

Multi-scale LSTM forecasting with attention-based feature fusion for anomaly detection in multivariate monitoring data

Liu Wei

weiliu_21@163.com

State Grid Jiangsu Electric Power Co Ltd, Information & Telecommunication Branch

Meizhao Liu

State Grid Jiangsu Electric Power Co Ltd, Information & Telecommunication Branch

Fei Xia

State Grid Jiangsu Electric Power Co Ltd, Information & Telecommunication Branch

Qiuyuan Zhong

State Grid Jiangsu Electric Power Co Ltd, Information & Telecommunication Branch

Shengjie Wei

State Grid Jiangsu Electric Power Co Ltd, Information & Telecommunication Branch

Wen Yang

State Grid Jiangsu Electric Power Co Ltd, Information & Telecommunication Branch

Research Article

Keywords: Anomaly Detection, Multivariate time series, Condition monitoring, Multi-scale temporal modeling, Attention mechanism, LSTM forecasting, Mahalanobis distance

Posted Date: September 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11004283/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

1 Multi-scale LSTM forecasting with
2 attention-based feature fusion for anomaly
3 detection in multivariate monitoring data

4 Liu Wei^{1*}, Meizhao Liu¹, Fei Xia¹, Qiuyuan Zhong¹,
5 Shengjie Wei¹, Wen Yang¹

6 ¹State Grid Jiangsu Electric Power Co Ltd, Information &
7 Telecommunication Branch, Nanjing, 210046, China.
8 Corresponding author: Liu Wei.

9 *Corresponding author(s). E-mail(s): weiliu_21@163.com;

10 **Abstract**

11 Engineering systems such as power equipment, industrial processes, and
12 computing infrastructure are monitored through multivariate time series
13 collected continuously from sensors and instruments. Anomalies in such
14 data appear at markedly different temporal scales, ranging from grad-
15 ual drifts that develop over long periods to transient impulses that last
16 only a few sampling intervals. Most existing detection methods rely on a
17 single temporal resolution or on fixed relationships among variables and
18 therefore cannot capture all of these patterns at the same time. This
19 paper presents MAF-LSTM, a forecasting-based anomaly detection frame-
20 work that combines multi-scale long short-term memory (LSTM) networks

with attention-based feature fusion. Parallel LSTM branches operate on input windows of different lengths to extract temporal dependencies over multiple horizons, and an attention module adaptively weights the resulting representations to produce a one-step-ahead forecast. Anomaly scores are computed from the forecasting residuals using the Mahalanobis distance, which accounts for the correlation structure of the residuals and improves robustness to distribution shift. The framework is evaluated on five widely used multivariate benchmark datasets from spacecraft telemetry, server monitoring, and industrial control systems. It achieves the highest F1-score on four of the five datasets when compared with four representative baselines, and an ablation study confirms that multi-scale modeling, attention-based fusion, and Mahalanobis scoring each contribute to the improvement.

Keywords: Anomaly Detection, Multivariate time series, Condition monitoring, Multi-scale temporal modeling, Attention mechanism, LSTM forecasting, Mahalanobis distance

1 Introduction

Multivariate time series are the primary form in which the operating state of engineering systems is recorded. Power equipment such as transformers and circuit breakers is monitored through oil temperature, partial discharge signals, and load currents; industrial processes are monitored through sensor and actuator readings; and computing infrastructure is monitored through server performance metrics. Anomalies in these measurements usually represent states that deviate from expected patterns and may indicate equipment malfunction or service disruption [33]. Anomaly detection in monitoring time series has therefore become an essential component of condition-based maintenance and reliable operation of such systems [14].

47 However, real-world time series data usually consists of several variables (or fea-
48 tures), which makes anomaly detection particularly challenging [12, 29], especially in
49 data-intensive applications such as power grid, smart factories or supply chain systems.
50 In these systems, massive amounts of data are continuously produced by distributed
51 sensors, where each measurement is characterized by at least three dimensions: time,
52 sensors, and monitor variables. Their complex spatio-temporal dependencies make
53 the patterns challenging to distinguish, and thus anomaly detection becomes more
54 arduous, especially under ultrafast detection demands [37].

55 To address this challenge, significant research focus on the area of multivariate
56 time series anomaly detection. As most of the historical data is unlabeled and anoma-
57 lies commonly do not have a predictable structure, anomaly detection is normally
58 approached as an unsupervised learning task [38]. There are three primary types of
59 existing methods, which can be divided into statistical-based, reconstruction-based,
60 and forecasting-based approaches. Statistical-based methods identify anomalies by
61 modeling the probabilistic distribution of time series data and detecting points that
62 deviate significantly from the expected patterns. The standard approaches include
63 Isolation Forest (iForest) [15], K-means [9], and other statistical hypothesis test-
64 ing methods [4, 17, 22]. Reconstruction-based models leverage the minimization of
65 input-output residuals to extract the noticeable characteristics of normal sequences
66 within a compressed feature space. Samples with large reconstruction errors are
67 regarded as anomaly. This category has evolved from traditional Autoencoders (AE)
68 [18] to advanced variants such as long short-term memory variational autoencoders
69 (LSTM-VAE) [23] and Generative Adversarial Networks (GAN) [3]. Forecasting-based
70 methods, on the other hand, build models that forecast future values based on his-
71 torical observations, and anomalies are detected when the prediction error exceeds

72 a certain threshold. Representative examples include autoregressive moving aver-
73 age (ARMA) [8], long short-term memory (LSTM) [10], and Hierarchical Temporal
74 Memory (HTM) [2].

75 While significant progress has been made, multivariate anomaly detection still
76 faces two key challenges. First, capturing the complex nonlinear patterns in real-
77 world time series remains a major challenge. Most existing approaches either extract
78 features within a fixed window or reconstruct the entire sequence, but still rely on
79 single-scale representations. However, recent studies have shown that this kind of
80 approach struggles to distinguish anomalies that manifest across different temporal
81 resolutions [24]. They are prone to confusion in cases such as local spikes versus
82 seasonal fluctuations, locally stable patterns versus global contextual deviations, and
83 short-lived anomalies versus coarse-scale smoothing. For example, in **Fig. 1**, under S_1
84 scale, current algorithms tend to distinguish P_0 as an anomaly, while under S_2 scale,
85 it is only a seasonal peak. Second, the correlations between variables change over time
86 [40]. Although some works utilize Graph Neural Network (GNN) to represent complex
87 relationships in multidimensional data [26, 31], most of them can only capture fixed
88 correlation, which cannot reflect the dynamic changes of dependencies among variables
89 timely.

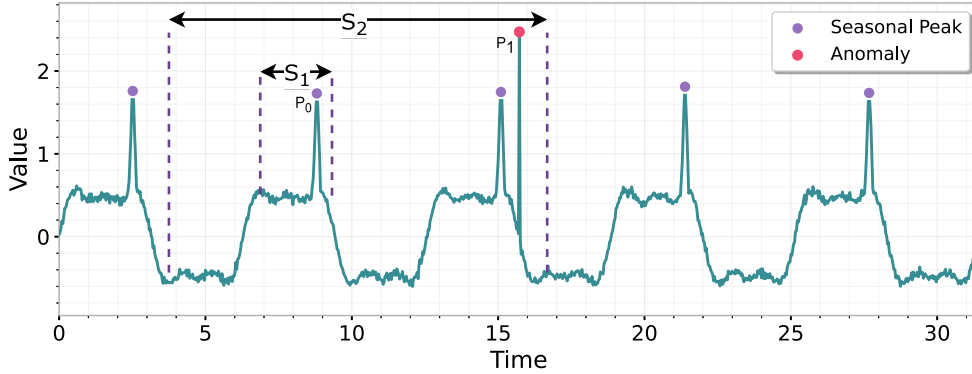


Fig. 1 Different scales in time series.

