Anomaly Detection in Smart Grid Behavior Monitoring via Federated Learning: A Privacy-Preserving Defense Against Cyber-Physical Attacks

Xiaozhi Deng[1,*], Yaoxin Pan[1] and Hongjia Fang[2]

[1]*Guangdong Grid Electric Power Dispatching Control Center, Guangzhou 510220, Guangdong, China*
[2]*China Southern Power Grid Digital Enterprise Technology (Guangdong) Co., Ltd., Guangzhou 510000, Guangdong, China*
E-mail: whxv22@163.com
[*]*Corresponding Author*

Received 04 September 2025; Accepted 21 October 2025

Abstract

The smart grid has achieved digitalization and interconnection of power systems through the integration of the Internet of Things (IoT), communication networks, and automation technologies. However, these advancements have also made the smart grid a prime target for cyberattacks. To ensure stable operation and protect user privacy, this study integrates long short-term memory (LSTM) networks, the K-means clustering algorithm, and federated learning to design a behavior anomaly detection method. The proposed method effectively safeguards the privacy and security of the smart grid against physical attacks, thereby ensuring its stable operation. Experimental results demonstrate that the fusion algorithm achieves a feature extraction accuracy of 98.7%, while the anomaly detection error rate remains as low as 2.3%. Furthermore, under attack scenarios, the method reduces the risk of privacy leakage by 92.1% and successfully resists over 90% of

Journal of Cyber Security and Mobility, Vol. 14_5, 1151–1172.
doi: 10.13052/jcsm2245-1439.1455

physical attacks. These findings indicate that the proposed detection method can not only accurately monitor abnormal behaviors in the smart grid but also provide robust protection for grid security in the event of an attack.

Keywords: Privacy attacks, federated learning, power grid safety, long short-term memory network, K-means clustering algorithm.

Introduction

The Smart Grid (SG) represents a transformative power system (PS) that emerges from the seamless integration of renewable energy sources, advanced materials, and sophisticated equipment, together with the incorporation of modern sensing, information, and control technologies into the traditional PS framework [1]. However, as society continues to develop, the scale of SG utilization is rapidly expanding, and the associated cybersecurity threats are escalating in severity [2–3]. The SG has become a frequent target of cyberattacks, which may result in power supply interruptions, equipment failures, and the exposure of user privacy information [4].

To address these risks, researchers have proposed a variety of methods for detecting SG behavior anomalies and safeguarding privacy, aiming to ensure secure and reliable grid operations. For example, Jithish J. et al. proposed a machine learning (ML)-based anomaly detection (AD) method capable of identifying abnormal behavior (AB) in the SG, achieving a detection precision of 92.3% when applied in practice [5]. Wang Z. et al. introduced an AD method based on a grid-to-vector framework, which improved upon the poor detection performance of existing techniques, reaching an average detection accuracy of 99.21% [6]. Similarly, Guha D. et al. developed a privacy-preserving approach for SG data based on autoencoders, which reduced privacy leakage (PL) risk by 43.4% [7]. To enhance resistance to data injection attacks, Takidin A. et al. proposed a generalized graph neural network-based method that successfully resisted more than 90% of such attacks [8].

Although many approaches have been developed for anomaly detection and privacy protection in SGs, most still exhibit limitations such as low accuracy in anomaly detection and insufficient resilience against physical attacks [9]. This highlights the necessity of optimizing current detection strategies.

Federated Learning (FL) is a distributed machine learning paradigm that enables collaborative model training across decentralized datasets without

requiring raw data sharing, thereby enhancing privacy protection [10]. The K-means clustering algorithm is a well-established clustering technique that partitions datasets into distinct groups [11]. Long Short-Term Memory (LSTM), a specialized form of recurrent neural network, excels at capturing long-range dependencies in sequential data [12]. Owing to their unique strengths, these algorithms have been widely applied across multiple domains. For instance, Zhou X. et al. proposed an FL-based detection method to improve the privacy and security of mobile robot communication networks, achieving a 34.5% enhancement in security [13]. Liu H. et al. applied K-means clustering to simplify complex problem-solving processes, improving computational efficiency by 32.3% [14]. In another study, Zhang X. et al. developed a load prediction model using LSTM to address the low accuracy of existing PS load prediction approaches, achieving a prediction accuracy of 92.3% [15].

In summary, while significant progress has been made in anomaly detection and privacy protection for SGs, existing methods still demonstrate weak resistance to physical attacks and inadequate protection against privacy leakage. To overcome these challenges, this study proposes an SG anomaly behavior detection model that integrates K-means clustering, LSTM networks, and FL. In the proposed framework, LSTM is deployed at power grid (PG) nodes to extract feature data, which are then transmitted to a central server via the FL framework for collaborative computation. This design reduces the risk of privacy leakage during data transmission while enhancing security and improving resilience against external physical attacks. The innovation of the research lies in deploying LSTM algorithm in the nodes of the power grid to obtain power grid feature data, and then transmitting the feature data to the central server for calculation through the FL framework, thereby reducing the risk of privacy leakage during data transmission and improving its security.

