>	96.24	89.95	<u>94.59</u>	<u>92.21</u>

DCDetector on two subsets of a pulp-and-paper manufacturing industry anomaly detection dataset: Paper-1 and Paper-2. For Paper-1, Introducing the proposed anomaly criterion to DCdetector slightly changes the results, with a precision of 31.43%, recall of 40.74%, and F1 score of 35.48%. There is a small reduction in precision and recall, which suggests that the proposed thresholding might be filtering out some true positives, leading to a lower overall performance. For Paper-2, proposed anomaly criteria improves the performance slightly, achieving a precision of 28.21%, recall of 38.60%, and F1 score of 32.59%. On the other hand, AE-FAR outperforms the other methods significantly for both Paper-1 and Paper-2 datasets. Although VAE-FAR shows lower performance compared to AE-VAR, it is seen that it gives much better performance than DCDetector. The proposed anomaly criterion for DCdetector shows minor improvements but does not suffice to compete with the performance of AE-FAR. These findings highlight the effectiveness of the proposed approach in handling the rare anomalies in the pulp-and-paper manufacturing industry datasets.

5 Conclusion

The proposed AE-FAR/VAE-FAR models effectively combine AE/VAE, an attention mechanism, and RNN to improve anomaly detection in time series data. The AE/VAE is employed to reconstruct the input data, aiming to capture the underlying normal patterns. They

Table 4 Results on pulp-and-paper manufacturing industry dataset. All results are presented as percentages; the best values are in **bold**, and the second-best are underlined.

Dataset	*Paper-1*			*Paper-2*		
Metric	**P**	**R**	**F1**	**P**	**R**	**F1**
DCdetector	35.29	44.44	39.34	23.17	33.33	27.34
DCdetector*	31.43	40.74	35.48	28.21	38.60	32.59
AE-FAR	**56.41**	**81.48**	**66.67**	**37.29**	**77.19**	**50.29**
VAE-FAR	50.0	70.37	58.46	32.73	63.16	43.11

consist of an encoder that maps the input data to a lower-dimensional latent space and a decoder that reconstructs the input from this latent representation. The attention mechanism is integrated to enhance the model's focus on significant parts of the input data. The RNN component processes the reconstructed data produced by the autoencoder and reconstruction errors. The RNN output is used to adjust the reconstructed input, providing the final output of the model. This integrated approach leads to more accurate anomaly detection results. The proposed models outperform state-of-the-art approaches overall on four datasets.

Future work. We need to focus on anomaly criterion selecting threshold factor β automatically, and better detection and localization for discontinuous anomalies such as pulp-and-paper dataset.

References

1 Archana Anandakrishnan, Senthil Kumar, Alexander Statnikov, Tanveer Faruquie, and Di Xu. Anomaly detection in finance: editors' introduction. In *KDD 2017 Workshop on Anomaly Detection in Finance*, pages 1–7. PMLR, 2018.

2 Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. Lof: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 93–104, 2000. doi:10.1145/342009.335388.

3 Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *Advances in neural information processing systems*, 33:9912–9924, 2020.

4 Cristian I Challu, Peihong Jiang, Ying Nian Wu, and Laurent Callot. Deep generative model with hierarchical latent factors for time series anomaly detection. In *International Conference on Artificial Intelligence and Statistics*, pages 1643–1654. PMLR, 2022. URL: https://proceedings.mlr.press/v151/challu22a.html.

5 Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020. URL: http://proceedings.mlr.press/v119/chen20j.html.

6 Ailin Deng and Bryan Hooi. Graph neural network-based anomaly detection in multivariate time series. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 4027–4035, 2021. doi:10.1609/AAAI.V35I5.16523.

7 Kyle Hundman, Valentino Constantinou, Christopher Laporte, Ian Colwell, and Tom Soderstrom. Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 387–395, 2018.

8 Bedeuro Kim, Mohsen Ali Alawami, Eunsoo Kim, Sanghak Oh, Jeongyong Park, and Hyoungshick Kim. A comparative study of time series anomaly detection models for industrial control systems. *Sensors*, 23(3):1310, 2023. doi:10.3390/S23031310.