90 To bridge this gap, we propose MAF-LSTM, a Multi-scale Attention-based
91 Feature-fusion LSTM framework. At its core, MAF-LSTM employs a multi-scale
92 LSTM forecasting network to capture temporal dependencies across different time
93 horizons, enabling the extraction of predictive representations that are highly sensi-
94 tive to abnormal behaviors. The learned temporal features are then combined with an
95 attention mechanism, which dynamically adjusts decision weights over time, allowing
96 the model to emphasize the most informative scales as conditions evolve. Further-
97 more, to address distribution shifts caused by data drift, MAF-LSTM integrates a
98 Mahalanobis-distance-based anomaly scoring module that can fit data distribution
99 and enhances detection robustness. We validate the effectiveness of the proposed
100 framework through extensive experiments.

101 Building upon the architecture of MAF-LSTM, this paper makes the following
102 contributions:

- 103 • We propose a multi-scale perspective for time series anomaly detection based on
104 forecasting method, an aspect seldom explored in prior research. This perspective
105 enables the model to identify anomalies that appear differently across temporal
106 resolutions.
- 107 • We design an LSTM-based prediction framework with attention mechanism to cap-
108 ture temporal dependencies at multiple scales. The proposed structure improves
109 anomaly detection accuracy and reduces both false alarms and missed detection.
- 110 • We evaluate MAF-LSTM on five real-world datasets. The results demonstrate
111 that our method outperforms state-of-the-art baselines in anomaly detection tasks,
112 achieving an average F1-score improvement of 2.52% over the best-performing
113 baseline.

114 The rest of this paper is organized as follows. Section 2 reviews related work on
115 forecasting-based and representation-based anomaly detection. Section 3 describes
116 the proposed MAF-LSTM framework, datasets, and experimental setup. Section 4

117 presents and discusses the experimental results on real-world datasets. Section 5
118 concludes the study and discusses future research directions.

119 2 Background and Related Work

120 Over the past decades, time series anomaly detection has been an active area of
121 research. Current methods can generally be divided into three main categories:
122 statistical-based, reconstruction-based, and forecasting-based. This section summa-
123 rizes the above categories. In addition, we investigate recent efforts that focuses on
124 multi-scale feature extraction and highlight their limitations.

125 2.1 Statistical-based Methods

126 Early research on time series anomaly detection primarily rely on statistical meth-
127 ods, which identify abnormal behaviors by modeling data distributions and detecting
128 significant deviations from common patterns. Such approaches typically assume that
129 time series follow a predefined distribution, such as Gaussian or Poisson [7], and data
130 points lying outside a certain confidence interval are classified as anomalies [28].

131 Beyond purely statistical methods, a variety of classical paradigms have been
132 proposed, including partition-based, clustering-based, and distance-based approaches.
133 The partition-based Isolation Forest (iForest) algorithm [15], for example, isolates
134 anomalies through recursive random partitioning, leveraging the principle that anoma-
135 lies are more easily separated in sparse regions of the feature space. Clustering-based
136 methods [6, 9] segment time series into subsequences and apply unsupervised clus-
137 tering (e.g., K-means or Density-Based Spatial Clustering of Applications with Noise
138 (DBSCAN)) to group similar temporal patterns, identifying segments that devi-
139 ate markedly from cluster centers as anomalous. Distance-based techniques [36], in
140 contrast, measure similarity between subsequences using metrics such as Euclidean
141 distance and flag points with unusually large neighbor distances as outliers.

2.2 Reconstruction-based Methods

Reconstruction-based methods constitute another major line of research in time series anomaly detection. These approaches typically employ encoder-decoder architectures, such as Autoencoders (AE) [18], Variational Autoencoders (VAE) [23], and Generative Adversarial Networks (GAN) [3], to learn low-dimensional representations of input data and reconstruct the original signals. Anomalies are identified by measuring reconstruction errors, as abnormal instances are generally harder to reproduce accurately.

Among them, Autoencoder-based deep learning models have been widely adopted. The LSTM-AE model [18] leverages recurrent units to capture temporal dependencies and evaluates anomalies based on reconstruction loss. Zong et al. [41] further proposed the Deep Autoencoding Gaussian Mixture Model (DAGMM), which integrates a deep autoencoder with a Gaussian mixture estimation network for joint optimization. Although effective in modeling spatial dependencies, DAGMM lacks temporal modeling capability. Variational Autoencoder-based methods extend this paradigm by incorporating probabilistic inference. LSTM-VAE [23] replaces feedforward layers with LSTM units to enhance sequential representation learning, while OmniAnomaly [27] employs a stochastic recurrent VAE and planar normalizing flows to capture complex non-Gaussian latent distributions, achieving robust multivariate anomaly detection.

2.3 Forecasting-based Methods

Forecasting-based approaches exploit temporal dependencies by leveraging historical observations to forecast future values, and detect anomalies based on the deviation between predicted and actual observations. These methods commonly employ recurrent neural networks (e.g., LSTM [10]) or Transformer-based architectures [5, 25] to predict the next time step or multiple future steps. When the prediction error exceeds

167 a predefined or adaptively learned threshold, the corresponding time point is regarded
168 as anomalous.

169 Malhotra et al. [19] introduced the LSTM-AD model, which employs stacked
170 LSTM layers to forecast future values and models the anomaly score as a Gaussian
171 distribution to determine the detection threshold. However, a single recurrent layer
172 often struggles to simultaneously capture both short-term and long-term dependen-
173 cies. Hundman et al. [11] proposed LSTM-NDT, which enhances threshold selection
174 through a non-parametric dynamically adjusted threshold (NDT) strategy that
175 updates anomaly thresholds based on the moving average of prediction errors. Song
176 et al. [25] further developed SAnD, a Transformer-based model employing multi-head
177 self-attention and temporal encoding to capture long-range dependencies, achiev-
178 ing superior performance over conventional recurrent models in healthcare anomaly
179 detection tasks.

180 **2.4 Leveraging Multi-Scale Features for Accurate Anomaly** 181 **Detection**

182 Previous research on time series anomaly detection has often overlooked multi-scale
183 features. For example, although Zhang et al. [39] use signature matrices with different
184 lookback lengths, they simply concatenate these matrices as inputs. This approach
185 lacks a mechanism to effectively fuse information across different scales during the
186 encoding and decoding process. OC4Seq [32] employs two separate recurrent neural
187 network (RNN) branches to respectively learn global and local representations, while
188 a weighted loss is introduced to balance their contributions. In this way, the model is
189 able to capture multiple levels of sequential patterns within discrete event streams. In
190 [30], a variational Transformer is introduced to learn correlations with self-attention
191 and captures multi-scale temporal information using improved positional encoding

192 and up-sampling. A residual VAE further models local features in the latent space
193 and improves the Kullback–Leibler (KL) divergence.

