Journal of Engineering, Project, and Production Management
2026, 16(4), 2025-237

# Multi-Scale Convolutional VAE for Anomaly Detection in Sensitive Power Marketing Data

Jinkai Sun[1], Chun Xiao[2], and Junfeng Yao[3]
[1] Specialist, Marketing Service Center, State Grid Shanxi Electric Power Company, Taiyuan, 030012, China, E-mail: JinkaiS_un@outlook.com (corresponding author).
[2] Senior Engineer, State Grid ShanXi Marketing Service Center, Taiyuan, 030012, China, E-mail: tyutxiaochun@163.com
[3] Senior Engineer, State Grid ShanXi Marketing Service Center, Taiyuan, 030012, China, E-mail: 785078849@qq.com

Project Management
Received October 20025; revised December 24, 2025; accepted March 22, 2026
Available online May 29, 2026

---

**Abstract:** Anomalous sensitive power marketing data can significantly impact the interests of various stakeholders in the development of power enterprises. To improve anomaly detection in such data, we propose a method based on the Multi-Scale Convolutional Variational Autoencoder (MSCVAE). First, a multi-scale attribute matrix is constructed to represent the system's state across varying time intervals within the multivariate time series of sensitive power marketing data. Next, a convolutional variational autoencoder functions as a generator of a reconstruction matrix from this attribute matrix, while an Attention-based Convolutional Long Short-Term Memory (ConvLSTM) is applied to find the temporal patterns. To address the class imbalance issue inherent in the sensitive power marketing dataset, a novel threshold-setting strategy derived from the confusion matrix is introduced. The proposed model is evaluated on two real-world sensitive power marketing datasets and compared with several benchmark models. The MSCVAE model exhibited superior performance relative to all benchmark models, which is evidenced by average increments of 42% in Precision, 40.9% in Recall, and 41.5% in F1-score.

**Keywords**: Sensitive power marketing data, multi-scale, ConvLSTM, anomaly detection.


DOI 10.32738/JEPPM-2025-237

---

## 1. Introduction

In the development of grids and electricity sector digitalization relying on the Internet, feature engineering and anomaly detection are necessary for preserving grid stability and operational efficiency. It is essential to uncover hidden patterns in sensitive power data through the examination of its irregular modes improves anomaly detection and optimization methods. Effective anomaly detection involves the analysis of power marketing regularities and bolsters stakeholder's capacity to manage anomalies in sensitive power datasets. For the above reason, developing efficient anomaly detection methods is becoming a key area of focus for energy industry research (Zhang et al., 2023). In this paper, the concept of "sensitive power marketing data" is applied to multivariate datasets that capture details of customer energy use, financial transactions and billing information, parameters for grid operation, and logs from demand response programs.

The process of anomaly detection in sensitive power marketing data begins with capturing anomalous features. Methods for feature engineering are used along with big data mining for semantic analysis and data reconstruction (Chen et al., 2022). Normal methods are spectral clustering algorithms, low-rank model-based anomaly detection (Jin et al., 2025), and corrections that leverage the Spark framework in power grid anomaly identification. However, traditional anomaly detection methods are becoming less effective in sensitive power marketing data due to the continuous evolution of power network's data pooling and the emergence of new equipment (Carlei et al., 2025; Li et al., 2024).

As the complexity of power marketing networks increases, so does the demand for advanced anomaly detection methods. To solve the limitations of conventional techniques, this paper proposes the latest anomaly detection method, the Multi-Scale Convolutional Variational Autoencoder (MSCVAE). By testing real-world sensitive power marketing data, MSCVAE can correctly identify anomalies and assess their severity. This enables a fast response to system anomaly and effectively minimizes operational risks (Chen and Cui 2025). To address the class imbalance in sensitive power marketing data, our principal contribution lies in a novel framework that merges multi-scale attribute matrices with an attention-augmented Convolutional Long Short-Term Memory (ConvLSTM) under a variational autoencoder structure.

Our proposed framework surpasses existing anomaly detection techniques in several key aspects, as evidenced by the following advancements:

1. The core of our design lies in a multi-scale attribute matrix that simultaneously encodes local and global behavioral patterns. This encoding leverages the calculation of pairwise relationships over multiple window sizes.

2. An attention-enhanced Convolutional Long Short-Term Memory (ConvLSTM) is embedded within our convolutional Variational Autoencoder (VAE) framework, facilitating the combined extraction of spatial features and long-range temporal patterns.

3. A novel thresholding strategy based on Error Rate (ERR) is developed, which demonstrates superior performance over traditional Receiver Operating Characteristic (ROC)-based techniques in scenarios with class imbalance.

These contributions, which serve to mitigate the deficiencies of established approaches, also strengthen the overall robustness and adaptability of anomaly detection frameworks as applied to power marketing.

## 2. Related Work

### 2.1. Deep Learning-based Anomaly Detection Methods

Three primary categories characterize deep learning-based time series anomaly detection: predictive, reconstructive, and representational. Reconstructive models often deliver superior accuracy in anomaly detection. Unlike predictive models, Reconstruction models have the advantage of accessing the entire time series dataset. This multivariable data access enables reconstructive models to effectively replicate scenarios and pinpoint discrepancies. These models are still the favored option, though they might lead to some latency in detection when high precision is of the essence, and minor delays, which can be accepted. So, reconstructive models emerge as the best fitting choice in scenarios where accuracy takes precedence.

