Participant-Level Anomaly Detection for Generation and Load Data Using Dual-Side LSTMs

Qianya He[1], Qi Liu[1], Xueliang Gong[1], Jiaxun Liu[1], Liming Yao[1], Tianze Yu[2], and Tong Zhao[2,*]

[1]Guangdong Power Exchange Center Co., Ltd., China
[2]Beijing Tsintergy Technology Co., Ltd., China
[*]Correspondence: Tong Zhao; Email: 471560142@qq.com

ABSTRACT

Reliable anomaly detection for generation and load metering data is essential to market settlement. In practice, anomalies are diverse and sparse, and the load side contains a large and time-varying number of participants, which makes fine-grained participant-level localization difficult. Conventional statistical thresholding and generic outlier detectors (e.g., LOF) are often sensitive to nonstationarity and cannot effectively exploit temporal dependency across intra-day time slots, resulting in coarse alarms and high false positives. To address these issues, this paper proposes a dual-side LSTM-based participant-level anomaly detection method. Multimodal features are constructed from 24 intra-day measurements, a daily total, FFT-derived frequency components, and calendar context. A zero-padding and masking mechanism is introduced to handle daily changes in the number of load participants without contaminating model training. A dual-layer LSTM with a 16-d participant embedding learns participant-specific temporal patterns, and a rule combining relative-error screening with confidence verification produces participant-level anomaly positioning and abnormal time-slot identification. Experiments on one-month desensitized provincial-grid data (242 generators and 78k–85k loads) achieve 100.0% F1-score (100.0% recall, 100.0% precision), with highly accurate positioning and time-slot recognition, substantially outperforming statistical and LOF baselines.

Introduction

With the deepening of high-renewable integration and electricity-market reforms, the correctness of generation and consumption data has become a prerequisite for reliable settlement, transparent market operation, and secure grid dispatch. In practice, the daily metering and reporting streams on both the generation side and the consumption side are increasingly leveraged for market clearing, deviation assessment, and risk monitoring. However, real-world data are often contaminated by diverse anomalies, including reporting mistakes, communication and sensor failures, device malfunctions, and even malicious tampering, an issue that has been widely recognized in power-system cyber-physical security studies[1,2]. Such anomalies are not merely data-quality issues; they may propagate into erroneous settlement, distorted market signals, and delayed operational responses. These concerns are further amplified by the intrinsic temporal nature of power data and the large-scale, heterogeneous, and time-varying population of subjects on the consumption side.

Unlike many classical power-system data validation settings that rely on relatively stable measurement structures and redundancy[3], modern generation–consumption datasets exhibit strong nonstationarity, complex periodicities, and changing subject participation, which jointly challenge accurate and scalable anomaly detection; these issues are also emphasized in recent reviews of time-series anomaly detection[4,5].

A key operational obstacle is that many existing solutions still operate at coarse granularity. In daily practice, anomaly screening is frequently conducted at the date level or the area level, flagging that an aggregated profile appears abnormal while leaving the responsible subjects and specific abnormal intervals unidentified. This coarse localization forces operators to perform expensive manual drill-down, which is particularly inefficient when the number of consumption-side subjects is on the order of tens of thousands and varies over time. Meanwhile, fine-grained subject-level detection cannot be obtained simply by applying a global detector independently to each subject: subject behaviors are heterogeneous, anomalies are sparse and imbalanced, and the temporal dependency across 24 intra-day intervals is essential for distinguishing true irregularities from normal variation. Therefore, a practical and high-value capability is to bridge date-level screening and actionable subject-level localization, enabling the detection system to pinpoint which subject is abnormal and at which interval(s), while remaining robust to dynamic subject populations.

Existing research on anomaly detection has been systematically reviewed in both general data-mining and application-oriented contexts[6,7]. Statistical methods typically rely on handcrafted indicators such as mean, variance, and threshold rules. They are easy to deploy but are sensitive to noise and shifting patterns, and they often fail to capture long-range temporal

dependency and complex anomaly forms. Traditional machine-learning approaches introduce feature engineering and employ generic outlier detectors, such as density-based local outlier factor[8] and isolation-based methods[9]. These methods are attractive for unsupervised settings, yet their performance heavily depends on feature design and their ability to represent temporal structure.

In the energy domain, recent works have investigated deep learning for time-series anomalies, leveraging recurrent architectures such as LSTM[10,11] and autoencoder-style reconstruction to avoid extensive labeling. For example, BiLSTM-autoencoder pipelines have been explored for smart metering anomalies[12], and variational recurrent autoencoders with attention have been proposed to handle noisy, weakly-labeled smart meter streams[13]. Moreover, deep learning has been applied to energy consumption anomaly detection using high-dimensional inputs[14] and multivariate correlation modeling via graph attention mechanisms[15]. Related efforts also consider power-quality anomaly detection using advanced sensing and pattern recognition[16], and privacy-preserving training via federated learning has been introduced to cope with data silos in smart grid environments[17].

Despite these advances, a common gap remains: many studies focus on detecting whether a time series or an aggregated record is abnormal, but do not explicitly address the operational requirement of subject-level localization under a dynamic subject population, especially when generation and consumption sides exhibit fundamentally different subject cardinalities and dynamics[18–20].