194 Despite its importance, few studies have integrated multi-scale features into
195 forecast-based anomaly detection. This paper focuses on addressing this gap through
196 our proposed framework.

197 **3 Methods**

198 In this section, we propose MAF-LSTM, a Multi-Scale Feature-fusion Anomaly Detec-
199 tion method for time series. MAF-LSTM employs a LSTM-based prediction structure
200 which effectively captures non-linear features at multiple scales, and finally make
201 comprehensive decisions through an attention-based feature fusion strategy.

202 **3.1 System Overview**

203 As illustrated in **Fig. 2**, the proposed MAF-LSTM consists of three key components.
204 The first component is the Multi-scale Feature Extraction module, which utilizes a
205 hierarchical LSTM ensemble to encode the standardized time series. This structure
206 captures temporal dependencies at various scales, preserving both short-term and
207 long-term features. These multi-scale features are then processed by the Attention-
208 based Fusion module. This component adaptively weights and integrates the features
209 into a unified representation for next-step prediction, ensuring that critical temporal
210 scales contribute more significantly to the output. Finally, the Anomaly Scoring mod-
211 ule quantifies the abnormal of the sequence. It calculates the prediction residuals and
212 employs a Mahalanobis-distance-based metric to generate the final anomaly scores.

213 **3.2 Data Collection and Preprocessing**

214 In practical industrial deployments, massive measurements are continuously collected
215 from heterogeneous field sensors and must be preprocessed to provide reliable and

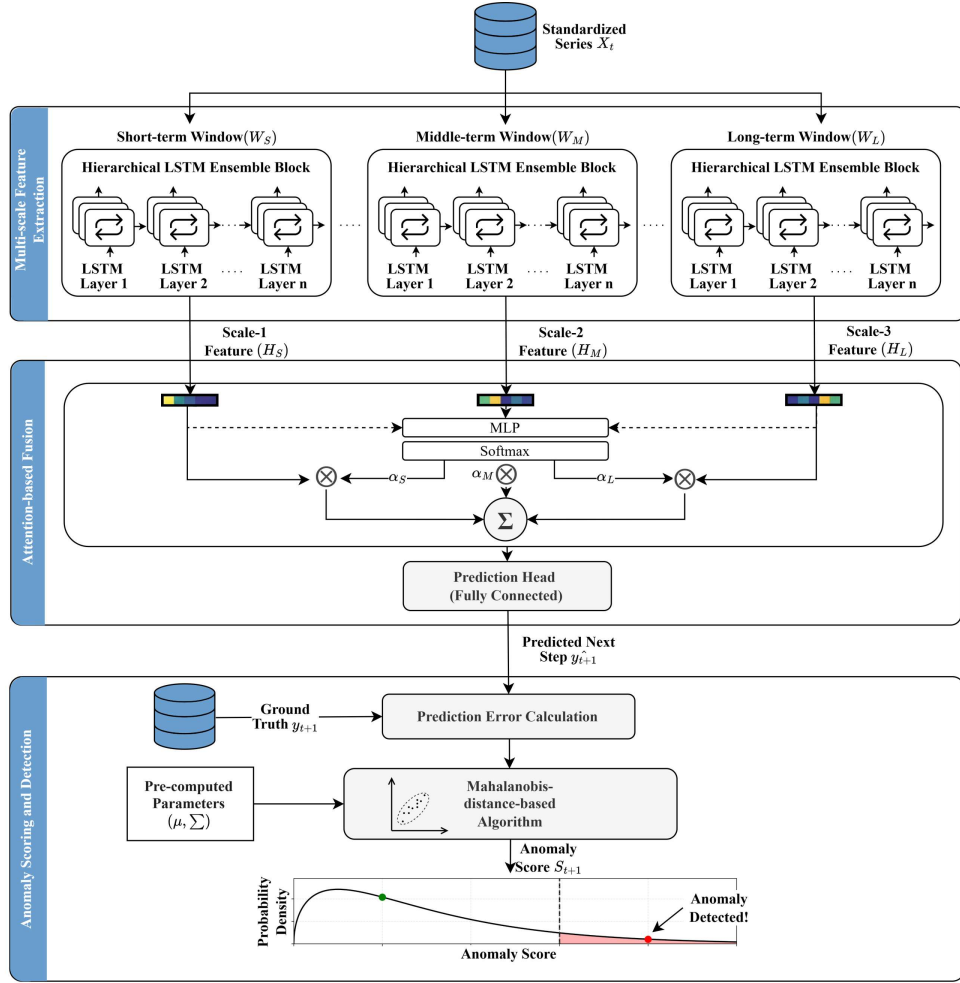


Fig. 2 Overall workflow of MAF-LSTM.

216 consistent inputs. Because these sensors operate asynchronously and often sample at
 217 irregular intervals, raw records are discontinuous and temporally misaligned [21]. To
 218 ensure consistency, we standardize timestamps, normalize measurement formats, and
 219 filter out erroneous or redundant records. Missing or inconsistent values are imputed
 220 based on historical observations, which improves robustness without distorting tem-
 221 poral trends. After internal consistency is guaranteed, measurements from different
 222 sources are temporally synchronized. For example, in power grid systems, switch-side
 223 and transformer-side observations are aligned by matching timestamps directly or by

224 nearest-neighbor assignment within a small tolerance window. The resulting dataset
 225 is clean, aligned, and restructured into a unified time-series representation indexed by
 226 timestamp and switch ID.

227 To stabilize model training, all numerical features were then normalized using
 228 Min-Max scaling, which can be formulated as

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \quad (1)$$

229 The normalized time series were further formatted according to the input structure
 230 required by the proposed multi-scale LSTM framework. Let the original sequence
 231 be denoted as $X = \{x_1, x_2, \dots, x_N\}$. To capture temporal dependencies at different
 232 resolutions, multiple window sizes $w_1, w_2, \dots, w_S$ were predefined. For a given window
 233 size w_s , the corresponding subsequence set is constructed as

$$X^{(s)} = \{X_1^{(s)}, X_2^{(s)}, \dots, X_{N-w_s+1}^{(s)}\}, \quad (2)$$

where $X_t^{(s)} = \{x_t, x_{t+1}, \dots, x_{t+w_s-1}\}.$

234 Each subsequence $X_t^{(s)}$ is treated as an input sample for the LSTM module at
 235 scale s , and the resulting segments are then organized into mini-batches for efficient
 236 training.

237 3.3 Multi-Scale LSTM for Feature Extraction

238 In order to better capture temporal dependencies across different time resolutions,
 239 we adopt a multi-scale feature extraction module based on sliding-window LSTM
 240 encoders. By aggregating information across multiple window sizes, the framework
 241 alleviates sensitivity to manually selected temporal scales, thereby improving both
 242 detection accuracy and robustness.

243 Building upon this design, the input sequence is processed under multiple window
 244 lengths $\{w_1, w_2, \dots, w_k\}$, where each window size corresponds to a distinct temporal

245 scale. Shorter windows emphasize local fluctuations and rapid transitions, while longer
 246 windows preserve long-range contextual dependencies. For each temporal scale, an
 247 independent LSTM network is employed to capture sequential dynamics, producing
 248 a hidden representation $h^{(s)}$ for scale s .