9 Dan Li, Dacheng Chen, Baihong Jin, Lei Shi, Jonathan Goh, and See-Kiong Ng. Mad-gan: Multivariate anomaly detection for time series data with generative adversarial networks. In *International conference on artificial neural networks*, pages 703–716. Springer, 2019. doi:10.1007/978-3-030-30490-4_56.

10 Zhihan Li, Youjian Zhao, Jiaqi Han, Ya Su, Rui Jiao, Xidao Wen, and Dan Pei. Multivariate time series anomaly detection and interpretation using hierarchical inter-metric and temporal embedding. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pages 3220–3230, 2021. doi:10.1145/3447548.3467075.

11 Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 eighth ieee international conference on data mining*, pages 413–422. IEEE, 2008. doi:10.1109/ICDM.2008.17.

12 Aditya P Mathur and Nils Ole Tippenhauer. Swat: A water treatment testbed for research and training on ics security. In *2016 international workshop on cyber-physical systems for smart water networks (CySWater)*, pages 31–36. IEEE, 2016. doi:10.1109/CYSWATER.2016.7469060.

13 Daehyung Park, Yuuna Hoshi, and Charles C Kemp. A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder. *IEEE Robotics and Automation Letters*, 3(3):1544–1551, 2018. doi:10.1109/LRA.2018.2801475.

14 Chitta Ranjan, Mahendranath Reddy, Markku Mustonen, Kamran Paynabar, and Karim Pourak. Dataset: rare event classification in multivariate time series. *arXiv preprint arXiv:1809.10717*, 2018.

15 Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *International conference on machine learning*, pages 4393–4402. PMLR, 2018.

16 Youjin Shin, Sangyup Lee, Shahroz Tariq, Myeong Shin Lee, Okchul Jung, Daewon Chung, and Simon S Woo. Itad: integrative tensor-based anomaly detection system for reducing false positives of satellite systems. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 2733–2740, 2020. doi:10.1145/3340531.3412716.

17 Ya Su, Youjian Zhao, Chenhao Niu, Rong Liu, Wei Sun, and Dan Pei. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2828–2837, 2019. doi:10.1145/3292500.3330672.

18 Shahroz Tariq, Sangyup Lee, Youjin Shin, Myeong Shin Lee, Okchul Jung, Daewon Chung, and Simon S Woo. Detecting anomalies in space using multivariate convolutional lstm with mixtures of probabilistic pca. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2123–2133, 2019. doi:10.1145/3292500.3330776.

19 Lawrence C Wong. *Time Series Anomaly Detection using Prediction-Reconstruction Mixture Errors*. PhD thesis, Massachusetts Institute of Technology, 2022.

20 Jiehui Xu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Anomaly transformer: Time series anomaly detection with association discrepancy. *arXiv preprint arXiv:2110.02642*, 2021. arXiv:2110.02642.

21 Yiyuan Yang, Chaoli Zhang, Tian Zhou, Qingsong Wen, and Liang Sun. Dcdetector: Dual attention contrastive representation learning for time series anomaly detection. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3033–3045, 2023. doi:10.1145/3580305.3599295.

22 Hang Zhao, Yujing Wang, Juanyong Duan, Congrui Huang, Defu Cao, Yunhai Tong, Bixiong Xu, Jing Bai, Jie Tong, and Qi Zhang. Multivariate time-series anomaly detection via graph attention network. In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 841–850. IEEE, 2020. doi:10.1109/ICDM50108.2020.00093.

23 Bin Zhou, Shenghua Liu, Bryan Hooi, Xueqi Cheng, and Jing Ye. Beatgan: Anomalous rhythm detection using adversarially generated time series. In *IJCAI*, volume 2019, pages 4433–4439, 2019. doi:10.24963/IJCAI.2019/616.

24 Bo Zong, Qi Song, Martin Renqiang Min, Wei Cheng, Cristian Lumezanu, Daeki Cho, and Haifeng Chen. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *International conference on learning representations*, 2018.