1 Privacy Protection Method for Cyber Physical Attacks Based on FL

1.1 FL algorithm combining K-means and LSTM

Intelligent grid behavior AD can detect abnormal situations in a timely manner by monitoring various parameters of the PS in real time, thereby avoiding power failures and ensuring the steadiness and dependability of the PS [16]. By detecting abnormal information in the PS, real-time management of the PS can be carried out, reducing the operation and maintenance costs of

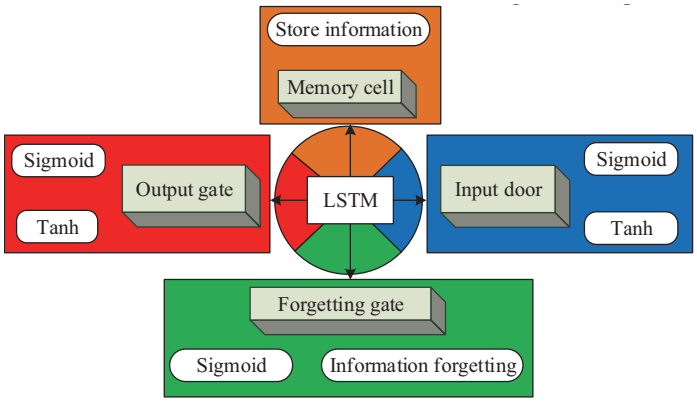


Figure 1 Basic structure of LSTM algorithm.

the PS and improving its operational and upkeep productivity [17]. However, the current intelligent grid AD has poor resistance to cyber physical attacks and needs to be optimized to guarantee the safety of private data in the grid. The core idea of the FL algorithm is to train the model on local data from multiple participants, and then aggregate the model parameters of each participant to obtain a global model [18]. This method can effectively avoid centralized storage and transportation of raw data, and effectively protect user privacy. In the FL algorithm, other algorithms are often used to extract feature information from the local data of the participants, integrate data features, and analyze data features using intelligent algorithms in the central server of FL. The LSTM algorithm can automatically learn the temporal dependencies in data through its own neural network, extract feature information from the data, and is often used in data feature extraction. So, the study deployed the LSTM algorithm in FL and utilized its feature extraction capability to extract data feature information from FL. The basic structure of the LSTM algorithm is presented in Figure 1.

In Figure 1, the LSTM algorithm mainly comprises memory cells, forget gates (FGs), input gates (IGs), and output gates (OGs). Among them, memory cells are the core part of LSTM, used to store information. The FG plays a pivotal role in determining which unimportant details to eliminate from the cell's memory state. It achieves this by employing a sigmoid layer (SL) that produces a value ranging from 0 to 1, indicating whether the information should be forgotten. Meanwhile, the IG is tasked with updating the memory cell's state. This objective is achieved by integrating an SL, which is responsible for choosing the information to be updated, with a

tanh layer (TL), which produces new candidate values that are subsequently incorporated into the memory cells. Similarly, the OG is responsible for deciding which information to convey. This gate also comprises an SL and a TL. The SL selects which portions of the cell state are to be outputted, while the TL reprocesses the cell state to yield a value within the range of -1 to 1. The two layers are multiplied to obtain the final output. In this process, the computation for the data in the IG is shown in equation (1).

$$I_t - \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \tag{1}$$

In equation (1), I_t is the activation value (AV) of the IG, σ is the sigmoid function, W_{xi} and W_{hi} are the weight matrices (WMs) of the hidden states of the IG at the current time step (TS) and the previous TS. x_t is the input of the current TS, h_{t-1} is the hidden state of the previous TS, and b_i is the bias term of the IG. The computation in the FG is shown in equation (2).

$$F_t - \sigma(W_{xf}xt + W_{hf}ht - 1 + b_f) \tag{2}$$

In equation (2), F_t is the AV of the FG, W_{xf} and W_{hf} are the WMs of the hidden states of the FG at the current TS and the previous TS, respectively. b_f is the bias term of the FG. The computation for the OG is shown in equation (3).

$$O_t - \sigma(W_{xo}xt + W_{ho}ht - 1 + b_o) \tag{3}$$

In equation (3), O_t is the AV of the OG, W_{xo} and W_{ho} are the WMs of the hidden states of the OG at the current TS and the previous TS. b_o is the bias term of the OG. The computation in memory cells is shown in equation (4).

$$G_t - \tanh(W_{xc}xt + W_{hc}ht - 1 + b_c) \tag{4}$$

In equation (4), $\tanh$ is the hyperbolic tangent function, G_t is the candidate memory cell value, W_{xc} and W_{hc} are the WMs of the hidden states of the memory cell at the current TS and the previous TS, and b_c is the bias term of the memory cell. Following the extraction of feature information from the data via the LSTM algorithm, it becomes crucial to classify the data within the PG in order to streamline subsequent analysis efforts focused on discerning whether the PG is under attack. The K-means can classify data with similar features into one category, thereby achieving the classification of PG behavior. So after feature extraction using LSTM algorithm in FL, K-means algorithm is used to classify the feature data. Figure 2 illustrates the fundamental procedure of the K-means algorithm.

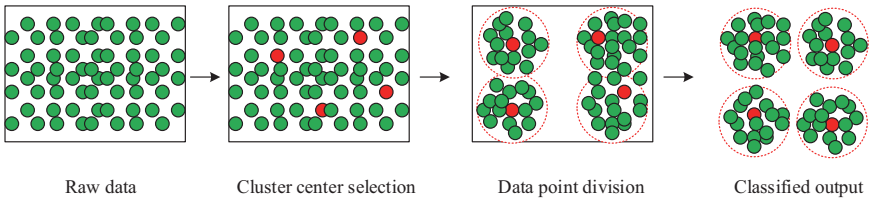


Figure 2 Fundamental procedure of K-means.