249 These scale-specific representations are subsequently aligned and aggregated
 250 through concatenation to form a unified feature vector. Formally,

$$H = [h^{(1)}, h^{(2)}, \dots, h^{(k)}]. \quad (3)$$

251 To prevent dominance from any single scale, a normalization layer is applied before
 252 feature fusion. The multi-scale representation is then fed to fusion module for next
 253 steps. In this way, the proposed multi-scale LSTM extractor enables the model to
 254 jointly capture both short-term irregularities and long-term deviations, which are
 255 essential for accurate anomaly detection in complex temporal patterns.

256 Although LSTM networks are adopted in our design to model sequence dynamics,
 257 the proposed multi-scale framework is inherently model-agnostic and can be seamlessly
 258 integrated with any prediction or reconstruction architecture that relies on window-
 259 based feature extraction.

260 3.4 Attention-based Model for Feature Fusion

261 To effectively integrate multi-scale temporal features while maintaining computational
 262 efficiency, we introduce a lightweight scale-level attention mechanism. This design is
 263 particularly suitable for large-scale power systems, where model complexity must be
 264 carefully controlled.

265 Given the multi-scale hidden representations $H = [h^{(1)}, h^{(2)}, \dots, h^{(k)}]$, these fea-
 266 tures are first concatenated and passed through a single-layer fully connected network
 267 with a tanh activation to produce a scale-level score vector $\mathbf{e} = [e_1, e_2, \dots, e_s]$, where
 268 $\mathbf{e}$ can be denoted as

$$\mathbf{e} = \tanh(W \cdot [h_t^1; h_t^2; \dots; h_t^s] + b). \quad (4)$$

Each element e_i in $\mathbf{e}$ reflects the contribution of the i -th temporal scale to the current time step t . To obtain a normalized importance distribution, the score vector is transformed via a softmax operation, yielding the attention weights $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_s]$, in which we have

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^s \exp(e_j)}. \quad (5)$$

The fused representation H_t is then computed as a weighted aggregation over all scales:

$$H_t = \sum_{i=1}^s \alpha_i \cdot h_t^i. \quad (6)$$

This formulation allows the model to adaptively emphasize informative temporal scales while down-weighting less relevant ones, thereby achieving dynamic multi-scale feature fusion.

Finally, the aggregated feature H_t is fed into a fully connected layer to generate the prediction $\hat{y}_t$ for the current time step according to

$$\hat{y}_t = \text{Dense}(H_t). \quad (7)$$

During training, both the LSTM encoders and the attention fusion module are optimized in an end-to-end manner under an unsupervised learning setting. The model parameters are updated based on the prediction error between $\hat{y}_t$ and the ground truth y_t , without requiring manually labeled anomalies. Considering the scarcity and labeling difficulty of anomaly data, this unsupervised framework is well-suited for practical deployment.

The training objective is to minimize the Mean Absolute Error (MAE), which is formulated as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (8)$$

Unlike the Mean Squared Error (MSE), MAE penalizes deviations linearly, preventing large errors from dominating the optimization and mitigating the effect of noisy samples in the training data.

3.5 Anomaly Score via Mahalanobis Distance and Detection

After the predictive model is trained, the test series $X_{test} = [x_1, x_2, \dots, x_N]$ is fed into the network for inference. For each time step t , the model generates a forecast value $\hat{y}_t$ based on the multi-scale fused representation H_t . The prediction residual, defined as $r_t = |y_t - \hat{y}_t|$, serves as the primary basis for identifying anomalies.

To quantify deviations in a statistically principled manner, we employ a Mahalanobis distance-based anomaly scoring strategy. Unlike simple point-wise error metrics (e.g. MSE), this method explicitly accounts for the inter-variable correlations within the residual space. By modeling the joint distribution, it provides a more robust characterization of multivariate deviations, distinguishing between logical anomalies and normal system noise.

We first estimate the statistical parameters using the entire validation set, which represents the system's normal operating modes. Let $\mathcal{E}_{val}$ denote the set of residual vectors from the validation data. We compute the global empirical mean vector μ and covariance matrix Σ as follows:

$$\begin{aligned} \mu &= \frac{1}{|\mathcal{E}_{val}|} \sum_{r \in \mathcal{E}_{val}} r, \\ \Sigma &= \frac{1}{|\mathcal{E}_{val}| - 1} \sum_{r \in \mathcal{E}_{val}} (r - \mu)(r - \mu)^T. \end{aligned} \quad (9)$$

This global covariance matrix Σ captures the holistic correlation structure of the system under normal conditions.

During inference, we obtain an anomaly score for each residual r_t according to the Mahalanobis distance within its corresponding interval, which is calculated as

$$S_t = \sqrt{(r_t - \mu)^T \Sigma^{-1} (r_t - \mu)}. \quad (10)$$

A higher S_t indicates a larger deviation from the normal residual distribution.

Finally, a data point x_t is classified as anomalous if its score exceeds a predefined threshold τ , which is empirically determined to maximize the F1-score on the validation set. Formally,

$$x_t \text{ is anomalous if } S_t > \tau. \quad (11)$$

In summary, this Mahalanobis-based method offers a distinct advantage over simple error-based metrics: it explicitly checks whether the variables are changing together in the correct way. While traditional methods only look at the magnitude of individual errors, our approach captures the correlations between features using the global covariance matrix. This allows the system to detect 'logical' anomalies—such as when two related sensors become out of sync—even if their individual values appear normal. This capability is critical for ensuring the safety of interconnected systems.

4 Experiments and Results

To evaluate the effectiveness of MAF-LSTM, we conducted extensive experiments on five real-world time series datasets from diverse industrial domains. The primary objective of these experiments is to validate the framework’s capability in enhancing anomaly detection accuracy across different practical scenarios.

4.1 Datasets

We evaluate our framework on five widely-used real-world multivariate time series datasets: MSL, SMAP, SMD, SWaT, and PSM. The datasets originate from three

types of engineering systems, namely spacecraft telemetry (MSL, SMAP), server monitoring (SMD, PSM), and an industrial control system (SWaT), and they share the characteristics that motivate the proposed framework: multiple correlated channels, long recording periods, and anomalies that span different temporal scales. Detailed descriptions of these datasets are provided as follows:

- **MSL (Mars Science Laboratory)** [11]: A real telemetry dataset from NASA’s Mars rover containing labeled point and contextual anomalies across 27 entities with 55 dimensions.
- **SMAP (Soil Moisture Active Passive)** [11]: A NASA dataset of soil moisture and remote sensing measurements with labeled anomalies over 55 entities and 25 dimensions.
- **SMD (Server Machine Dataset)** [13]: A dataset consists of multivariate time series collected from 28 servers in a large-scale internet company over five weeks, each with 38 dimensions.
- **SWaT (Secure Water Treatment)** [20]: A cyber-physical system dataset collected from a real water treatment testbed, containing 11 days of sensor and actuator data with 51 dimensions.
- **PSM (Pooled Server Metrics)** [1]: A real-world IT system dataset from eBay that includes multivariate server performance metrics with 26 dimensions for anomaly detection.