The application of traditional models, including ARIMA (Carlei et al., 2025) and Holt-Winters (Li et al., 2024), has focused on spotting irregularities in energy consumption data. For customer segmentation and flagging outliers, cluster-based methods are commonly employed. Although more advanced solutions like Isolation Forest and One-Class SVM enable model-free anomaly detection, their performance can be limited when dealing with intricate temporal relationships in high-dimensional data.

Deep learning has greatly propelled progress in energy system anomaly detection. Long Short Term Memory (LSTM)-autoencoders capture long-term consumption trends, and Convolutional Neural Networks (CNNs) identify short-term fluctuations in power data. Hybrid frameworks (e.g., CNN-LSTM, Graph Neural Network's (GNNs)) jointly represent spatial and temporal grid characteristics. For smart grid monitoring, Variational Autoencoders (Carvajal Rico et al., 2025) provide a reconstructive detection approach, and Generative Adversarial Networks are emerging as powerful tools for synthesizing realistic load curves, thereby facilitating stronger detection models.

Current studies on sensitive power marketing data fail to adequately address three critical issues: they cannot effectively model multi-scale temporal patterns, achieve limited robustness under class-imbalanced conditions, and do not fully employ attention mechanisms to capture essential temporal-spatial characteristics.

### 2.2. Variational Autoencoder (VAE)

Autoencoders (AEs) are essential neural network architectures (Carvajal Rico et al., 2025) designed to map input data to corresponding outputs with minimal errors. They are composed of three primary layers: the encoder layer, compressing the input data $x \in R^{dx}$ into a latent variable $z \in R^{dz}$, the latent variable layer itself, and the decoder layer, which reconstructs the data back to its original dimensions. The primary objective of an autoencoder is the accurate reconstruction of input data from its latent variables.

However, conventional autoencoders have a notable drawback: their hidden layers may lack continuity, potentially causing substantial computational challenges. To overcome these limitations, the VAE was developed as an enhanced alternative (Ji et al., 2025).

The VAE consists of three main components: an encoder, a decoder, and a loss function. The latent variable z is a random variable with a distribution defined by a Prior function $p(z)$. This Prior is often speculated as a multivariate Standard Gaussian distribution $N(0,1)$. During the training, data x is extracted from the conditional probability distribution $p(x \mid z)$. The decoder, which acts as a generative model, generates the decoded variable distribution based on the encoded latent variables. The encoder is defined by the conditional probability function $p(z \mid x)$, which models how the latent variables are distributed among the encoded variables.

A VAE's loss function consists of two parts: the first part is the reconstruction loss, which is used to maximize the probability of reconstruction. The second part is the regularization term, aimed at ensuring that the probability $q(z \mid x)$ of the latent variable fit the preset prior $p(z)$. The loss function is shown in Eq. (1).

$$L\left(x_i\right) = -E_{z\sim q\left(z|\, x_i\right)}\left[\log p\left(x_i|\, z\right)\right] + KL\left(q\left(z|\, x_i\right) \| \, p\left(z\right)\right) \quad (1)$$

In this context, KL means the Kullback-Leibler (KL) divergence, an metric used to measure the difference between the encoder's distribution $q(z \mid x)$ and the prior distribution $p\left(z\right)$.

Most systems determine the optimal threshold based on receiver operating characteristic ROC curves. The ROC curve depicts the relationship between True Positive Rate (TPR) and False Positive Rate (FPR). By analyzing the ROC curve, it is possible to optimize specificity and sensitivity, thereby locking in the most suitable threshold. However, in the sensitive dataset of power marketing, the problem of category imbalance is prominent and has become a major challenge. Because most unsupervised learning techniques are typically based on the assumption of class equilibrium distribution, which is often broken in practical scenarios.

## 3. Method Overview

### 3.1. Problem Definition

The problem of multivariate time series anomaly detection for sensitive power marketing data can be formally described as follows: when n time series historical data of length $X = \left(x_1, x_2, x_3, \ldots, x_n\right)^T \in R^{n\times T}$ are given, the anomalies in sensitive power marketing data mainly manifest as data points that do not match historical patterns or deviate significantly from most previous data. During the training phase, only normal data is used for model training. During the testing phase, normal and abnormal data are utilized for model detection and assessment.

### 3.2. Method Overview

This study introduces a Multi-Scale Convolutional Variational Autoencoder (MSCVAE) framework for anomaly detection in sensitive power marketing data. This research employs an unsupervised approach for anomaly detection. The proposed framework is structured around four core modules:

1) Multi-Scale Representation: Input multivariate time-series data are converted into hierarchical attribute matrices, encoding system states across varying temporal resolutions.

2) Robust Reconstruction: Convolutional encoder-decoder architecture, enhanced by variational inference, generate reconstructed representations while improving framework robustness.

3) Spatiotemporal Modeling: An attention-enhanced ConvLSTM module (Wei et al., 2025) is incorporated to jointly model inter-feature relationships and temporal dependencies inherent in power marketing time series.

Threshold Optimization: A novel Anomaly Response Rate (ERR)-based thresholding strategy replaces conventional ROC-driven approaches, significantly boosting detection performance in imbalanced scenarios.

To capture the deviations between reconstructed and original matrices, Mean Square Error (MSE) serves as our reconstruction loss. This choice enables effective measurement of reconstruction errors, whose magnitude in our multivariate time series framework directly corresponds to the severity of detected anomalies.

By integrating multi-scale analysis with variational autoencoding, the framework captures both local anomalies and global system behavior. The attention mechanism prioritizes critical temporal patterns, while the ERR thresholding adapts dynamically to data imbalance, addressing key limitations of existing methods.