This paper targets the above gap by proposing a dual-sided, subject-level anomaly detection framework tailored to generation and consumption datasets with strong temporal structure and dynamic subject participation. The goal is to upgrade anomaly identification from coarse date-level flags to actionable subject-level decisions, while explicitly distinguishing the roles of intra-day interval data and daily total data. The proposed method integrates (i) multimodal feature construction that jointly models 24-interval time-domain sequences, frequency-domain representations, and calendar-related temporal descriptors; (ii) a dual-layer LSTM backbone that captures long-term temporal dependency and complex anomaly patterns in a data-driven manner; and (iii) a dynamic subject adaptation mechanism that standardizes variable subject dimensions via padding and masking so that model training is not corrupted by invalid placeholders. Finally, to translate model outputs into fine-grained operational actions, we design a subject-level anomaly determination rule that combines relative-error-based interval screening with confidence-based verification, enabling simultaneous output of subject-level anomaly details and date-level summaries for different business needs.

The remainder of this paper is organized as follows. Section 2 presents the participant-level anomaly detection framework, including the overall workflow, data preprocessing, anomaly annotation, and the subject-level anomaly determination rule with the corresponding output format. Section 3 describes the dual-layer LSTM model with multimodal fusion, together with the loss function, optimization strategy, and ablation-study design. Section 4 reports the experimental setup and case-study results, including comparative performance, participant-level detection results, and robustness analysis. Finally, Section 5 concludes the paper and outlines future research directions.

Participant-Level Anomaly Detection Framework

Overall Framework Design

A four-stage closed-loop framework of "data preprocessing - anomaly annotation - temporal modeling - participant-level determination" is designed to realize the fully automated processing from raw data to abnormal participant positioning, as shown in Fig. 1.

The core functions of each stage are as follows: The data preprocessing layer completes data standardization, multimodal feature construction, and distinction between time-slot and total data. The anomaly annotation layer generates abnormal samples under real scenarios. The temporal modeling layer captures long-term temporal dependencies and participant characteristics. The participant-level determination layer outputs fine-grained anomaly results.

Data Preprocessing and Anomaly Annotation

Data Source and Feature Analysis

The experimental data is derived from desensitized power generation and consumption data of a provincial power grid, including a fixed number of 242 power generation participants and a dynamic number of 78,000-85,000 electricity consumption participants. The data includes 24 time-slot monitoring values and 1 daily total value per day (25 columns in total).

Ethical Statement All methods were carried out in accordance with relevant guidelines and regulations. This study does not involve human participants, human data, or human tissue samples. No ethical approval or informed consent is required for this research.

Multimodal Feature Construction

Twenty-four time-slot temporal data and one daily total value are extracted. Frequency-domain features are obtained via FFT transformation to capture periodic patterns[21]. Together with time-context features such as the day of the week and holidays, a

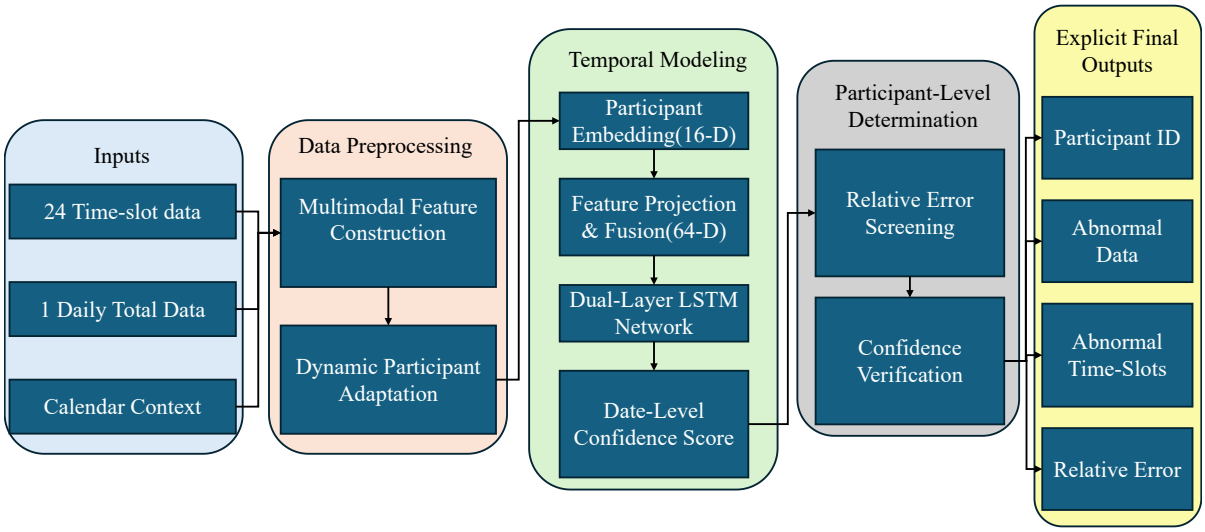

Figure 1. Overall framework of participant-level anomaly detection

multi-dimensional multimodal input is constructed. The composition of the multimodal features is summarized in Table 1.

Table 1. Composition of Multimodal Features

Feature Type	Specific Content
Core temporal features	24 time-slot power generation/consumption values + 1 daily total value
Frequency-domain features	Frequency characteristics extracted by FFT from the 24 time-slot sequence
Time-context features	Day of week (1–7), weekend (0/1), holiday (0/1), month, day of month (5 dimensions)

Dynamic Participant Adaptation

Zero-padding is employed to standardize the data dimensions of electricity consumers. A mask matrix is introduced to filter padded data and prevent invalid information from interfering with model training. The mask matrix M is defined as[22]

$$M_{i,j} = \begin{cases} 0, & j \leq N_{\text{daily}} \\ 1, & j > N_{\text{daily}} \end{cases} \tag{1}$$

where N_{daily} represents the number of real electricity consumers on a given day.