Table 1 summarizes the key statistics of these datasets.

4.2 Experimental Setup

We implemented MAF-LSTM using Python 3.6, with Tensorflow 1.9 for the LSTM and Attention model. All experiments were conducted on a computing platform equipped with an Intel Core i7-12700H CPU (2.30 GHz), an NVIDIA GeForce RTX 4050 Laptop GPU, and 16 GB of RAM.

Table 1 Statistics of the datasets.

Dataset	Dim	Length	Anomaly (%)
MSL	55	132046	10.72
SMAP	25	562800	13.13
SMD	38	1416825	4.16
SWaT	51	946719	11.98
PSM	26	220322	27.8

For all deep learning models, the Adam optimizer is employed with an initial learning rate of 0.001. Each LSTM module in the proposed multi-scale feature fusion model consists of 2 hidden layers, each with a hidden size of 64. The multi-scale sliding windows are set to $\{20, 30, 50, 70, 100\}$, and the batch size is 32. We employ the early stopping strategy within 50 epochs during training to prevent overfitting. For all datasets, we follow their original training/test split, and 10% of the training data is further used as a validation set for hyperparameter tuning. All metrics are normalized using min-max scaling within each dataset.

During inference, the model assigns an anomaly score to each time point, where a larger value indicates a higher degree of abnormality. We further adopt the point adjustment strategy [34] to enhance anomaly detection performance. In real-world scenarios, when contextual anomalies occur, it is sufficient for the model to detect at least one abnormal point within the anomalous interval, rather than every anomalous time step. This strategy has been proven effective in many previous works [16, 35].

For comparison, we evaluated MAF-LSTM against four baseline approaches representing different categories of anomaly detection strategies:

- **iForest** [15]: A widely adopted unsupervised algorithm that identifies anomalies by isolating instances through recursive random partitioning.

- 373 • **DAGMM** [41]: A deep autoencoding Gaussian mixture model that jointly opti-
 374 mizes dimensionality reduction and density estimation for unsupervised anomaly
 375 detection.
- 376 • **LSTM-VAE** [23]: A recurrent variational autoencoder that models temporal
 377 dependencies in time series to detect anomalies based on reconstruction probability.
- 378 • **OmniAnomaly** [27]: A representative stochastic model that combines a gated
 379 recurrent unit (GRU) and VAE to capture the temporal dependence of multivariate
 380 time series.

381 For performance evaluation, we compute **Precision(P)**, **Recall(R)**, and **F1 score**
 382 by setting a threshold on the anomaly score for each compared algorithm. These
 383 metrics are defined as follows:

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}, \quad F_1 = \frac{2 \times P \times R}{P + R}. \quad (12)$$

384 TP , FP , and FN respectively denote the numbers of *true positives*,
 385 *false positives*, and *false negatives*. To ensure a fair comparison, we search for
 386 the optimal threshold within a feasible range and report the best performance
 387 corresponding to the highest F1-score.

388 4.3 Performance Comparison

389 **Table 2** shows the results on the MSL, SMAP, SMD, SWaT and PSM. As shown
 390 in results, iForest performs worse than the deep learning methods, mainly because it
 391 cannot capture the non-linear temporal dependencies in multivariate sequences. MAF-
 392 LSTM performs well on most datasets, while OmniAnomaly usually performs better
 393 than the other baselines but still produces sub-optimal results overall.

394 In general, MAF-LSTM outperforms all other methods on most datasets (except
 395 SMAP), achieving an average F1-score improvement of 2.52% over the best-
 396 performing baseline. These results demonstrate that integrating multi-scale features

Table 2 Performance on MSL, SMAP, SMD, SWaT and PSM.

Dataset	Metric	MAF-LSTM	iForest	DAGMM	LSTM-VAE	OmniAnomaly
MSL	Precision	92.19	50.68	87.51	85.76	87.01
	Recall	89.70	81.43	61.93	70.94	82.50
	F1-Score	90.93	62.48	72.53	77.65	84.69
SMAP	Precision	85.37	50.37	86.42	89.90	91.39
	Recall	58.23	51.07	58.79	61.75	79.99
	F1-Score	69.24	50.72	69.98	73.20	85.31
SMD	Precision	84.32	41.36	66.28	73.33	83.54
	Recall	98.92	70.29	50.87	89.08	86.72
	F1-Score	91.04	52.08	57.56	80.44	85.10
SWaT	Precision	84.51	48.19	87.86	75.00	79.02
	Recall	91.27	44.83	36.84	87.49	83.40
	F1-Score	87.76	46.45	51.91	80.76	81.15
PSM	Precision	96.63	75.08	90.38	70.61	86.47
	Recall	92.49	90.45	72.13	85.29	73.26
	F1-Score	94.51	82.05	80.23	77.25	79.32

Bold values indicate the highest value in each row.

with an attention mechanism effectively enhances anomaly detection performance. The improvement on the PSM dataset is more noticeable than that on the other datasets, on which MAF-LSTM achieves an F1-score 12.46 percentage points higher than that of iForest.

4.4 Ablation Study

Furthermore, we perform an ablation study to evaluate the contribution of multi-scale feature fusion and the attention mechanism. Specifically, each component is removed or replaced, and the model is retrained to observe the performance difference. Four ablation variants are considered:

- 406 • **MAF-LSTM:** Our proposed full model, which includes both multi-scale feature
407 extraction and the attention-based feature fusion module.
- 408 • **MAF-LSTM (Eu):** Based on the proposed MAF-LSTM model, this variant
409 replaces the Mahalanobis distance with the Euclidean distance for anomaly scoring.
410 Given two vectors $\mathbf{x} = (x_1, x_2, \dots, x_n)$ and $\mathbf{y} = (y_1, y_2, \dots, y_n)$, the Euclidean
411 distance is defined as:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} = \|\mathbf{x} - \mathbf{y}\|_2. \quad (13)$$

- 412 • **Multi-scale Equal Model:** A simplified framework that replaces the attention
413 fusion module by assigning equal weights to all scales.
- 414 • **Single-scale Model:** A simplified framework that replaces the multi-scale LSTM
415 module with a fixed single LSTM.

416 We evaluate each variant using F1-score, AU-ROC (Area Under the Receiver Oper-
417 ating Characteristic Curve), and AU-PR (Area Under the Precision-Recall Curve).
418 F1-score reflects the balance between precision and recall. AU-ROC measures the
419 ability to distinguish anomalies from normal points across different thresholds, while
420 AU-PR focuses on the precision-recall trade-off under imbalanced conditions. These
421 metrics provide a comprehensive assessment of detection accuracy and robustness.

422 The results on the MSL dataset are reported in **Table 3**. Furthermore, the varia-
423 tion of F1, AU-ROC and AU-PR with window size are reported in **Fig. 3**. As shown,
424 removing either module leads to a clear degradation, and both components contribute
425 more than a 4% improvement in F1-score. This shows that the multi-scale feature
426 fusion and attention mechanism are not just extra modules, but are important for
427 improving detection accuracy.

428 To further understand the contribution of each component, we analyze the abla-
429 tion results from three perspectives. First, using multi-scale windows gives better

Table 3 Comparisons of variants using different structures.