### 3.3. Data Normalization

The RevIN method (Fan and Chen 2025; Sylligardos et al., 2023; Sarfraz et al., 2024; Zhou et al., 2023) addresses this issue by first normalizing the input data and then performing inverse normalization at the output layer, thus preserving and restoring the original distribution of the input data (Song et al., 2023; Yin et al., 2025).

### 3.4. Multi-Scale Attribute Matrix Generation

To capture the system state embodied in the correlations across power marketing time series, attribute matrices are derived. $M_i \in R^{n\times n}$ by computing pairwise inner products of multivariate time series segments, where the window size is denoted as $w_i$. This process captures the relationships between different time series within a time window, specifically from time $t - w_i$ to $t$. To account for various scales of sensitive power marketing data, we extract multi-scale attribute matrix's using different window sizes $w_i$, thereby capturing data from multiple time steps. The method for generating these attribute matrices is based on the proposed approach, which considers both shape-related along with correlations the scale characteristics characterizing the dual time series. The detailed procedure for generating the multi-scale attribute matrix is outlined in Algorithm 1.

In the context of sensitive power marketing data, anomaly events may vary in duration. To account for this, we set m=3 and select multi-scale time window sizes of 30, 80, and 120 time steps. By performing anomaly detection across these different scales, we significantly enhance the detection accuracy.

### 3.5. Attention-based ConvLSTM

The convolutional encoder generates spatial feature maps that exhibit temporal dependencies on preceding time intervals. Conventional ConvLSTM architecture often experiences performance deterioration with extended sequence lengths, primarily due to challenges in preserving long-range temporal relationships. To mitigate this limitation, Xingjian et al. integrated LSTM mechanisms within CNN frameworks to model temporal patterns in video data. Our framework implements an attention-enhanced ConvLSTM variant to dynamically focus on critical temporal features across varying time intervals.

Algorithm 1: Multi-Scale Attribute Matrix Generation

Input: Multivariate time series $X \in R^{n \times T}$, Time series length T, Window sizes $w_1, w_2, \ldots, w_m$, Time series dimension n, Number of scales m;

Output: Multi-scale attribute matrix $\left[X_1, X_2, \ldots X_m\right]$, where $M^i \in R^{n \times n}$

$for\ i\ in\ m\ do$ :

$for\ t\ in\ T\ do$ :

$$S_{sub} \leftarrow \left[x\left[t-w_i+1\right], x\left[t-w_i+2\right], \ldots, x\left[t\right]\right]$$

$$M_i = \frac{S_{sub} \times S_{sub}^T}{w}$$

$return$ $\left[X_1, X_2, \ldots, X_m\right]$

Fig. 1 depicts the architecture of our modified ConvLSTM unit. At temporal iteration t, the $l$-th encoder layer's output $P^{t,l}$ serves as input to the corresponding ConvLSTM layer. Concurrently, the hidden state $H^{t,l}$ from the prior iteration $t-1$ (at layer $l-1$) is incorporated. The mathematical formulation governing the $l$-th ConvLSTM layer at temporal position $t$ is defined by the following operational sequence, as shown in Eq. (2).

$$\begin{gathered} i^{t,l} = \sigma\left(W_i^l * [H^{t-1,l}, P^{t,l}] + b_i^l\right) \\ f^{t,l} = \sigma\left(W_f^l * [H^{t-1,l}, P^{t,l}] + b_f^l\right) \\ g^{t,l} = tanh(W_c^l * [H^{t-1,l}, P^{t,l}] + b_c^l) \\ C^{t,l} = f^{t,l} \odot C^{t-1,l} + i^{t,l} \odot g^{t,l} \\ o^{t,l} = \sigma(W_o^l * [H^{t-1,l}, P^{t,l}] + b_o^l) \\ H^{t-1,l} = o^{t,l} \odot tanh(C^{t,l}) \end{gathered} \tag{2}$$

Where $\sigma$ denote the sigmoid function, and $W_{pi}^l$, $W_{ci}^l$ and $W_{hi}^l$ represent the convolutional filter kernels for the input gate, the previous time step's cell state $C^{t-1,l}$, and the hidden state from the previous time step, respectively. The input $P^{t,l}$, cell state $C^{t,l}$, hidden state $H^{t,l}$, candidate memory $G^{t,l}$, and the gates $i^{t,l}$, $f^{t,l}$, and $o^{t,l}$ are all represented as three-dimensional tensors. The symbol $*$ represents convolution operation, while $\odot$ represents the Hadamard product.

To further enhance the temporal modeling, this study integrated the attention mechanism as seen in Fig. 1.

### 3.6. MSCVAE Model

Sensitive power marketing data is inherently a multivariate time series with temporal dependencies. To accurately detect anomalies in such data, we propose a novel anomaly detection model, Multi-Scale Convolutional Variational Autoencoder (MSCVAE), which is based on an encoder-decoder architecture, as depicted in Fig. 2.

Encoder: We employ a convolutional encoder to process the spatial representations of the sensitive power marketing data's attribute matrix. Multiple attribute matrices, each representing different scales, are input into a multi-layer convolutional encoder. In our framework, four convolutional encoder layers are used to extract features from these attribute matrices. Let $X^t$ represent the input to the first layer and let $P^{t,l-1} \in R^{n_{l-1} \times n_{l-1} \times d_{l-1}}$ denote the feature map at the $(l-1)$-th layer.