In Figure 2, the K-means algorithm starts by randomly picking K data points from the dataset as the initial cluster centroids. Then, it assigns each data point to the closest centroid by measuring the distance between every point and the centroids. Finally, it recalculates the centroid of each cluster by taking the average of all points in that cluster, and this new mean is employed as the updated center of the cluster. These steps of data point allocation and cluster center updating are repeated until the cluster centers cease to change or the max iteration count is attained. The final output encompasses the cluster centers and the division of clusters. In the above process, the calculation method for the distance between data points and cluster centers is shown in equation (5).

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (xi - yi)^2} \tag{5}$$

In equation (5), $d(x, y)$ is the Euclidean distance between two points x and y, xi and yi are the coordinates of points x and y in the ith dimension, respectively, n is the total number of dimensions of points. The computation for cluster center update in K-means is shown in equation (6).

$$\mu_k = \frac{1}{|C_k|} \sum_{x \in C_k} x \tag{6}$$

In equation (6), μ_k is the cluster center, $|C_k|$ is the quantity of samples in cluster C_k, and x is the samples in cluster C_k. The K-means-LSTM algorithm can extract features of various data in the SG and classify them. Within the framework of the FL algorithm, the LSTM algorithm operates as the feature extraction mechanism for participants in the FL system. Concurrently, the K-means algorithm serves as the data analysis tool at the central server of the FL system, with the objective of detecting anomalous information within the dataset. The basic process of the FL algorithm based on K-means-LSTM is shown in Figure 3.

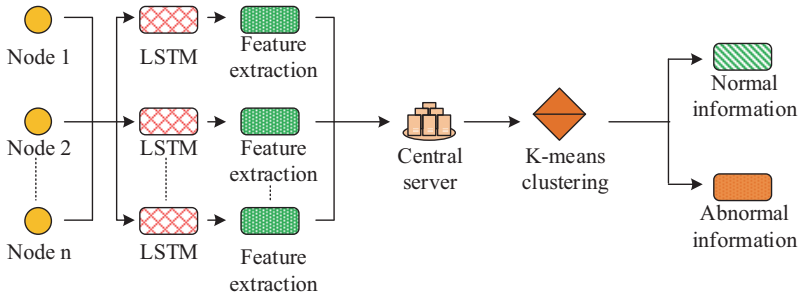


Figure 3 K-means-LSTM-FL algorithm flow.

As shown in Figure 3, the FL algorithm can perform distributed processing on datasets from different locations in the SG. LSTM is deployed to each data node, and the LSTM algorithm is employed to extract the data features of each node. The extracted feature information is then sent to the central server, which utilizes the K-means to cluster and analyze the feature data from each node. In this way, grid behavior data with similar features are classified into one category, making it easier to identify abnormal data in the future.

1.2 Grid behavior AD method based on FL

When checking for PG anomalies, it is necessary to distinguish between abnormal data and normal data. Consequently, the K-means-LSTM-FL algorithm is applied to identify unusual activities within the PG. The basic structure of the PG behavior AD method based on K-means-LSTM-FL algorithm is shown in Figure 4.

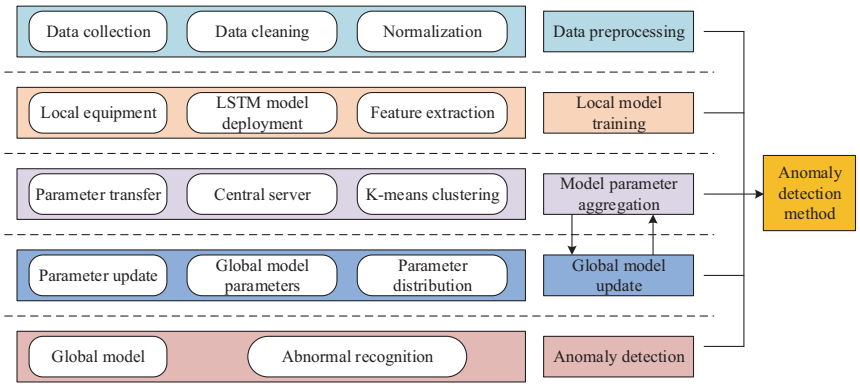


Figure 4 Basic structure of AD approach.

In Figure 4, the detection method comprises five modules: data preprocessing (DP), local model training (LMT), model parameter aggregation (MPA), global model update (GMU), and AD. In the DP section, once data from the PG is gathered via sensors and other equipment, preprocessing steps like data cleansing and normalization are essential to enhance data quality. The expression for normalization operation is shown in equation (7).

$$p' = \frac{p - \min(P)}{\max(P) - \min(P)} \tag{7}$$

In equation (7), p' is the normalized data, p is the original data, $\min(P)$ and $\max(P)$ is the minimal and maximal values in the dataset. In LMT, after deploying LSTM models at each data collection point in the PG, the LSTM model is employed to extract features from the data collected at each collection point, including voltage fluctuation characteristics, current change characteristics, and other feature information related to PG behavior. MPA is to send the feature information extracted by LSTM to a central server, where K-means is used to classify PG behavior data with similar features, in order to achieve classification and induction of PG behavior. Within the GMU, the central server computes weighted averages of the model parameters from each node in the PG, thereby generating global model parameters. These global parameters are subsequently disseminated to every node, serving as a guide for model training and contributing to an improvement in the accuracy of feature extraction at each node.