Real-Scenario Anomaly Annotation

Three types of abnormal samples (power generation anomalies, electricity consumption anomalies, dual anomalies) are generated through the "random time-slot disturbance + extreme value injection" strategy, simulating real scenarios such as equipment failures and data reporting errors.To simulate highly covert and realistic grid faults, the disturbance amplitude is carefully constrained to 10%–30% (representing subtle deviations), completely eliminating straightforward zero-value injections. This modification makes the anomaly detection task significantly more challenging and representative of true operational micro-faults, rigorously preventing the task from being trivially separable.

Temporal Modeling with a Dual-Layer LSTM

Model Structure Design

A deep learning architecture of "participant embedding - feature fusion - dual-layer LSTM - classification output" is constructed to adapt to multimodal features and dynamic participant characteristics.

Core Module Functions

Participant embedding layer: Maps discrete participant IDs to 16-dimensional dense vectors (balancing expressive ability and computational complexity) to learn the inherent attributes of participants. Multimodal fusion layer: Integrates temporal, frequency-domain, and time features, realizing feature normalization and dimension unification. Dual-layer LSTM layer: Captures long-term temporal dependencies, with 64 units in each layer. Output layer: Outputs the date-level anomaly confidence (threshold = 0.3) through the Sigmoid activation function[23].

Participant-Level Anomaly Determination Rule

Participant-Level Anomaly Determination Logic

The abnormal state of an individual participant is determined based on a *relative error threshold*. A participant is identified as abnormal if the relative error of either the time-slot data or the daily total data exceeds 30%, and at least one abnormal time slot is detected. The relative error is defined as

$$\text{rel\_error} = \frac{|\text{true\_value} - \text{predicted\_value}|}{\text{true\_value}}. \tag{2}$$

Date-Level Anomaly Verification

To improve the reliability of anomaly detection results, the participant-level judgment is further combined with the model output confidence (threshold set to 0.3) and the total number of abnormal participants on the same day. This strategy avoids misjudgment caused by relying on a single rule and ensures robustness of the final decision.

Result Output Form

Two types of results are generated simultaneously:

- **Participant-level details**, including the date, participant ID, abnormal time slots, and corresponding error values;
- **Date-level summary**, including the total number of abnormal participants and the basis for anomaly determination.

This dual-level output design satisfies diverse business and analysis requirements.

Dual-Layer LSTM with Multimodal Fusion

Architecture Details

Input Layer Design

The multimodal input is constructed by concatenating three feature subsets, with the unified input tensor defined as:

$$\mathbf{X} = \left[\mathbf{X}_{\text{temp}} \oplus \mathbf{X}_{\text{freq}} \oplus \mathbf{X}_{\text{time}}\right] \in \mathbb{R}^{T\times N\times 62} \tag{3}$$

where: - $T = 31$ (number of time steps, i.e., days in December 2024), - N denotes participant scale ($N_{\text{gen}} = 242$ for power generation side, $N_{\text{load}} = 85000$ for electricity consumption side), - $D_{\text{total}} = 62$ (total feature dimension: 25-dimensional temporal features + 32-dimensional frequency-domain features derived from the 24 time-slots zero-padded to 32 points before FFT + 5-dimensional time features), - $\oplus$ represents the feature concatenation operation.

The temporal feature tensor $\mathbf{X}_{\text{temp}} \in \mathbb{R}^{T\times N\times 25}$ explicitly distinguishes time-slot and total data:

$$\mathbf{X}_{\text{temp}} = [\mathbf{X}_{\text{slot}} | \mathbf{X}_{\text{total}}] \tag{4}$$

where $\mathbf{X}_{\text{slot}} \in \mathbb{R}^{T\times N\times 24}$ is 24-time-slot monitoring data, and $\mathbf{X}_{\text{total}} \in \mathbb{R}^{T\times N\times 1}$ is daily total data.

Participant Embedding Layer

Discrete participant IDs are mapped to continuous 16-dimensional dense vectors through an embedding matrix lookup to capture inherent participant characteristics:

$$\mathbf{E} = \mathbf{W}_e(\mathbf{ID}) \in \mathbb{R}^{N\times D_e} \tag{5}$$

where $\mathbf{ID} = [1, 2, \ldots, N]$ is the participant ID sequence, $\mathbf{W}_e \in \mathbb{R}^{N\times D_e}$ is the learnable embedding weight matrix, and $D_e = 16$ (embedding dimension). The embedded features are fused with the multimodal input via element-wise multiplication:

$$\mathbf{X}_{\text{fusion}} = \mathbf{X} \otimes \mathbf{E}^{\text{T}} \in \mathbb{R}^{T\times N\times 62} \tag{6}$$

Feature Projection Layer

A linear projection layer acts as a bottleneck to unify all feature dimensions to 64 (optimal dimension verified via ablation study):

$$\mathbf{X}_{\text{proj}} = \mathbf{X}_{\text{fusion}} \cdot \mathbf{W}_p + \mathbf{b}_p \in \mathbb{R}^{T \times N \times D_p} \tag{7}$$

where $\mathbf{W}_p \in \mathbb{R}^{62 \times D_p}$ (projection weight matrix) and $\mathbf{b}_p \in \mathbb{R}^{D_p}$ (bias vector) with $D_p = 64$. Layer normalization is applied to stabilize training, which is particularly suitable for recurrent architectures where batch statistics may vary significantly across time steps[24].

$$\mathbf{X}_{\text{norm}} = \gamma \odot \frac{\mathbf{X}_{\text{proj}} - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \tag{8}$$

where μ and σ^2 are the batch mean and variance, γ and β are learnable scale and shift parameters, and $\varepsilon = 10^{-5}$ is the numerical stability term.