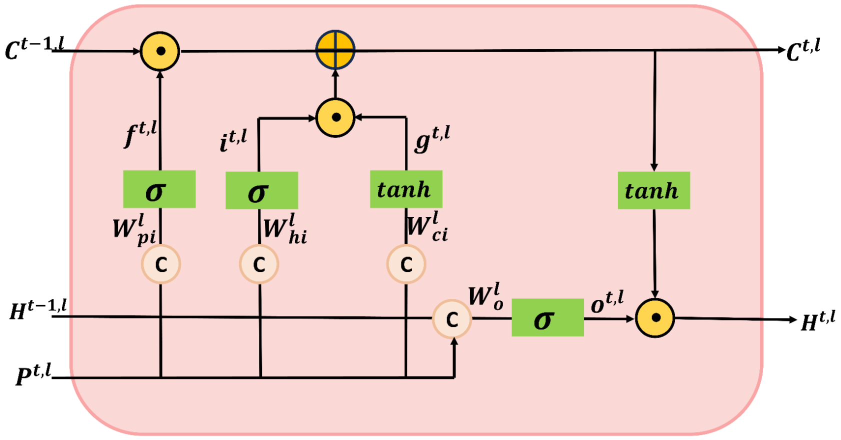



**Fig. 1.** ConvLSTM structure

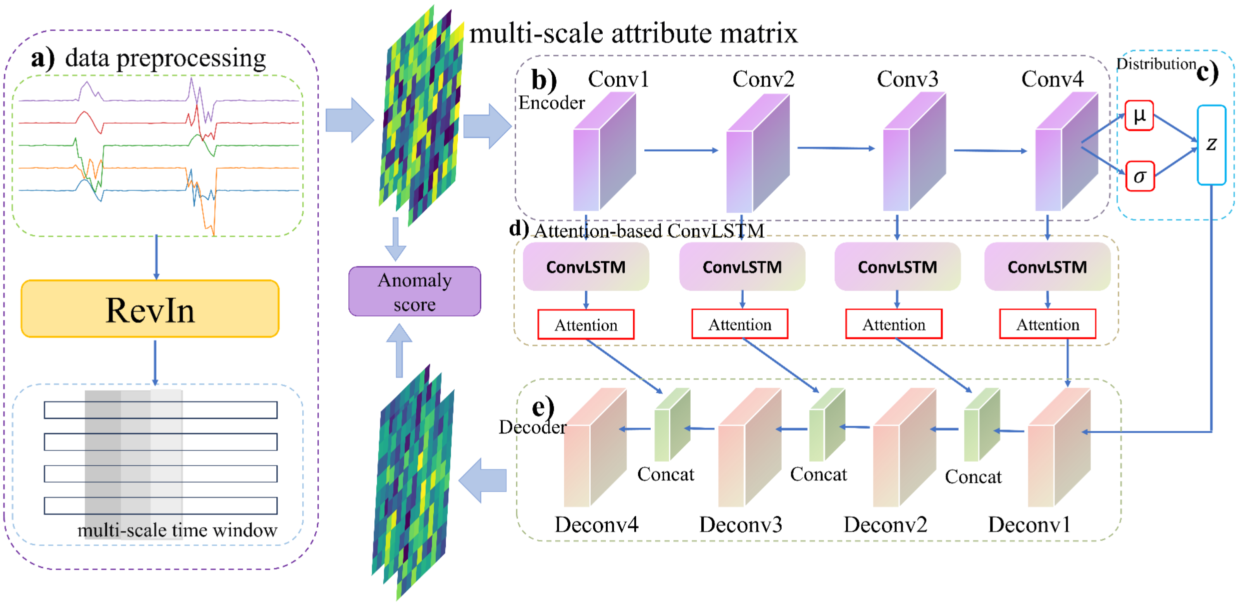



**Fig. 2.** MSCVAE model

The output of the $l$-th layer is computed as follows using Eq. (3).

$$P^{t,l} = f\left(W_e^l * P^{t,l-1} + b_e^l\right) \tag{3}$$

Where $*$ means the convolution operation, $W_e^l$ represents the filter kernel at the $l$-th layer, $b_e^l$ is the bias term, $P^{t,l}$ is the output of the $l$-th layer, $P^{t,l-1}$ is the output from the $(l-1)$-th layer, and $f(\cdot)$ denotes the function of activation. Part (b) of Fig. 2 illustrates the encoding process as shown in Eq. (4).

$$\hat{P}^{t,l-1} = \begin{cases} f\left(\hat{W}_d^{t,l} \hat{H}^{t,l} + \hat{b}_d^{t,l}\right), l = 4 \\ f\left(\hat{W}_d^{t,l}\left[\hat{H}^{t,l} \oplus \hat{P}^{t,l}\right] + \hat{b}_d^{t,l}\right), l = 3,2,1 \end{cases} \tag{4}$$

where $\circledast$ represents the deconvolution operation, $\oplus$ denotes the merging operation, and $f(\cdot)$ is the function of activation. $W_d^{t,l}$ and $b_d^{t,l}$ represent the filter kernel and bias term of the $l$-th layer of the convolutional decoder. The decoding process is performed in reverse, starting from layer $l = 4$ and proceeding layer by layer down to $l = 1$, applying deconvolution at each stage to reconstruct the attribute matrices for each layer. Specifically, deconvolution is applied to the results of the Attention-based ConvLSTM at layer $l = 4$, denoted as $\hat{H}^{t,4}$, to reconstruct the matrix at layer l=3, denoted as $\hat{P}^{t,3}$. The deconvolution operation continues, with the concatenation operation at each layer, to reconstruct the sequence of attribute matrices from layer $l-1$ to layer $1$. Part (e) of Fig. 2 illustrates the decoding process.