The formula for weighted average is shown in equation (8).

$$\theta_{global} = \frac{1}{\sum_{a=1}^{A} w_a} \sum_{a=1}^{A} w_a \theta_a \tag{8}$$

In equation (8), θ_{global} is the global model parameter, θ_a is the model parameter of the ath node, w_a is the weight of the ath node, and A is the number of nodes. The K-means-LSTM-FL algorithm gradually improves the model's capacity for generalization through multiple rounds of parameter aggregation and GMUs. The AD module leverages a global model to pinpoint AB within the PG, thereby safeguarding the secure operation of the system. It accomplishes this by closely monitoring the clustering results to ascertain the presence of AB in the PG. Firstly, a threshold for normal behavior is set for each category. Upon clustering the characteristic behaviors of the PG utilizing the K-means-LSTM-FL approach, if the feature values surpass the threshold established for the corresponding clustering category, it can be concluded that AB is present within the PG.

1.3 Privacy protection method based on abnormal detection of PG behavior

K-means-LSTM-FL PG AD can detect AB of the PG during network physical attacks, thereby ensuring the security of the PG. A cyber-physical attack denotes a malicious technique that disrupts a system by leveraging network-based methods, such as tampering with system data, pilfering sensitive information, and altering configurations, ultimately impairing the system's normal operational capabilities. The K-means-LSTM-FL AD model can protect data privacy through the FL framework. In this framework, the data nodes of the PG only send the feature data and parameters processed by LSTM to the central server, and do not send the original PG behavior data, which can effectively avoid the risk of PL and attack during data transmission. The specific structure of the privacy protection method based on the K-means-LSTM-FL AD model is shown in Figure 5.

As illustrated in Figure 5, this method comprises a data node layer, an edge computing layer, and a central server layer. Within the data node layer, each node in the SG is designated as a data node. Data collection devices are employed to gather various types of data from these nodes, and the collected data are stored locally at each data node. Subsequently, an LSTM model is utilized to extract features from the data nodes. The feature data, extracted by the LSTM algorithm, are then transmitted to the central server via a secure

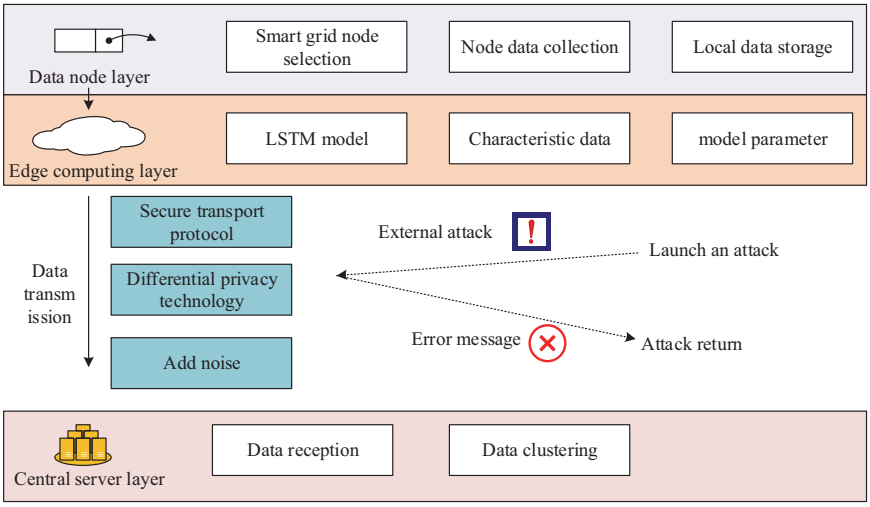


Figure 5 Structure of privacy protection method.

transmission protocol, with only the feature information being conveyed during this process. Furthermore, differential privacy technology is introduced, which represents a robust privacy protection mechanism capable of adding an appropriate amount of noise to the feature data. This prevents attackers from obtaining accurate information from the data while simultaneously ensuring its availability. The definition of differential privacy is shown in equation (9).

$$\frac{\Pr[M(D) \in S]}{\Pr[M(D') \in S]} \leq e^{\varsigma} \tag{9}$$

In equation (9), D and D' are two adjacent data in the transmitted feature dataset, S is any possible output, ς is the privacy budget parameter, the smaller the value, the more robust the privacy protection becomes. M is a randomized algorithm, where the input is denoted as D and the output as S. $\Pr[M(D) \in S]$ denotes the probability that algorithm M outputs the set S when the input is the dataset D. $\Pr[M(D') \in S]$ denotes the probability that algorithm M outputs the set S when the input is an adjacent dataset D'. If the given expression is satisfied, then algorithm M is said to satisfy the differential privacy of ς. Differential privacy noise is added before data transmission and abnormal detection result output. By setting the noise at an appropriate level, the accuracy of PG abnormal detection can be ensured, while the risk of PL during the detection process can be effectively minimized. This method achieves privacy protection in PG behavior monitoring and abnormal detection, and also enhances resilience against external network physical attacks.

2 Analysis of Privacy Protection Effect of SG

2.1 Performance analysis of K-means-LSTM-FL algorithm

To analyze the privacy protection effect of the introduced K-means-LSTM-FL PG AD model, the capability of the K-means-LSTM-FL algorithm needed to be analyzed first, including its feature extraction and clustering effects on PG data. During algorithm experimentation, the environmental setup is detailed in Table 1.