Method	Multi-scale	Attention Fusion	Mahalanobis	F1-score	AU-ROC	AU-PR
MAF-LSTM	✓	✓	✓	90.93	70.32	20.41
MAF-LSTM (Eu)	✓	✓	×	74.79	61.88	20.04
Multi-scale Equal	✓	×	✓	86.81	67.40	18.32
Single-scale	×	×	✓	86.49	65.68	18.10

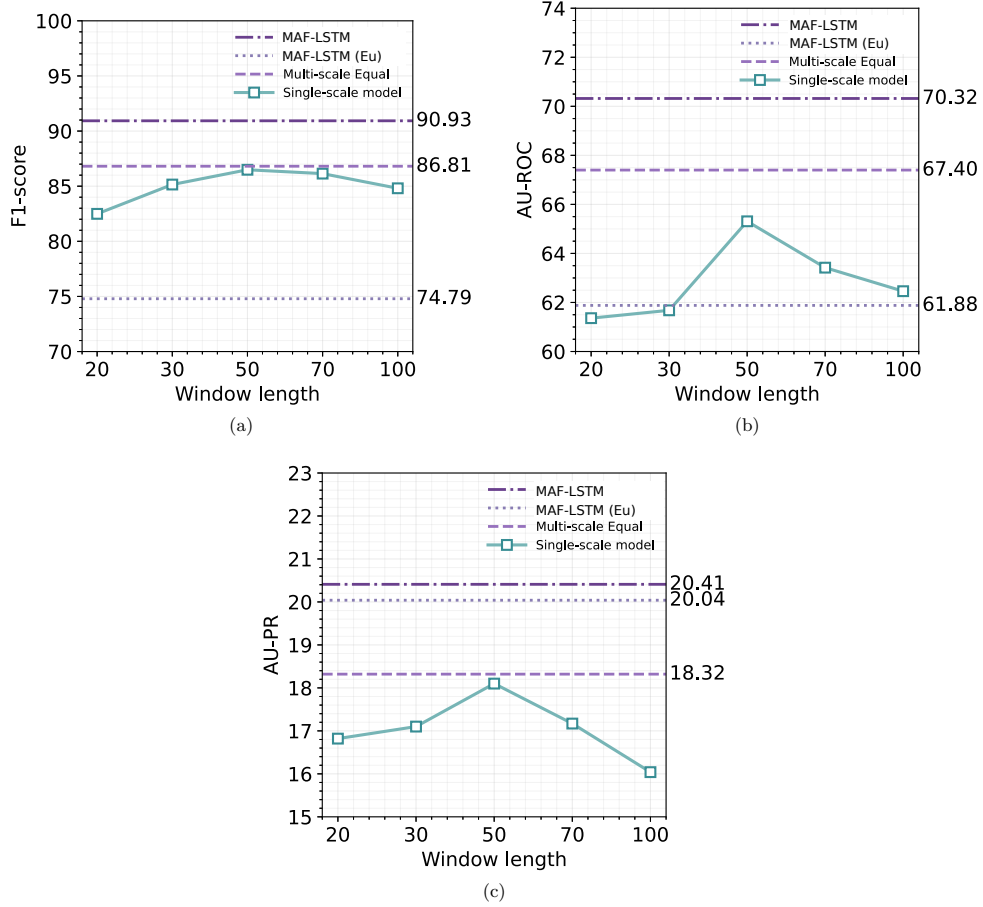


Fig. 3 Performance under varying window lengths. (a) F1-score; (b) AU-ROC; (c) AU-PR.

430 results than the single-scale model. In **Fig. 3**, the results of single-scale model increase
 431 before window 50, and decrease after it. This shows that for single scale models, an
 432 appropriate window length is required: one that is sufficient to capture larger-scale
 433 patterns but not so large as to introduce redundancy or noise. Without specifically
 434 choosing the most appropriate scale, MAF-LSTM improves the F1-score by 4.44 com-
 435 pared with the best result of single-scale model, showing that features from different
 436 time scales are complementary and help the detector capture richer patterns. In addi-
 437 tion, adding the attention-based fusion module is more effective than simply add scale
 438 representations. Compared with the multi-scale equal variant, MAF-LSTM increases
 439 the F1-score by 4.12. This means the attention mechanism helps the model focus on
 440 important features, instead of treating all scales equally. Finally, using Mahalanobis
 441 distance gives higher F1-scores than Euclidean distance. All methods that use the
 442 proposed anomaly score computing perform better than MAF-LSTM (Eu), showing
 443 that Mahalanobis distance fits the residual distribution more accurately and provides
 444 more reliable anomaly scores.

445 **4.5 Discussion and limitations**

446 The experimental results support the two design premises of the framework. The
 447 ablation study shows that processing the measurement sequence at multiple scales
 448 improves detection over any single scale, and that adaptively weighting the scales
 449 through attention is more effective than combining them with equal weights. The
 450 largest single contribution comes from Mahalanobis scoring, which indicates that
 451 accounting for the correlation structure of the forecasting residuals is at least as impor-
 452 tant as the choice of temporal representation. These findings are consistent with the
 453 multi-scale nature of anomalies in engineering monitoring data described in Section 1.

454 Several limitations should be noted. First, although the framework is motivated
 455 in part by the monitoring of power equipment, it has been evaluated only on public

456 benchmark datasets from other engineering domains; validation on measurement data
457 from power equipment in service is the subject of ongoing work. Second, the window
458 lengths of the multi-scale branches were fixed at 20, 30, 50, 70, 100 for all datasets. An
459 adaptive selection of scales according to the sampling rate and the dominant periods
460 of the measured variables may further improve performance and reduce computational
461 cost. Third, the framework did not achieve the best result on SMAP, which indicates
462 that the benefit of multi-scale forecasting depends on the temporal extent of the
463 anomalies present in the data. Fourth, the covariance matrix used in Mahalanobis
464 scoring is estimated from the training residuals and is not updated online; over long
465 operating periods, gradual changes in equipment behavior may require the matrix to
466 be re-estimated periodically.

467 5 Conclusions

468 This paper presents MAF-LSTM, a multi-scale anomaly detection framework designed
469 for forecasting-based anomaly detection in multivariate monitoring data. The pro-
470 posed method employs a multi-scale LSTM forecasting network to capture temporal
471 dependencies across different horizons, while an attention-based feature fusion module
472 dynamically adjusts feature importance over time. By leveraging the Mahalanobis dis-
473 tance of prediction errors as anomaly scores, MAF-LSTM enables accurate detection
474 of abnormal patterns. Extensive experiments on five real-world multivariate datasets
475 demonstrate that MAF-LSTM outperforms state-of-the-art approaches in terms of
476 detection accuracy and robustness. In future work, we will further explore adaptive
477 scale selection mechanisms and investigate the extension of MAF-LSTM to real-time
478 streaming environments.