### 3.7. Loss Function

The MSE is utilized as the loss function to optimize the parameters of the model. The MSE is widely used in anomaly detection tasks due to its ability to effectively quantify the deviation of predicted outputs from actual measurements. It is computed as follows in Eq. (5):

$$\mathrm{Loss}(\mathrm{t}) = \sum_{t=1}^{T}\sum_{c=1}^{m} P^{t,c} - \hat{P}^{t,c2} \tag{5}$$

where $P^{t,c} \in R^{n\times n}$. This paper uses the mini-batch stochastic gradient descent algorithm in conjunction with the Adam optimizer to mitigate the loss function described above. After training for a sufficient number of epochs, the learned neural network parameters are applied to reconstruct the attribute matrices for the training and validation sets. To perform anomaly detection, the reconstructed multi-scale attribute matrix $\left[M_1, M_2, \ldots M_m\right]$ is subtracted from the original multi-scale attribute matrix $\left[\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_m\right]$, yielding a different matrix. This difference matrix is then used for anomaly detection and diagnosis based on the observed discrepancies.

### 3.8. Threshold Setting Strategy

In the anomaly detection process, the VAE assigns an anomaly score to each set of attribute matrices. The core purpose of threshold setting is to achieve automatic classification of samples by determining the optimal discriminant boundary. The confusion matrix is commonly used to evaluate the performance of the model, showing the relationship between predicted and actual classes of test samples, producing the four fundamental metrics: True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN).

We propose a new threshold setting strategy based on ERR, where ERR is calculated based on TP, FP, FN, and TN to obtain Eq. (6). The objective of this strategy is to minimize the ERR, thereby reducing the number of misclassified samples. Consequently, the optimal threshold is selected by minimizing the ERR, leading to improved anomaly detection performance.

$$ERR = \frac{FP}{FP + TP + TN} \tag{6}$$

## 4. Experiment

### 4.1. Dataset Description

To assess the effectiveness of the proposed approach, we applied it comprehensively to two real-world power marketing datasets obtained from an electric utility company. From this data, we created two distinct datasets: Electric Sensitive Power Marketing Dataset 1 (ESPMD1) and Electric Sensitive Power Marketing Dataset 2 (ESPMD2).

### 4.2. Experimental Setup

The experiments were carried out in Python 3.8 with the PyTorch (PyTorch is an open-source deep learning framework) 2.4.1 framework. To maintain fairness in comparisons, all baseline models were tested under identical conditions as the proposed method. The tests were performed on an NVIDIA Tesla V100-PCIE GPU. The experimental hyperparameters are configured as follows:

**Table 1.** Experimental settings and parameter values

| Parameter | Value |
|---|---|
| Learning rate | 0.001 |
| Batch size | 64 |
| Number of scales | 3 |
| Window sizes | {30, 80, 120} |
| Latent dimension | 32 |
| ConvLSTM layers | 2 |
| Attention heads | 4 |

A three-step data preprocessing procedure is implemented. First, the time series is resampled to achieve a uniform 15-minute interval for temporal alignment. Following this, anomaly labeling is performed: the training set is constructed exclusively from normal samples, whereas anomalies in the test set are flagged according to expert-defined rules. Finally, the RevIN technique (Fan and Chen 2025) is employed to normalize the input data, standardizing the sequences before processing and restoring their original distribution post-reconstruction.

The optimal configuration for the multi-scale module was determined to be the window combination (30, 80, 120), following an evaluation of candidate sizes {20, 30, 50, 80, 120} on a validation set. This selection was based on its superior performance in both F1 and G-mean metrics.

Furthermore, five anomaly detection techniques were selected for benchmarking to verify the framework's efficacy. The compared methods consist of the traditional Local Outlier Factor (LOF) [15], K-Nearest Neighbors (KNN) (Yang et al., 2025), the reconstruction-based VAE, the state-of-the-art MIXAD (Ren et al., 2026) model, and a Generative Adversarial Network (GAN) (Vishwakarma and Kumar, 2025) model.

### 4.3. Evaluation Metrics

During testing, each anomaly detection model was trained exclusively on a normal-sample training subset (Li and Jung 2023). After training, this method's model was assessed using a validation set containing both normal and anomalous samples (Abdulhadi et al., 2023). Model performance was measured with the application of four primary metrics (accuracy, recall, F1-Score, and G-mean), where scores closer to 1 indicate superior detection capability (Yan et al., 2025). The resulting confusion matrices and evaluation metrics for both datasets are detailed in Table. 2 and Table. 3, with top-performing values emphasized in bold.