When analyzing the performance of the algorithm, the datasets from the CIFAR-10/CIFAR-100 database were selected as the experimental datasets. These datasets consisted of 60,000 32×32 color images across 10 classes, with 6,000 images per class. There were 50,000 training images and 10,000 test images, and the data source website was “https://www.cs.toronto.edu/~k

Table 1 Experimental setup

Project	Index	Allocation
Hardware environment	CPU	Intel Core i7-12700K
	GPU	NVIDIA RTX 309
	RAM	64GB DDR4
	OS	Windows 10
Software environment	Deep learning framework	PyTorch 1.12
	ML library	Scikit-learn 1.1
	Data analysis software	Python 3.9

riz/cifar.html. In this study, image information comprising 500 images from each of four classes (airplane (A), automobile (B), bird (C), and cat (D)) was randomly selected from the dataset to serve as the experimental dataset. Only use this dataset to validate the performance of the K-means LSTM-FL algorithm. In the course of the experiment, the LSTM algorithm's local training iterations were configured to 20, the TS to 100, and the learning rate to 0.005. For the K-means, the quantity of clusters was set to 4, and the maximal iteration count was set to100. The experiment was conducted with the above configurations. Firstly, an analysis was carried out on the algorithm's feature extraction effectiveness for the four classes of data, and the results are shown in Figure 6.

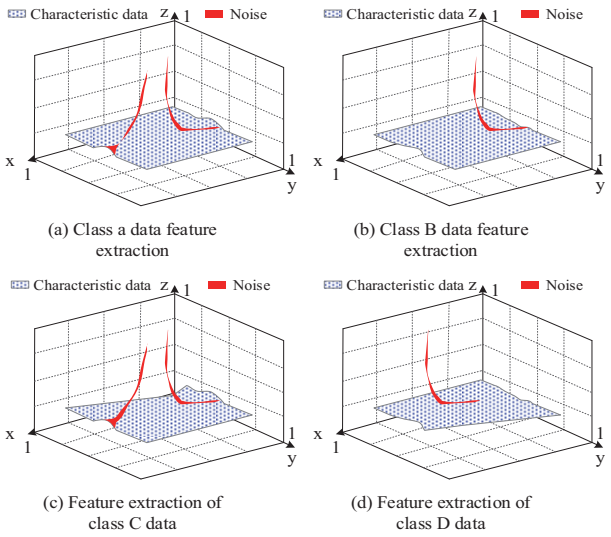
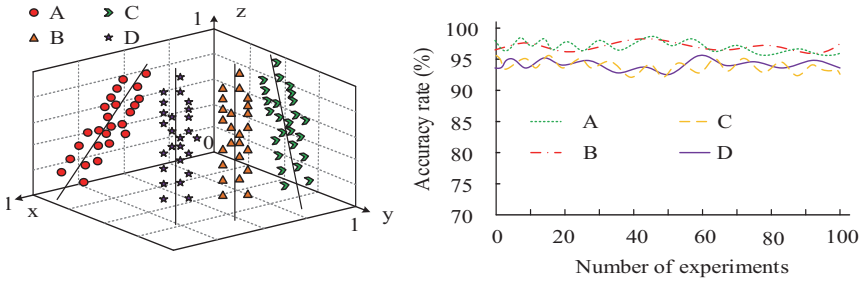


Figure 6 Effect of feature information extraction.



(a) Classification effect (b) Specific classification accuracy

Figure 7 Algorithm classification effect.

As shown in Figure 6(a), after feature extraction of Class A data, the extracted data contained less noise information and had a higher accuracy in feature extraction. From Figure 6(b), 6(c), and 6(d), the K-means-LSTM-FL algorithm also performed well in feature extraction for the other three types of data, with an accuracy rate of 98.7% for all four types of datasets. Further analysis was conducted on the clustering performance of the K-means-LSTM-FL algorithm, and the outcomes are presented in Figure 7.

As shown in Figure 7(a), the algorithm performed well in classifying four types of data: A, B, C, and D, and could accurately classify each type of data. According to the specific results in Figure 7(b), the average accuracy of the algorithm for classifying four types of datasets was 96.2%, 95.6%, 96.1%, and 95.7%, respectively, with relatively high classification accuracy. In order to further verify the robustness of the K-means-LSTM-FL algorithm, the algorithm was tested in environments with 20% Gaussian noise, 20% network delay, and 10% data offset. The results are shown in Table 2.

According to Table 2, the K-means-LSTM-FL algorithm has a classification accuracy of over 95% and a high extraction accuracy of over 96% for feature information when classifying data in environments with interference. From the above results, the proposed K-means-LSTM-FL algorithm could accurately extract and classify feature data.

Table 2 Robustness test results

Environment	Classification accuracy	Feature extraction accuracy	Classification time consumption
20% Gaussian noise	$95.4 \pm 1.8\%$	$97.4 \pm 1.2\%$	1.3 ± 0.1s
20% network latency	$93.5 \pm 0.9\%$	$96.6 \pm 2.6\%$	1.5 ± 0.2s
10% data offset	$95.2 \pm 1.8\%$	$98.5 \pm 2.6\%$	1.4 ± 0.1s

2.2 Effect of abnormal detection of PG behavior

After validating the feature extraction effectiveness and clustering effectiveness of the K-means-LSTM-FL algorithm, an analysis was then conducted on the detection effectiveness of the PG AD method based on this algorithm. The experimental setup was the same as that in Section 2.1. The study chose the dataset on the website "https://gitcode.com/Universal-Tool/ac80a" as the dataset for this study, which contains detailed information on the daily electricity load, temperature, humidity, wind speed, and rainfall of the PG for a period of 13 months. During the experiment, 1000 records were randomly selected from the dataset to confirm the effectiveness of the PG behavior AD model. Firstly, an analysis was carried out on the AD effectiveness of the PG behavior AD method, and the results are shown in Figure 8.