Abbreviations

AE	Autoencoder
ARMA	Autoregressive Moving Average
AU-PR	Area Under the Precision–Recall Curve
AU-ROC	Area Under the Receiver Operating Characteristic Curve
CPU	Central Processing Unit
DAGMM	Deep Autoencoding Gaussian Mixture Model
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
FN	False Negative
FP	False Positive
GAN	Generative Adversarial Network
GNN	Graph Neural Network
GPU	Graphics Processing Unit
GRU	Gated Recurrent Unit
HTM	Hierarchical Temporal Memory
iForest	Isolation Forest
KL	Kullback–Leibler
LSTM	Long Short-Term Memory
LSTM-AD	Long Short-Term Memory Anomaly Detection
LSTM-AE	Long Short-Term Memory Autoencoder
LSTM-NDT	Long Short-Term Memory with Nonparametric Dynamic Thresholding
LSTM-VAE	Long Short-Term Memory Variational Autoencoder
MAE	Mean Absolute Error
MAF-LSTM	Multi-scale Attention-based Feature-fusion Long Short-Term Memory
MSE	Mean Squared Error
MSL	Mars Science Laboratory
NASA	National Aeronautics and Space Administration
NDT	Nonparametric Dynamic Thresholding
PSM	Pooled Server Metrics
RAM	Random Access Memory
RNN	Recurrent Neural Network
SAnD	Simply Attend and Diagnose
SMAP	Soil Moisture Active Passive
SMD	Server Machine Dataset
SWaT	Secure Water Treatment
TP	True Positive
VAE	Variational Autoencoder

481 **Ethics approval and consent to participate**

482 Not applicable.

483 **Consent for publication**

484 Not applicable.

485 **Data availability**

486 The five datasets analyzed during the current study are publicly available. MSL and
487 SMAP are available from the NASA Telemnom repository [11]. SMD is available
488 from the OmniAnomaly repository [27]. SWaT is available from the iTrust Centre for
489 Research in Cyber Security, Singapore University of Technology and Design, upon
490 request [20]. PSM is available from the eBay RANSynCoders repository [1].

491 **Competing interests**

492 The authors declare that they have no competing interests.

493 **Funding**

494 Sponsored by "Research on Deep Learning-based Measurement Data Anomaly Detec-
495 tion Technology (2025 State Grid Jiangsu Electric Power Co., Ltd., Information &
496 Communication Branch 'Technology Project Package for the Production Frontline')".

497 **Authors' contributions**

498 All authors contributed to the study conception and design. WL and ML implemented
499 the framework and performed the experiments. FX, QZ, SW, and WY contributed
500 to data analysis and manuscript revision. All authors read and approved the final
501 manuscript.

502 Acknowledgments

503 Not applicable.

504 References

- 505 [1] Abdulaal A, Liu Z, Lancewicki T (2021) Practical Approach to Asynchronous
506 Multivariate Time Series Anomaly Detection and Localization. In: Proceedings
507 of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining.
508 ACM, Virtual Event Singapore, pp 2485–2494, [https://doi.org/10.1145/344754](https://doi.org/10.1145/3447548.3467174)
509 [8.3467174](https://doi.org/10.1145/3447548.3467174)
- 510 [2] Ahmad S, Lavin A, Purdy S, et al (2017) Unsupervised real-time anomaly detec-
511 tion for streaming data. Neurocomputing 262:134–147. [https://doi.org/10.1016/](https://doi.org/10.1016/j.neucom.2017.04.070)
512 [j.neucom.2017.04.070](https://doi.org/10.1016/j.neucom.2017.04.070)
- 513 [3] Audibert J, Michiardi P, Guyard F, et al (2020) USAD: UnSupervised Anomaly
514 Detection on Multivariate Time Series. In: Proceedings of the 26th ACM
515 SIGKDD International Conference on Knowledge Discovery & Data Mining.
516 ACM, Virtual Event CA USA, pp 3395–3404, [https://doi.org/10.1145/3394486.](https://doi.org/10.1145/3394486.3403392)
517 [3403392](https://doi.org/10.1145/3394486.3403392)
- 518 [4] Breunig MM, Kriegel HP, Ng RT, et al (2000) LOF: identifying density-based
519 local outliers. SIGMOD Rec 29(2):93–104. [https://doi.org/10.1145/335191.335](https://doi.org/10.1145/335191.335388)
520 [388](https://doi.org/10.1145/335191.335388)
- 521 [5] Chen Z, Chen D, Zhang X, et al (2021) Learning Graph Structures with Trans-
522 former for Multivariate Time Series Anomaly Detection in IoT. IEEE Internet
523 Things J 9(12):9179–9189. <https://doi.org/10.48550/arXiv.2104.03466>
- 524 [6] Ester M, Kriegel HP, Sander J, et al (1996) A density-based algorithm for
525 discovering clusters in large spatial databases with noise. In: Proc. 2nd Int.
526 Conf. Knowledge Discovery and Data Mining (KDD’96), Portland, OR, USA, pp

- 226–231, <https://doi.org/10.5555/3001460.3001507>
- [7] Fox AJ (1972) Outliers in Time Series. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 34(3):350–363. <https://doi.org/10.1111/j.2517-6161.1972.tb00912.x>
- [8] H. L. S. (1939) A Study in the Analysis of Stationary Time Series. *J Inst Actuar* 70(1):113–115. <https://doi.org/10.1017/S0020268100011574>
- [9] Hardin J, Rocke DM (2004) Outlier detection in the multiple cluster setting using the minimum covariance determinant estimator. *Computational Statistics & Data Analysis* 44(4):625–638. [https://doi.org/10.1016/S0167-9473\(02\)00280-3](https://doi.org/10.1016/S0167-9473(02)00280-3)
- [10] Hochreiter S, Schmidhuber J (1997) Long Short-Term Memory. *Neural Computation* 9(8):1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [11] Hundman K, Constantinou V, Laporte C, et al (2018) Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. *arXiv preprint arXiv:1802.04431*, <https://doi.org/10.48550/arXiv.1802.04431>
- [12] Li J, Izakian H, Pedrycz W, et al (2021) Clustering-based anomaly detection in multivariate time series data. *Applied Soft Computing* 100:106919. <https://doi.org/10.1016/j.asoc.2020.106919>
- [13] Li Z, Chen W, Pei D (2018) Robust and Unsupervised KPI Anomaly Detection Based on Conditional Variational Autoencoder. In: 2018 IEEE 37th International Performance Computing and Communications Conference (IPCCC), pp 1–9, <https://doi.org/10.1109/IPCCC.2018.8710885>
- [14] Liang H, Song L, Du J, et al (2022) Consistent Anomaly Detection and Localization of Multivariate Time Series via Cross-Correlation Graph-Based Encoder–Decoder GAN. *IEEE Trans Instrum Meas* 71:1–10. <https://doi.org/10.1109/TIM.2021.3139696>