**Table 2.** Anomaly detection results for ESPMD1

| Model | Confusion Matrix | | | | Evaluation Metrics | | | |
|---|---|---|---|---|---|---|---|---|
| | TP | FP | FN | TN | Precision | Recall | F1-score | Gmean |
| KNN | 82 | 23 | 13 | 52 | 0.7810 | 0.8632 | 0.8200 | 0.7736 |
| LOF | 54 | 51 | 32 | 33 | 0.5143 | 0.6270 | 0.5654 | 0.4967 |
| VAE | 89 | 11 | 4 | 66 | 0.8900 | 0.9570 | 0.9223 | 0.9057 |
| MIXAD | 77 | 28 | 19 | 46 | 0.7333 | 0.8021 | 0.7662 | 0.7061 |
| GAN | 97 | 1 | 6 | 66 | 0.6967 | 0.9898 | 0.9652 | 0.9166 |
| Ours | 99 | 1 | 4 | 66 | 0.9705 | 0.8970 | 0.9323 | 0.9201 |

**Table 3.** Anomaly detection results for ESPMD2

| Model | Confusion Matrix | | | | Evaluation Metrics | | | |
|---|---|---|---|---|---|---|---|---|
| | TP | FP | FN | TN | Precision | Recall | F1-score | Gmean |
| KNN | 164 | 92 | 94 | 140 | 0.6406 | 0.6357 | 0.6381 | 0.6193 |
| LOF | 152 | 104 | 123 | 111 | 0.5938 | 0.5527 | 0.5725 | 0.5342 |
| VAE | 259 | 16 | 23 | 192 | 0.9418 | 0.9184 | 0.9300 | 0.9208 |
| MIXAD | 178 | 78 | 77 | 157 | 0.6953 | 0.6980 | 0.6967 | 0.6829 |
| GAN | 157 | 99 | 103 | 131 | 0.6133 | 0.6038 | 0.6085 | 0.5865 |
| Ours | 296 | 9 | 34 | 151 | 0.99 | 0.9612 | 0.9754 | 0.973 |

The comparative analysis reveals that traditional anomaly detection algorithms underperform in multivariate time series modeling tasks. In contrast, the MIXAD model, which employs GNNs to autonomously learn inter-feature correlations, outperforms the VAE in all evaluated metrics. The observed performance gap, where GAN excels on ESPMD2 but not on ESPMD1, is likely due to the former's stable regional consumption patterns, a characteristic that favors GAN's synthesis approach. In contrast, the significant fluctuations in ESPMD1 prove less effective for GAN's reconstruction-based methodology. Nevertheless, MIXAD's anomaly detection accuracy remains marginally lower than our proposed model, primarily due to its incomplete modeling of temporal dynamics. Interestingly, while the GAN yields unsatisfactory results on ESPMD1, it achieves competitive performance on ESPMD2. This discrepancy likely stems from regional variations in electricity consumption patterns between the two datasets.

### 4.4. Ablation Experiments

To assess the contributions of each component within the proposed model, ablation studies were carried out to study two critical modules: the Multiscale Attribute Matrix and the Attention-based ConvLSTM. Three MSCVAE-CAD variants were examined: (1) MSCVAEw-CAD, in which the Attention-based ConvLSTM is removed. (2) CVAEa-CAD, with the Multiscale Attribute Matrix removed. (3) CVAEw-CAD, with both modules removed. Fig. 3 presents the F1 scores and G-means across the two cable datasets. The results demonstrate that the complete MSCVAE-CAD framework significantly outperforms all three ablated variants, underscoring the effectiveness of both the Multiscale Attribute Matrix and the Attention-based ConvLSTM in enhancing anomaly detection performance. Notably, MSCVAEw-CAD and CVAEw-CAD exhibit comparable performance, while CVAEa-CAD achieves markedly better results than both, further affirming the critical role of the Attention-based ConvLSTM.

Overall, these findings highlight that both modules are essential for achieving robust anomaly detection in time series of multivariate, especially in complex cable monitoring scenarios.

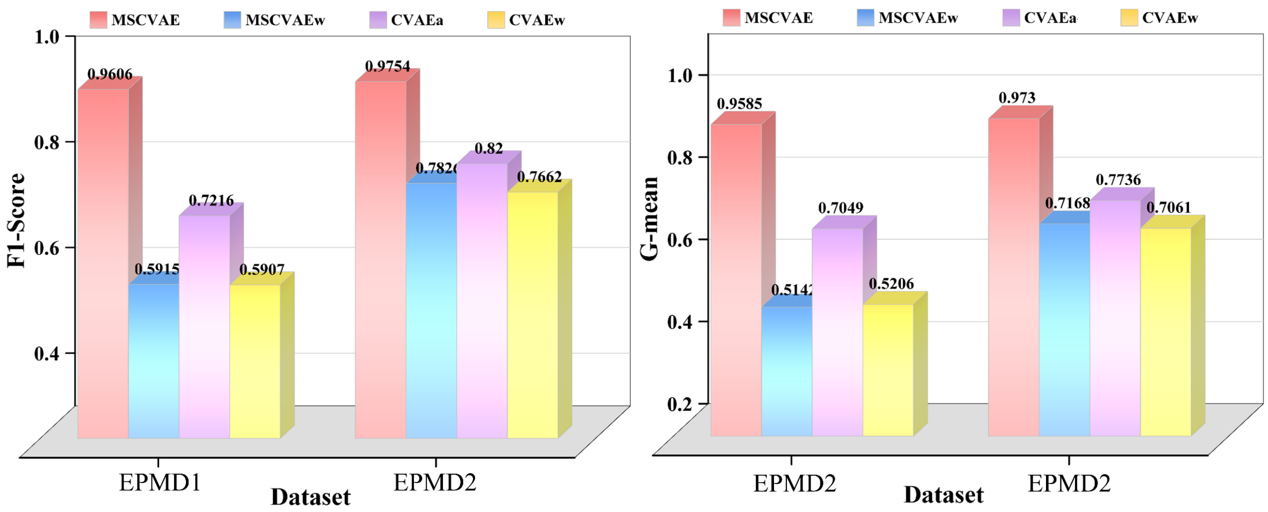
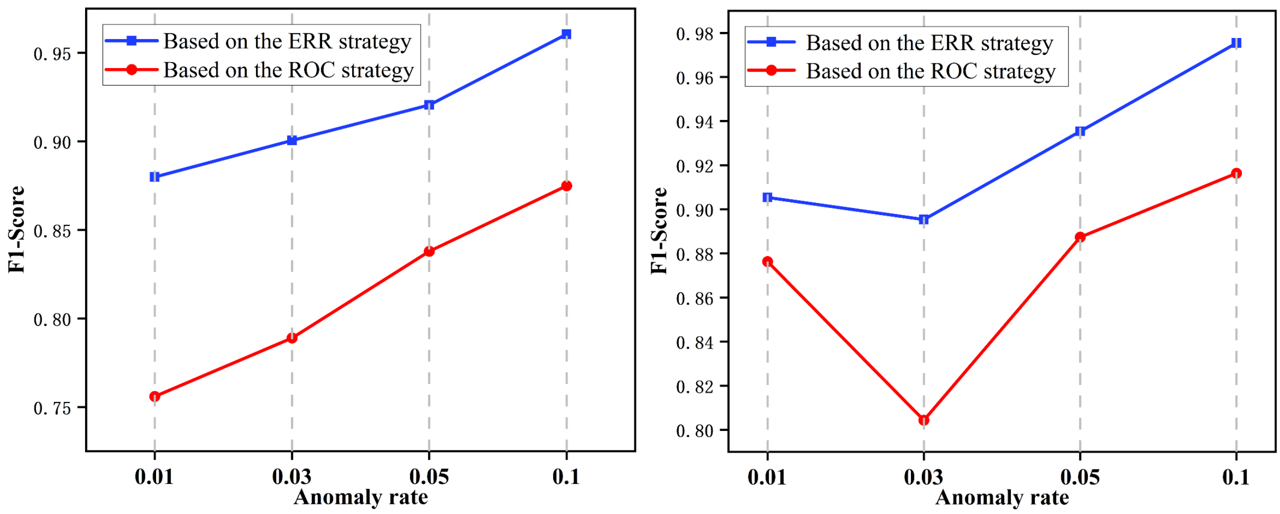