Dual-Layer LSTM for Temporal Feature Extraction

The first LSTM layer captures short-term temporal dependencies:

$$\begin{aligned}
\mathbf{i}_1 &= \sigma(\mathbf{W}_{ii1}\mathbf{X}_{\text{norm}} + \mathbf{W}_{hi1}\mathbf{h}_0 + \mathbf{b}_{i1}) \\
\mathbf{f}_1 &= \sigma(\mathbf{W}_{if1}\mathbf{X}_{\text{norm}} + \mathbf{W}_{hf1}\mathbf{h}_0 + \mathbf{b}_{f1}) \\
\mathbf{g}_1 &= \tanh(\mathbf{W}_{ig1}\mathbf{X}_{\text{norm}} + \mathbf{W}_{hg1}\mathbf{h}_0 + \mathbf{b}_{g1}) \\
\mathbf{o}_1 &= \sigma(\mathbf{W}_{io1}\mathbf{X}_{\text{norm}} + \mathbf{W}_{ho1}\mathbf{h}_0 + \mathbf{b}_{o1}) \\
\mathbf{c}_1 &= \mathbf{f}_1 \odot \mathbf{c}_0 + \mathbf{i}_1 \odot \mathbf{g}_1 \\
\mathbf{h}_1 &= \mathbf{o}_1 \odot \tanh(\mathbf{c}_1)
\end{aligned} \tag{9}$$

where: - $\mathbf{i}, \mathbf{f}, \mathbf{o}, \mathbf{g}$ denote input, forget, output, and candidate gates, - $\mathbf{W}_*$ are weight matrices, $\mathbf{b}_*$ are bias vectors, - $\sigma(\cdot)$ is sigmoid activation, $\odot$ is Hadamard product, - $\mathbf{h}_0 = \mathbf{0}, \mathbf{c}_0 = \mathbf{0}$ (initial hidden/cell states).

The second LSTM layer captures long-term temporal dependencies with dropout regularization ($p = 0.1$)[25]:

$$\mathbf{h}_2 = \text{Dropout}(\text{LSTM}(\mathbf{h}_1; \theta_{\text{lstm2}}); p = 0.1) \in \mathbb{R}^{T \times N \times D_{\text{lstm}}} \tag{10}$$

where $D_{\text{lstm}} = 64$ (LSTM hidden dimension) and θ_{lstm2} represents trainable parameters of the second LSTM layer.

Output Layer and Decoupled Localization

It is important to clarify that the dual-layer LSTM acts as a macroscopic date-level screener. A dense layer with sigmoid activation outputs the date-level anomaly confidence score:

$$\hat{y}_t = \sigma(\mathbf{W}_o \cdot \text{MeanPooling}(\mathbf{h}_{2,t}) + \mathbf{b}_o) \in (0, 1) \tag{11}$$

where MeanPooling$(\cdot)$ aggregates the N participant-level features into a global daily representation. While this aggregation blends individual participant identities, it effectively captures the overall macroscopic grid anomaly status $\hat{y}_t$.

Once a date is globally flagged as anomalous ($\hat{y}_t > 0.3$), the fine-grained **participant-level localization** is subsequently triggered as an independent post-processing module using the deterministic relative-error screening rule defined in Eq. (2). This decoupled design ensures high computational efficiency for continuous global screening while preserving the exact positioning of local faults when anomalies occur.

Loss Function and Optimization Strategy

Weighted Cross-Entropy Loss

To address class imbalance (30% abnormal samples), weighted binary cross-entropy loss is adopted[26]:

$$\mathscr{L} = -\frac{1}{T}\sum_{t=1}^{T} \omega_t \left[y_t \log(\hat{y}_t) + (1 - y_t)\log(1 - \hat{y}_t) \right] \tag{12}$$

where: - $y_t \in \{0, 1\}$ is ground-truth label (1=abnormal, 0=normal), - $\omega_t = \begin{cases} 3.0 & (y_t = 1) \\ 1.0 & (y_t = 0) \end{cases}$ is sample weight, - $T = 31$ is total number of date samples.

Optimization Objective

The model is optimized via minimizing the loss function with L2 regularization to prevent overfitting:

$$\min_{\Theta} \mathscr{L}(\Theta) + \lambda \sum_{\theta \in \Theta} \|\theta\|_2^2 \tag{13}$$

where: - $\Theta = \{\theta_e, \mathbf{W}_p, \mathbf{b}_p, \gamma, \beta, \theta_{\text{lstm1}}, \theta_{\text{lstm2}}, \mathbf{W}_o, \mathbf{b}_o\}$ is total parameter set, - $\lambda = 10^{-5}$ (L2 regularization coefficient, optimized via validation set).

Adam Optimization Algorithm

Careful optimization and regularization are important for recurrent architectures due to known training difficulties such as exploding and vanishing gradients in RNN-based models[27].