As shown in Figure 8(a), this detection method could accurately detect abnormal voltage behavior and could detect all abnormal data. In Figures 8(b) and 8(c), this detection method could accurately detect current and power anomalies in the PG. Further analysis was conducted on the specific detection error and detection time of the detection method, and the results are shown in Figure 9.

From Figure 9(a), the detection error of this detection method for voltage anomaly information in SGs was 2.1%, and the detection time was also relatively short, only 3.2 seconds. From Figures 9(b) and 9(c), the detection error of this detection method for current anomalies and grid power anomalies was also relatively low, at 2.5% and 2.3%, respectively, and its detection time was also relatively short. The average detection error was 2.3%. Further analysis of the detection performance of this method on attacks such as Advanced Persistent Threat (APT) and Zero day attacks is shown in Table 3.

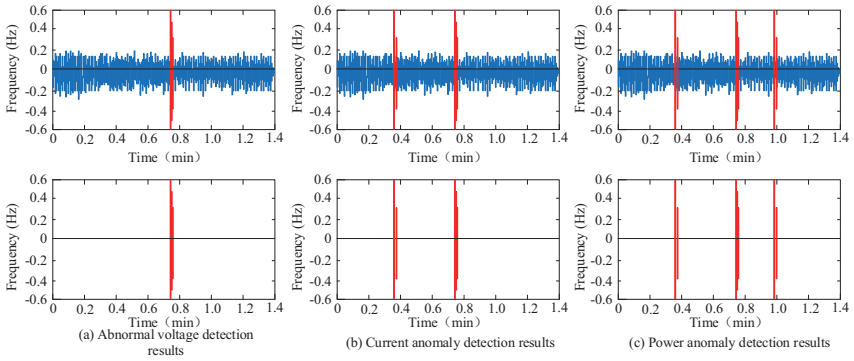


Figure 8 Analysis of abnormal detection effect.

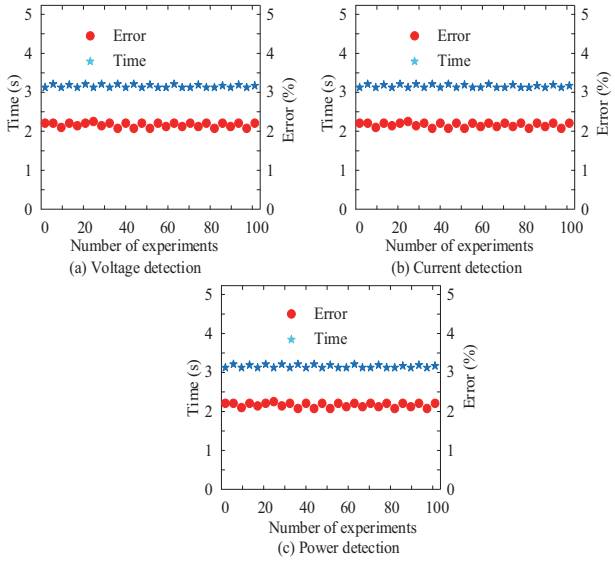


Figure 9 Detection error and detection time.

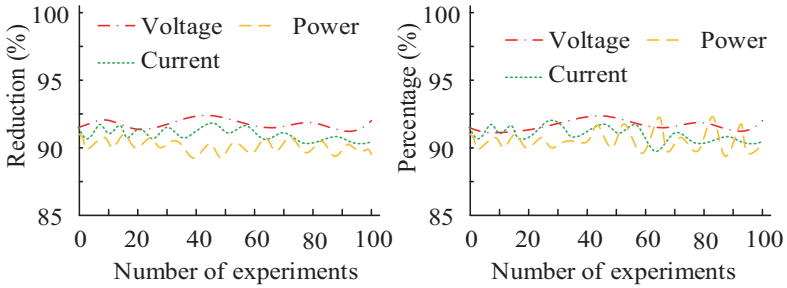
Table 3 Analysis of test results

Indicator	APT	Zero-day attacks
False positive rate	3.2%	2.9%
False negative rate	2.7%	3.2%
Testing time consumption	2.3s	2.4s

According to Table 3, the detection method proposed in the study also performs well in detecting abnormal information generated by APT and Zero day attacks, with false positive and false negative rates both below 4%, and detection time less than 3 seconds. From the above results, the proposed method for detecting AB in the PG based on K-means-LSTM-FL algorithm could detect abnormal information in the PG.

2.3 Privacy protection effect

After validating the effectiveness of the AD model, an analysis was subsequently conducted on the validity of the SG privacy protection approach based on this model. During the analysis, data from the "Smart Grid Real-Time Load Monitoring Dataset" were still utilized as the observation dataset. Additionally, a portion of the data was randomly selected from the same



(a) Privacy disclosure reduction (b) Physical attack mitigation effect

Figure 10 Analysis of privacy protection effect.

"Smart Grid Real-Time Load Monitoring Dataset" to serve as the attack dataset to simulate external physical attacks. Firstly, an analysis was carried out on the PL risks and the effectiveness of resisting external physical attacks of this method, and the results are shown in Figure 10.

As shown in Figure 10(a), when using the proposed method for detecting AB in the PG to protect the privacy data of the PG, the leakage of privacy information of PG voltage, current, and power was significantly reduced, with an average reduction of 92.1%. In Figure 10(b), the introduced approach could resist physical attacks of over 90% on the voltage, current, and power of the SG when resisting external attacks. Finally, to confirm the advantage of the privacy protection approach based on the K-means-LSTM-FL algorithm, A comparison experiment was carried out to evaluate the proposed privacy protection approach against two alternatives: the SG method utilizing Homomorphic Encryption combined with Long Short-Term Memory (HE-LSTM) networks, and the standard Federated Average (FedAvg) algorithm-based technique. The outcomes are presented in Table 4.