**Fig. 3.** Ablation experiments

### 4.5. Comparison of Threshold Setting Strategies

To strategically conduct a systematic analysis of the newly developed threshold configuration method, we compare two approaches, an ERR-based strategy and a ROC-based strategy, using two power marketing datasets. Both strategies are applied within the same anomaly detection framework across varying anomaly rates, and the results are presented in Fig. 4. The ERR strategy consistently outperforms the ROC-based strategy in terms of detection performance, indicating its superior ability to determine more appropriate threshold values. Moreover, an elevated anomaly rate correlates with a reduced performance gap between the two evaluated strategies. Overall, the proposed ERR-based strategy demonstrates greater effectiveness than the traditional ROC-based approach, particularly in handling imbalanced datasets.




a) EPMD1

b) EPMD2

**Fig. 4.** Abnormal strategy experiment

### 5. Conclusion

This study introduces a MSCVAE framework for anomaly detection in electric sensitive power marketing data, addressing key challenges in multivariate time-series analysis. The proposed method first converts each variable’s time series into multi-scale attribute matrix, then employs a CVAE for robust data reconstruction. To enhance temporal modeling, we integrate an Attention-based ConvLSTM module, which simultaneously captures inter-variable correlations and long-term temporal dependencies. Additionally, we propose a novel ERR-based thresholding strategy, which surpasses conventional ROC-based methods, particularly in handling class-imbalanced datasets. More experiments used for two real-world electric power marketing datasets confirm that this paper model achieves superior detection accuracy compared to state-of-the-art baselines.

The current MSCVAE architecture lacks explicit mechanisms for handling extreme class imbalance, such as specialized sampling techniques or cost-sensitive loss functions. Consequently, future research will investigate the integration of synthetic oversampling methods and weighted loss functions to enhance its applicability in highly imbalanced scenarios.

**Author Contributions**

Jinkai Sun contributed to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation and manuscript editing. Chun Xiao visualization, supervision, project administration, and funding acquisition. Junfeng Yao contributed to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition.

**Funding**

This work was supported by the Science and Technology Project of State Grid Corporation of China (No. 5700-202341299A-1-1-ZN).

**Institutional Review Board Statement**

Not applicable.

**Declaration of Artificial Intelligence (AI) Tools**

The authors used AI tools solely for language editing and readability improvement. The authors reviewed and verified all content and take full responsibility for the accuracy and integrity of the manuscript.

**References**

Abdulhadi, O. M., Ghaleb, F. F. M., Hassanien, A. E., Abdelfattah, A. M. H., and Zakaria, W. (2023). Deep clustering-based anomaly detection and health monitoring for satellite telemetry. *Big Data and Cognitive Computing*, 7(1), 39. doi: 10.3390/bdcc7010039.

Carlei, V., Furia, D., Cascioli, P., and Manenti, P. (2025). Macro-founded machine learning models for power market price trend detection. *Finance Research Letters,* 83, 107660. doi: 10.1016/j.frl.2025.107660.

Carvajal Rico, J., Alaeddini, A., Faruqui, S. H. A., Fisher Hoch, S. P., and Mccormick, J. B. (2025). A generative framework for predictive modeling of multiple chronic conditions using graph variational autoencoder and bandit-optimized graph neural network. *IEEE Journal of Biomedical and Health Informatics*.