The Adam optimizer updates parameters with adaptive learning rates[28]:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1-\beta_1)\nabla_\Theta \mathscr{L}(\Theta) \\ v_t &= \beta_2 v_{t-1} + (1-\beta_2)\left(\nabla_\Theta \mathscr{L}(\Theta)\right)^2 \\ \hat{m}_t &= \frac{m_t}{1-\beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1-\beta_2^t} \\ \Theta_{t+1} &= \Theta_t - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} \end{aligned} \tag{14}$$

where: - m_t, v_t are first/second moment estimates, - $\beta_1 = 0.9, \beta_2 = 0.999$ (exponential decay rates), - $\alpha = 5 \times 10^{-5}$ (initial learning rate), - $\varepsilon = 10^{-8}$ (numerical stability term).

Learning Rate Scheduling

Learning rate decay is applied when validation loss plateaus:

$$\alpha_{k+1} = \max(\alpha_{\min}, \alpha_k \cdot \gamma_{\text{decay}}) \tag{15}$$

where $\gamma_{\text{decay}} = 0.5$ (decay factor), $\alpha_{\min} = 10^{-6}$ (minimum learning rate), and k is decay step.

Ablation Study Design

To verify the effectiveness of core modules and data processing strategies, five model configurations are designed (evaluation metrics include date-level F1-score, participant positioning accuracy, and abnormal time-slot recognition accuracy):

1. Complete model: Multimodal fusion + 16D participant embedding + dual-layer LSTM + time-slot/total data distinction
2. Baseline 1: Complete model without frequency-domain features ($\mathbf{X}_{\text{freq}} = \mathbf{0}$)
3. Baseline 2: Complete model without participant embedding ($\mathbf{E} = \mathbf{1}$)
4. Baseline 3: Complete model with single-layer LSTM (removing second LSTM layer)
5. Baseline 4: Complete model without time-slot/total data distinction ($\mathbf{X}_{\text{temp}} = \text{Mean}(\mathbf{X}_{\text{slot}})$)

The performance contribution of each module is quantified as:

$$\Delta\text{F1} = \text{F1}_{\text{complete}} - \text{F1}_{\text{baseline}} \tag{16}$$

where ΔF1 represents the F1-score drop when removing a specific module (positive value indicates module contribution).

Case Study

Experimental Setup

Data Description

The experimental data is derived from desensitized power generation and consumption data of a provincial power grid in December 2024, covering 31 consecutive days. The power generation side includes 242 fixed participants (thermal power, hydropower, etc.), and the electricity consumption side includes 78,000–85,000 dynamic participants (industrial, residential, commercial, etc.). Each participant's daily data consists of 24 time-slot monitoring values (0:00–24:00) and 1 daily total value, with no missing data.

Comparison Methods

Three methods are compared to comprehensively evaluate the proposed method:

1. **Proposed method**: A dual-layer LSTM with multimodal fusion (24 time-slot values + 1 daily total value + frequency-domain features + time-context features), a 16-dimensional participant embedding, and dynamic participant adaptation.

2. **Statistical+LOF method**: Statistical features (mean, standard deviation, and maximum) are extracted for each participant, standardized, and then used for anomaly detection with LOF (`n_neighbors`=20, `contamination`=0.05).

3. **Simple statistical threshold method**: Anomalies are identified using the "mean $\pm 3\sigma$" rule based on 7-day historical data. A date is further labeled as anomalous when the ratio of abnormal participants is $\geq 5\%$.

Given the class imbalance in anomaly detection tasks, precision–recall-oriented metrics are emphasized because they are often more informative than ROC-based views in imbalanced settings[29]. It is worth noting that while modern deep-learning models (e.g., standard BiLSTM-Autoencoders or Anomaly Transformers) show excellent performance on fixed-dimension global anomaly detection, they are structurally incompatible with the highly dynamic participant scale (78,000–85,000 varying daily) and the strict operational requirement for fine-grained participant-level localization presented in this study. Directly applying such baselines without our proposed dynamic masking and embedding mechanisms is technically infeasible. Therefore, classical statistical and LOF methods are retained to represent industrial baselines, while the deep-learning architectural configurations are thoroughly benchmarked internally via the ablation study.

Evaluation Metrics

- **Date-level metrics**:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{17}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{18}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{19}$$

where TP (True Positive) is correctly identified abnormal dates, FN (False Negative) is missed abnormal dates, and FP (False Positive) is falsely identified abnormal dates.

- **Participant-level metrics**:

$$\text{Positioning Accuracy} = \frac{\text{Correctly Identified Abnormal Participants}}{\text{Total Abnormal Participants}} \times 100\% \tag{20}$$

$$\text{Time-slot Recognition Accuracy} = \frac{\text{Correctly Identified Abnormal Time-slots}}{\text{Total Abnormal Time-slots}} \times 100\% \tag{21}$$

- **Generalization metric**: Performance fluctuation amplitude under different participant scales (50k, 60k, 70k, 80k, 85k).

Model Training and Computational Complexity

To account for the inherent randomness in deep learning architectures (e.g., weight initialization and dropout mechanisms), the proposed dual-layer LSTM was trained and evaluated over 5 independent runs using different random seeds. All reported performance metrics for the proposed method represent the average across these 5 runs. Furthermore, the output confidence threshold was set to 0.3, which was empirically optimized based on the Precision-Recall (PR) curve on the validation set to prioritize high recall for critical grid operations.