According to Table 4, the privacy protection method proposed in this study for SGs could reduce PL by $92.4 \pm 1.2\%$ compared to HE-LSTM and FedAvg privacy protection methods, and could protect most privacy information. The privacy protection method proposed in the study could

Table 4 Analysis of privacy protection effect

Method	K-means-LSTM-FL	HE-LSEM	FedAvg
Privacy disclosure reduction	$92.4 \pm 1.2\%$	$87.4 \pm 0.8\%$	$82.9 \pm 2.7\%$
Physical attack resistance	$90.5 \pm 2.3\%$	$84.5 \pm 1.8\%$	$80.3 \pm 2.9\%$
Success rate of privacy attack	$9.3 \pm 0.9\%$	$12.1 \pm 1.7\%$	$13.4 \pm 1.9\%$
Attack prediction error	$0.2 \pm 0.1\%$	$1.4 \pm 0.3\%$	$2.7 \pm 0.7\%$

achieve a resistance effect of $90.5 \pm 2.3\%$ against physical attacks, higher than HE-LSTM's $84.5 \pm 1.8\%$ and FedAvg's $82.9 \pm 2.7\%$. When attacking the privacy protected PG based on K-means-LSTM-FL, the success rate of the attack was only $9.3 \pm 0.9\%$, which was relatively low. From the above results, the privacy protection method based on K-means-LSTM-FL algorithm proposed in the study could protect data privacy in SGs and effectively resist external physical attacks.

3 Discussion

To enhance the confidentiality and security of SG data and improve resilience against external physical attacks, this study integrates K-means, LSTM, and FL to construct an anomaly behavior (AB) detection model for the SG. Based on this model, privacy in the power grid (PG) is effectively protected, while resistance to external attacks is strengthened.To validate the proposed approach, the performance of the K-means-LSTM-FL algorithm was evaluated. Results showed that the algorithm achieved a feature extraction accuracy of 98.7% across multiple data types. For data classification, the algorithm attained accuracies of 96.2%, 95.6%, 96.1%, and 95.7% for categories A, B, C, and D, respectively, demonstrating strong classification capability. These results were comparable to the experimental findings of Abdi N. et al. [19], whose feature extraction accuracy was 92.1%, slightly lower than the method proposed in this study. The reduced accuracy in their approach can be attributed to limitations of the convolutional kernel's local receptive field, which constrained global information modeling and potentially caused the loss of early key features.

Further evaluation of the detection capability revealed that the proposed method maintained high detection accuracy when identifying abnormal information in voltage, current, and power data, with an average detection error of only 2.3%. In comparison, Fan Y. et al. [20] reported an error rate of 5.8% using the LSYM model, a performance gap likely due to the weaker feature extraction ability of their approach.The proposed method also demonstrated strong privacy protection. Results indicated a 92.1% reduction in privacy data leakage, while the SG successfully resisted more than 90% of external physical attacks. Compared to HE-LSTM and FedAvg privacy-preserving methods, the proposed approach exhibited significantly stronger resilience to external threats. A deeper analysis revealed that the LSTM component effectively captured long-term dependencies in PG data through its gate mechanisms, while K-means clustering enhanced feature extraction

by leveraging local similarity and global temporal patterns. Within the K-means-LSTM-FL framework, each PG node trained the model locally and shared only parameters rather than raw data, thereby reducing the risk of leakage during transmission and storage. Moreover, K-means clustering isolated abnormal data to minimize attackers' influence on training, LSTM extracted temporal features to detect hidden attack patterns, and FL preserved privacy by avoiding centralized data aggregation. By contrast, HE-LSTM and FedAvg approaches primarily addressed either privacy or data distribution alone, making them less effective against composite attacks.

4 Conclusion

To address the challenges of weak privacy protection and poor attack resistance in SGs, this study proposed an anomaly detection method that integrates K-means clustering, LSTM networks, and FL. Experimental results confirmed that the proposed method achieved high feature extraction and classification accuracy, with low detection error, enabling effective detection of abnormal information. When applied to SGs, the method substantially reduced privacy leakage and significantly enhanced resilience against external physical attacks.The findings demonstrate that the proposed K-means-LSTM-FL anomaly detection method enables real-time monitoring of PGs while ensuring data security. However, in this study, when using the K-means LSTM-FL algorithm to detect smart grids, the presence of a large amount of noise, outliers, and missing values in the smart grid may have a certain impact on the performance of the K-means LSTM-FL algorithm. Therefore, in the future, median filtering and linear interpolation can be used in data preprocessing to further process the data and reduce its impact on the K-means LSTM-FL algorithm.

Acknowledgements

This research project was supported by the Science and Technology Project of China Southern Power Grid Co., Ltd. (Project No.: 030000KC23120108 (GDKJXM20231533).

References

[1] Takiddin A, Ismail M, Atat R, Davis KR, Serpedin E. Robust graph autoencoder-based detection of false data injection attacks against data

poisoning in smart grids. IEEE Transactions on Artificial Intelligence. 2023, 5(3): 1287–1301. DOI: 10.1109/TAI.2023.3286831.

[2] Wu L, Fu S, Luo Y, Yan H, Shi H, Xu M. A robust and lightweight privacy-preserving data aggregation scheme for smart grid. IEEE Transactions on Dependable and Secure Computing. 2023, 21(1): 270–283. DOI: 10.1109/TDSC.2023.3252593.