Chen, B., and Cui, W. (2025). Data-driven power marketing strategy optimization and customer loyalty promotion. *Energy Informatics,* 8(1), 53. doi: 10.1186/s42162-025-00510-y.

Fan, C., and Chen, H. (2025). Research on eXplainable artificial intelligence in the CNN-LSTM hybrid model for energy forecasting. *Journal of Building Engineering,* 111, 113150. doi: 10.1016/j.jobe.2025.113150.

Ji, T., Xu, P., Guo, D., Sun, L., Ma, K., Wang, Y., and Han, X. (2025). Variational autoencoder based anomaly detection in large-scale energy storage power stations. *Energies,* 18(11), 2770. doi: 10.3390/en18112770.

Jin, J., Qin, H., Sun, W., Qian, Y., Wang, H., Wang, Z., Zhou, J., and Huang, X. (2025). Foreign object intrusion detection and early warning in substation video surveillance based on deep learning. *International Journal of Power and Energy Systems,* 45(1), 46-57. doi: 10.2316/J.2025.203-0588.

Li, G., and Jung, J. J. (2023). Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges. *Information Fusion*, 91, 93-102. doi: 10.1016/j.inffus.2022.10.008.

Li, X., Zhuang, M., Yang, W., Zhu, X., and Wang, Q. (2024). Neural network based data quality monitoring and real-time analysis method for energy storage power plants. *International Journal of Power and Energy Systems,* 44(10), 1-15. doi: 10.2316/J.2024.203-0534.

Ren, C., Li, M., Chen, C., Guan, X., and Huang, G., Q. (2026). Multi-modal digital twins for industrial anomaly detection: Framework, method, and application. *Robotics and Computer-Integrated Manufacturing,* 97, 103068. doi: 10.1016/j.rcim.2025.103068.

Sarfraz, M. S., Chen, M.-Y., Layer, L., Peng, K., and Koulakis, M. (2024). Position paper: Quo vadis, unsupervised time series anomaly detection? *In Proceedings of the International Conference on Machine Learning.*

Song, J., Kim, K., Oh, J., and Cho, S. (2023). Memto: Memory-guided transformer for multivariate time series anomaly detection. *Advances in Neural Information Processing Systems,* 36, 57947-57963.

Sylligardos, E., Boniol, P., Paparrizos, J., Trahanias, P. E., and Palpanas, T. (2023). Choose wisely: An extensive evaluation of model selection for anomaly detection in time series. *Proceedings of the VLDB Endowment,* 16(11), 3418-3432. doi: 10.14778/3611479.3611536.

Vishwakarma, D., and Kumar, S. (2025). Automatic query expansion for enhancing document retrieval system in healthcare application using GAN based embedding and hyper-tuned DAEBERT algorithm. *Data & Knowledge Engineering,* 160, 102468. doi: 10.1016/j.datak.2025.102468.

Wei, L., Peng, X., and Cao, Y. (2025). Enhanced fault diagnosis of rolling bearings using an improved inception-lstm network. *Nondestructive Testing and Evaluation,* 40(7), 3274-3293. doi: 10.1080/10589759.2024.2402549.

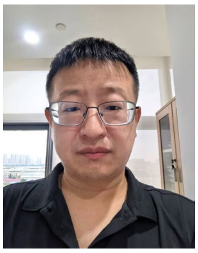
Yan, T., Tu, J., Li, C., and Zhang, F. (2025). Topology-preserved information bottleneck for multiview anomaly detection. *IEEE Transactions on Neural Networks and Learning Systems.* doi: 10.1109/TNNLS.2025.3579412.

Yang, H., Xia, H., Liu, S., Chen, S., Li, L., Liao, X., and Hu, X. (2025). A study on time-series prediction and analysis of acidity of Daqu based on multivariate data fusion and KNN-Attention-LSTM-XGBoost modeling. *Bioprocess and Biosystems Engineering*. doi: 10.1007/s00449-025-03187-5.

Yin, P., Wang, F., Liu, Y., Song, P., Zhao, J., Yue, Z., and Liu, J. (2025). Reconstruct defect features in ultrasonic scan data by multi-time local outlier factors. *Measurement,* 255, 117998. doi: 10.1016/j.measurement.2025.117998.

Zhang, M., Zhi, H., Zhao, J., Li, R., Chang, X., Fan, R., and Xu, K. (2023). Research on power quality assessment based on ubiquitous power IoT. *Applied Mathematics and Nonlinear Sciences,* 8(1), 2873-2892. doi: 10.2478/amns.2023.1.00007.

Zhou, T., Niu, P., Sun, L., and Jin, R. (2023). One fits all: Power general time series analysis by pretrained lm. *Advances in Neural Information Processing Systems,* 36, 43322-43355.



Jinkai Sun is a specialist at the State Grid Shanxi Marketing Service Center, with a focused expertise in marketing information security and artificial intelligence. His work involves addressing and integrating AI-driven solutions to improve service efficiency.

Xiao Chun, (1987.11.5) female, Han Nationality, was born in Yongji, Shanxi Province. She currently works at the State Grid ShanXi Marketing Service Center, where she serves as a senior engineer. She earned her master’s degree in electrical engineering, with her primary research focus on power system operation and control, as well as electrical power marketing technology.

Yao Junfeng (1980.03.26), male, of Han nationality, hails from Xinzhou, Shanxi Province. Mr. Junfeng is a senior engineer at the State Grid ShanXi Marketing Service Center. He completed his undergraduate studies in electrical engineering, and his research focuses on power system operation and control, as well as power marketing technology.