- [15] Liu FT, Ting KM, Zhou ZH (2012) Isolation-Based Anomaly Detection. ACM Trans Knowl Discov Data 6(1):1–39. <https://doi.org/10.1145/2133360.2133363>
- [16] Liu Q, Liu C, Sun S, et al (2025) GDformer: Going Beyond Subsequence Isolation for Multivariate Time Series Anomaly Detection. arXiv preprint arXiv:2501.18196, <https://doi.org/10.48550/arXiv.2501.18196>
- [17] Ma J, Perkins S (2003) Time-series novelty detection using one-class support vector machines. In: Proceedings of the International Joint Conference on Neural Networks, pp 1741–1745, <https://doi.org/10.1109/IJCNN.2003.1223670>
- [18] Malhotra P, Ramakrishnan A, Anand G, et al (2016) LSTM-based Encoder-Decoder for Multi-sensor Anomaly Detection. arXiv preprint arXiv:1607.00148, <https://doi.org/10.48550/arXiv.1607.00148>
- [19] Malhotra P, et al (2015) Long short term memory networks for anomaly detection in time series. In: Proceedings, URL https://books.google.co.uk/books?hl=zh-CN&lr=&id=USGLCgAAQBAJ&oi=fnd&pg=PA89&dq=Long+short+term+memory+networks+for+anomaly+detection+in+time+series.&ots=FuibgmNWWI&sig=TL06_vIAriN9B3fdlX4HPhVhA2E
- [20] Mathur AP, Tippenhauer NO (2016) 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), pp 31–36, <https://doi.org/10.1109/CySWater.2016.7469060>
- [21] Neil D, Pfeiffer M, Liu SC (2016) Phased LSTM: Accelerating Recurrent Network Training for Long or Event-based Sequences. In: Adv. Neural Inf. Process. Syst. (NeurIPS), URL https://proceedings.neurips.cc/paper_files/paper/2016/file/5bce843dd76db8c939d5323dd3e54ec9-Paper.pdf
- [22] Pan X, Tan J, Kavulya S, et al (2010) Ganesha: blackBox diagnosis of MapReduce systems. SIGMETRICS Perform Eval Rev 37(3):8–13. <https://doi.org/10.1145/>

- 579 [23] Park D, Hoshi Y, Kemp CC (2018) A Multimodal Anomaly Detector for Robot-
580 Assisted Feeding Using an LSTM-based Variational Autoencoder. *IEEE Robot*
581 *Autom Lett* 3(3):1544–1551. <https://doi.org/10.48550/arXiv.1711.00614>
- 582 [24] Qingning L, et al (2023) Multi-Scale Anomaly Detection for Time Series with
583 Attention-based Recurrent Autoencoders. In: *Proceedings of The 14th Asian*
584 *Conference on Machine Learning*. PMLR, pp 674–689, URL [https://proceeding](https://proceedings.mlr.press/v189/qingning23a.html)
585 [s.mlr.press/v189/qingning23a.html](https://proceedings.mlr.press/v189/qingning23a.html), accessed 28 October 2025
- 586 [25] Song H, Rajan D, Thiagarajan JJ, et al (2018) Attend and Diagnose: Clinical
587 Time Series Analysis using Attention Models. In: *Proc. AAAI Conf. Artificial*
588 *Intelligence*, <https://doi.org/10.48550/arXiv.1711.03905>
- 589 [26] Song Y, Xin R, Chen P, et al (2023) Identifying performance anomalies in fluctu-
590 ating cloud environments: A robust correlative-GNN-based explainable approach.
591 *Future Generation Computer Systems* 145:77–86. [https://doi.org/10.1016/j.futu](https://doi.org/10.1016/j.future.2023.03.020)
592 [re.2023.03.020](https://doi.org/10.1016/j.future.2023.03.020)
- 593 [27] Su Y, Zhao Y, Niu C, et al (2019) Robust Anomaly Detection for Multivariate
594 Time Series through Stochastic Recurrent Neural Network. In: *Proceedings of the*
595 *25th ACM SIGKDD International Conference on Knowledge Discovery & Data*
596 *Mining*. ACM, Anchorage AK USA, pp 2828–2837, [https://doi.org/10.1145/32](https://doi.org/10.1145/3292500.3330672)
597 [92500.3330672](https://doi.org/10.1145/3292500.3330672)
- 598 [28] Tsay RS (1988) Outliers, level shifts, and variance changes in time series. *Journal*
599 *of Forecasting* 7(1):1–20. <https://doi.org/10.1002/for.3980070102>
- 600 [29] Wang F, Jiang Y, Zhang R, et al (2025) A Survey of Deep Anomaly Detection
601 in Multivariate Time Series: Taxonomy, Applications, and Directions. *Sensors*
602 25(1):190. <https://doi.org/10.3390/s25010190>

- [30] Wang X, Pi D, Zhang X, et al (2022) Variational transformer-based anomaly detection approach for multivariate time series. *Measurement* 191:110791. <https://doi.org/10.1016/j.measurement.2022.110791>
- [31] Wang Y, Zhang J, Guo S, et al (2021) Decoupling Representation Learning and Classification for GNN-based Anomaly Detection. In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, Virtual Event Canada, pp 1239–1248, <https://doi.org/10.1145/3404835.3462944>
- [32] Wang Z, Chen Z, Ni J, et al (2021) Multi-Scale One-Class Recurrent Neural Networks for Discrete Event Sequence Anomaly Detection. In: *Proc. 27th ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD)*, pp 3726–3734, <https://doi.org/10.1145/3447548.3467125>
- [33] Xiong W, et al (2014) Anomaly secure detection methods by analyzing dynamic characteristics of the network traffic in cloud communications. *Information Sciences* 258:403–415. <https://doi.org/10.1016/j.ins.2013.04.009>
- [34] Xu H, et al (2018) Unsupervised Anomaly Detection via Variational Auto-Encoder for Seasonal KPIs in Web Applications. *arXiv preprint arXiv:1802.03903*, <https://doi.org/10.48550/arXiv.1802.03903>
- [35] Yang Y, Zhang C, Zhou T, et al (2023) DCdetector: Dual Attention Contrastive Representation Learning for Time Series Anomaly Detection. In: *Proc. 29th ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD)*, pp 3033–3045, URL <https://arxiv.org/pdf/2306.10347>
- [36] Yeh CCM, et al (2016) Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets. In: *2016 IEEE 16th International Conference on Data Mining (ICDM)*, Barcelona, pp 1317–1322, <https://doi.org/10.1109/ICDM.2016.0179>

- 629 [37] Yu X, et al (2025) Adversarial Transformer-Based Anomaly Detection for Mul-
630 tivariate Time Series. *IEEE Trans Ind Inf* 21(3):2471–2480. <https://doi.org/10.1109/TII.2024.3507211>
631
- 632 [38] Zamanzadeh Darban, Webb GI, Pan S, et al (2025) Deep Learning for Time
633 Series Anomaly Detection: A Survey. *ACM Comput Surv* 57(1):1–42. <https://doi.org/10.1145/3691338>
634
- 635 [39] Zhang C, et al (2019) A Deep Neural Network for Unsupervised Anomaly Detec-
636 tion and Diagnosis in Multivariate Time Series Data. In: *Proc. AAAI Conf.*
637 *Artificial Intelligence*, pp 1409–1416, URL <https://aaai.org/ojs/index.php/AAAI/article/view/3942/3820>
638
- 639 [40] Zhao Z, Xiao Z, Tao J (2024) MSDG: Multi-Scale Dynamic Graph Neural
640 Network for Industrial Time Series Anomaly Detection. *Sensors* 24(22):7218.
641 <https://doi.org/10.3390/s24227218>
- 642 [41] Zong B, Song Q, Min MR, et al (2018) Deep autoencoding Gaussian mix-
643 ture model for unsupervised anomaly detection. In: *Proc. Int. Conf. Learning*
644 *Representations (ICLR)*, URL <https://openreview.net/forum?id=BJJLHbb0->