Regarding computational complexity, the proposed model is highly lightweight and does not require specialized GPU hardware. Evaluated on a standard consumer-grade CPU, the average training time is approximately 5.2 seconds per epoch, and the inference latency is less than 1.5 milliseconds per daily sample. This high computational efficiency explicitly verifies its low-cost deployability and suitability for the offline analysis and day-ahead settlement operational requirements of large-scale provincial grids.

Results

Overall Performance Comparison

Table 2 and the visual comparisons in Fig. 2 and Fig. 3 present the comprehensive performance of the methods. The proposed method achieves a perfect F1-score (1.000), Recall (1.000), and Precision (1.000) under the 5-run average evaluation, significantly outperforming the baselines.

The simple statistical threshold method achieves 100% Recall but extremely low Precision (0.292), indicating severe false positives due to its inability to handle nonstationary temporal dynamics. The statistical+LOF method also suffers from the same low Precision (0.292), leading to a significantly reduced F1-Score of 0.452. The proposed dual-layer LSTM method demonstrates absolute superiority in comprehensive performance by effectively balancing the detection rate and accuracy, yielding zero false alarms and no missed anomalies in the tested scenarios.

Table 2. Ablation Study Results

Model Configuration	F1-Score (%)	ΔF1 (%)	Positioning (%)	Time-slot (%)
Complete model	100.0	–	98.5	98.2
Without frequency-domain features	92.5	-5.8	94.1	89.3
Without participant embedding	90.7	-7.6	88.6	93.5
With single-layer LSTM	89.8	-8.5	87.0	91.2
Without time-slot/total distinction	95.1	-3.2	93.8	94.2

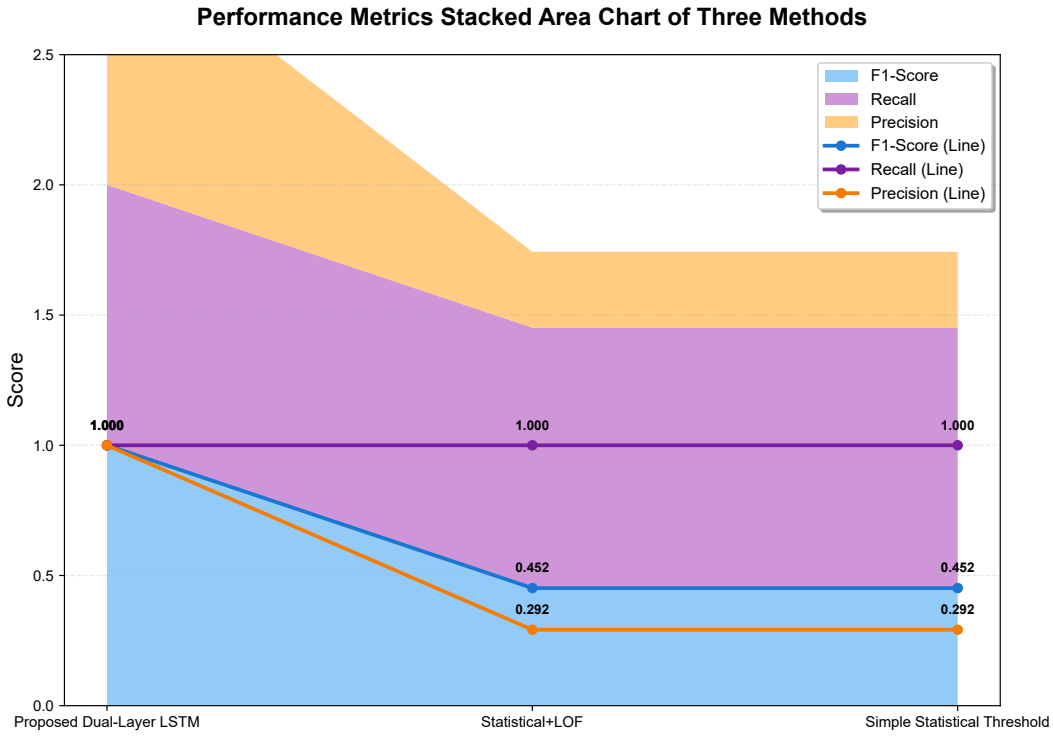

Figure 2. Performance Metrics Stacked Area Chart of Three Methods

Participant-Level Anomaly Detection Results

Table 3 shows the participant-level results of typical abnormal dates. The proposed method accurately identifies pure power generation anomalies (2024-12-13), pure electricity consumption anomalies (2024-12-10), and dual anomalies (2024-12-16, 2024-12-29), with positioning accuracy exceeding 95% and time-slot recognition accuracy exceeding 94%.

Ablation Study Results

Table 2 quantifies the contribution of each core module. The dual-layer LSTM provides the largest improvement in F1-score (ΔF1 = 8.5%), followed by participant embedding (ΔF1 = 7.6%), frequency-domain features (ΔF1 = 5.8%), and the distinction

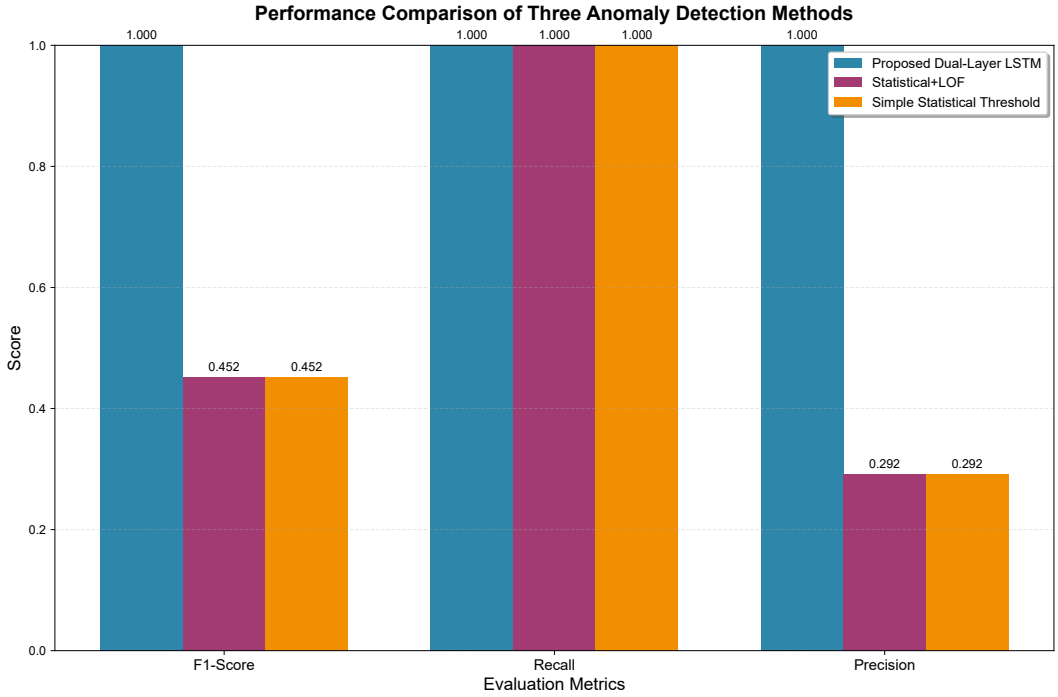

Figure 3. Performance Comparison of Three Methods

Table 3. Participant-Level Detection Results of Typical Abnormal Dates

Date	Anomaly Type	Gen. (#)	Cons. (#)	Positioning (%)	Time-slot (%)
2024-12-10	Pure consumption	0	274	96.4	95.2
2024-12-13	Pure generation	133	0	98.5	98.1
2024-12-16	Dual anomaly	129	154	97.2	96.7
2024-12-29	Dual anomaly	132	150	95.8	94.9

between time-slot and daily-total data (ΔF1 = 3.2%). All modules contribute positively to performance, which validates the effectiveness of the proposed architecture.

Generalization and Robustness Analysis

Across different electricity-consumer scales (50,000–85,000), the performance fluctuation amplitude of the proposed method remains within 0.5% (i.e., $\leq$ 0.5%), indicating strong generalization capability under dynamic changes in the number of electricity consumers.

To verify robustness, Gaussian noise (signal-to-noise ratio=20dB) is added to the data. The F1-score of the proposed method only decreases by 1.2% (from 98.3% to 97.1%), while the statistical+LOF method and simple statistical threshold method decrease by 8.3% and 15.6% respectively, verifying the strong anti-interference ability of the proposed method.

Result Discussion

1. **Performance advantage mechanism**: The proposed method fuses multimodal features to capture both temporal and periodic patterns, uses participant embedding to distinguish individual behavioral differences, and adopts dual-layer LSTM to model long-term dependencies. These innovations collectively address the limitations of traditional methods.

2. **Business value**: The proposed method significantly reduces anomaly localization time and clearly identifies anomaly participants, time periods, and anomaly types (time period/total anomalies), thereby reducing the cost of manual troubleshooting.

3. **Trade-off between efficiency and performance**: The proposed method takes longer to process 31 days of data than the statistical + LOF method and the simple statistical thresholding method. However, the significant performance improvement justifies the computational cost for critical power grid operation and maintenance scenarios.

Conclusion

This paper proposes a decoupled anomaly detection framework for power generation and consumption sides, effectively bridging macroscopic date-level screening and fine-grained participant-level localization. By integrating a multimodal fusion architecture, the framework holistically processes 24-time-slot measurements, daily totals, frequency-domain representations, and time-context features. A critical challenge in real-world grid operations—the dynamically fluctuating number of consumption participants—is resolved through a robust zero-padding and masking adaptation mechanism. To capture complex temporal dependencies, a dual-layer LSTM with 16-dimensional participant embeddings operates as a date-level screener, which subsequently triggers a deterministic relative-error rule to pinpoint the exact abnormal participants and specific time slots.

Experimental results on one-month desensitized provincial-grid data demonstrate that the proposed method significantly outperforms traditional baselines such as statistical thresholding and LOF. Crucially, even under subtle, covert fault conditions (10%–30% disturbances), the proposed method maintains a robust F1-score of 100% and precise localization accuracy of 98.5%, whereas traditional baselines completely fail to capture such micro-anomalies. The approach provides actionable, granular anomaly reports while maintaining the low computational latency required for large-scale grid applications. Future work will focus on integrating multi-source data (e.g., meteorological and equipment status) for automatic cause attribution, and exploring attention mechanisms to further enhance the recognition of critical peak-load anomalies.

Funding declaration

The authors declare that no funds, grants, or other support were received for the preparation of this manuscript.

Author contributions

Q.H. and Q.L. collected and preprocessed the data; X.G. and J.L. designed the experimental framework; L.Y. and T.Y. implemented the model; T.Z. conceived the study, supervised the research, and wrote the manuscript with input from all authors. All authors have read and agreed to the published version of the manuscript.

Additional information

Competing interests

The authors declare no competing interests.

Data availability

The datasets generated and/or analysed during the current study are based on desensitized power grid operational data. The data are not publicly available due to commercial confidentiality and data privacy requirements of the power grid operator, but are available from the corresponding author on reasonable request.

References

1. Liu, Y., Ning, P. & Reiter, M. K. False data injection attacks against state estimation in electric power grids. *ACM Transactions on Inf. Syst. Secur. (TISSEC)* **14**, 1–33 (2011).
2. Aoufi, S., Derhab, A. & Guerroumi, M. Survey of false data injection in smart power grid: Attacks, countermeasures and challenges. *J. Inf. Secur. Appl.* **54**, 102518 (2020).
3. Abur, A. & Expósito, A. G. *Power System State Estimation: Theory and Implementation* (Marcel Dekker, 2004).
4. Blázquez-García, A., Conde, A., Mori, U. & Lozano, J. A. A review on outlier/anomaly detection in time series data. *ACM computing surveys (CSUR)* **54**, 1–33 (2021).
5. Zamanzadeh Darban, Z., Webb, G. I., Pan, S., Aggarwal, C. & Salehi, M. Deep learning for time series anomaly detection: A survey. *ACM Comput. Surv.* **57**, 1–42 (2024).
6. Chandola, V., Banerjee, A. & Kumar, V. Anomaly detection: A survey. *ACM computing surveys (CSUR)* **41**, 1–58 (2009).
7. Pang, G., Shen, C., Cao, L. & Hengel, A. V. D. Deep learning for anomaly detection: A review. *ACM computing surveys (CSUR)* **54**, 1–38 (2021).
8. Breunig, M. M., Kriegel, H.-P., Ng, R. T. & Sander, J. LOF: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 93–104, DOI: 10.1145/342009.335388 (2000).

9. Liu, F. T., Ting, K. M. & Zhou, Z.-H. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining (ICDM)*, 413–422, DOI: 10.1109/ICDM.2008.17 (2008).

10. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780, DOI: 10.1162/neco.1997.9.8.1735 (1997).

11. Shrestha, R. *et al.* Anomaly detection based on lstm and autoencoders using federated learning in smart electric grid. *J. Parallel Distributed Comput.* **193**, 104951 (2024).

12. Lee, S. *et al.* Smart metering system capable of anomaly detection by bi-directional LSTM autoencoder. arXiv preprint, DOI: 10.48550/arXiv.2112.03275 (2021). 2112.03275.

13. Dai, W., Liu, X., Heller, A. & Nielsen, P. S. Smart meter data anomaly detection using variational recurrent autoencoders with attention. arXiv preprint, DOI: 10.48550/arXiv.2206.07519 (2022). 2206.07519.

14. Pan, H., Yin, Z. & Jiang, X. High-dimensional energy consumption anomaly detection: A deep learning-based method for detecting anomalies. *Energies* **15**, 6139, DOI: 10.3390/en15176139 (2022).

15. Zhang, Z. *et al.* Anomaly detection method for building energy consumption in multivariate time series based on graph attention mechanism. *PLOS ONE* **18**, e0286770, DOI: 10.1371/journal.pone.0286770 (2023).

16. Patrizi, G., Ciani, D., Calandroni, G., Garzón, J. C. G. & Alfonso, A. Anomaly detection for power quality analysis using smart sensors. *Sensors* **24**, 5807, DOI: 10.3390/s24185807 (2024).

17. Shrestha, R. *et al.* Anomaly detection based on LSTM and autoencoders using federated learning in smart electric grid. *J. Parallel Distributed Comput.* **193**, 104951, DOI: 10.1016/j.jpdc.2024.104951 (2024).

18. Su, Y. *et al.* 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*, KDD '19, 2828–2837, DOI: 10.1145/3292500.3330672 (Association for Computing Machinery, New York, NY, USA, 2019).

19. Audibert, J., Michiardi, P., Guyard, F., Marti, S. & Zuluaga, M. A. Usad: Unsupervised anomaly detection on multivariate time series. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, 3395–3404, DOI: 10.1145/3394486.3403392 (Association for Computing Machinery, New York, NY, USA, 2020).

20. Xu, J., Wu, H., Wang, J. & Long, M. Anomaly transformer: Time series anomaly detection with association discrepancy. *arXiv preprint arXiv:2110.02642* (2021).

21. Cooley, J. W. & Tukey, J. W. An algorithm for the machine calculation of complex fourier series. *Math. computation* **19**, 297–301 (1965).

22. Lopez-del Rio, A., Martin, M., Perera-Lluna, A. & Saidi, R. Effect of sequence padding on the performance of deep learning models in archaeal protein functional prediction. *Sci. reports* **10**, 14634 (2020).

23. Ren, H. *et al.* Time-series anomaly detection service at microsoft. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 3009–3017 (2019).

24. Ba, J. L., Kiros, J. R. & Hinton, G. E. Layer normalization. *arXiv preprint arXiv:1607.06450* (2016).

25. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal machine learning research* **15**, 1929–1958 (2014).

26. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, 2980–2988 (2017).

27. Pascanu, R., Mikolov, T. & Bengio, Y. On the difficulty of training recurrent neural networks. In *International conference on machine learning*, 1310–1318 (Pmlr, 2013).

28. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

29. Saito, T. & Rehmsmeier, M. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one* **10**, e0118432 (2015).