[3] Vahidi S, Ghafouri M, Au M, Kassouf M, Mohammadi A, Debbabi M. Security of wide-area monitoring, protection, and control (WAMPAC) systems of the smart grid: A survey on challenges and opportunities. IEEE Communications Surveys & Tutorials. 2023, 25(2): 1294–1335. DOI: 10.1109/COMST.2023.3251899.

[4] An D, Zhang F, Cui F, Yang Q. Toward data integrity attacks against distributed dynamic state estimation in smart grid. IEEE Transactions on Automation Science and Engineering. 2023 Jan 17; 21(1): 881–194. DOI: 10.1109/TASE.2023.3236102

[5] Jithish J, Alangot B, Mahalingam N, Yeo KS. Distributed anomaly detection in smart grids: a federated learning-based approach. IEEE Access. 2023, 11(6): 7157–7179. DOI: 10.1109/ACCESS.2023.3237554.

[6] Wang Z, Jiang W, Xu J, Xu Z, Zhou A, Xu M. Grid2Vec: learning node representations of digital power systems for anomaly detection. IEEE Transactions on Smart Grid. 2024, 15(5): 5031–5042. DOI: 10.1109/TSG.2024.3377223.

[7] Guha D, Chatterjee R, Sikdar B. Anomaly detection using lstm-based variational autoencoder in unsupervised data in power grid. IEEE Systems Journal. 2023, 17(3): 4313–4323. DOI:10.1109/jsyst.2023.3266554.

[8] Takiddin A, Atat R, Ismail M, Boyaci O, Davis KR, Serpedin E. Generalized graph neural network-based detection of false data injection attacks in smart grids. IEEE Transactions on Emerging Topics in Computational Intelligence. 2023, 7(3): 618–630. DOI: 10.1109/TETCI.2022.3232821.

[9] Hu P, Gao W, Li Y, Guo X, Hua F, Qiao L. Anomaly detection and state correction in smart grid using EKF and data compensation techniques. IEEE Sensors Journal. 2024, 24(8): 12995–13009. DOI: 10.1109/JSEN.2024.3372973.

[10] Fu M, Shi Y, Zhou Y. Federated learning via unmanned aerial vehicle. IEEE Transactions on Wireless Communications. 2023, 23(4): 2884–2900. DOI: 10.1109/TWC.2023.3303492.

[11] Rykov A, De Amorim RC, Makarenkov V, Mirkin B. Inertia-based indices to determine the number of clusters in K-Means: an

experimental evaluation. IEEE Access. 2024, 12(5): 11761–11773. DOI: 10.1109/ACCESS.2024.3350791.

[12] Helmy I, Tarafder P, Choi W. LSTM-GRU model-based channel prediction for one-bit massive MIMO system. IEEE Transactions on Vehicular Technology. 2023, 72(8): 11053–11057. DOI: 10.1109/TVT.2023.3262951.

[13] Zhou X, Liang W, Kevin I, Wang K, Yan Z, Yang L T, Jin Q. Decentralized P2P federated learning for privacy-preserving and resilient mobile robotic systems. IEEE Wireless Communications. 2023, 30(2): 82–89. DOI: 10.1109/MWC.004.2200381.

[14] Liu H, Chen J, Dy J, Fu Y. Transforming complex problems into K-means solutions. IEEE transactions on pattern analysis and machine intelligence. 2023, 45(7): 9149–9168. DOI: 10.1109/TPAMI.2023.3237667.

[15] Zhang X, Chau TK, Chow YH, Fernando T, Iu HH. A novel sequence to sequence data modelling based CNN-LSTM algorithm for three years ahead monthly peak load forecasting. IEEE Transactions on Power Systems. 2023, 39(1): 1932–1947. DOI: 10.1109/TPWRS.2023.3271325.

[16] Yu P, Huang W, Zhang R, Qian X, Li H, Chen H. GuardGrid: A Queriable and Privacy-Preserving Aggregation Scheme for Smart Grid via Function Encryption. IEEE Internet of Things Journal. 2025, 12(11): 17622–17633. DOI: 10.1109/JIOT.2025.3539724.

[17] Sarieddine K, Sayed MA, Jafarigiv D, Atallah R, Debbabi M, Assi C. A real-time cosimulation testbed for electric vehicle charging and smart grid security. IEEE Security & Privacy. 2023, 21(4): 74–83. DOI: 10.1109/MSEC.2023.3247374.

[18] Liu Y, Kang Y, Zou T, Pu Y, He Y, Ye X, Ouyang Y, Zhang YQ, Yang Q. Vertical federated learning: Concepts, advances, and challenges. IEEE Transactions on Knowledge and Data Engineering. 2024, 36(7): 3615–3634. DOI: 10.1109/TKDE.2024.3352628.

[19] Abdi N, Albaseer A, Abdallah M. The role of deep learning in advancing proactive cybersecurity measures for smart grid networks: A survey. IEEE Internet of Things Journal. 2024, 11(9): 16398–16421. DOI: 10.1109/JIOT.2024.3354045.

[20] Fan Y, Liu J, Ye H, Lyu Z. TA-LSTM: a time and attribute aware LSTM for deep flight track clustering. IEEE Transactions on Aerospace and Electronic Systems. 2023, 59(5): 7047–7060. DOI: 10.1109/TAES.2023.3285203.

Biographies

Xiaozhi Deng (October 1985–), male, graduated from South China University of Technology with a master's degree in information engineering. After graduation, I worked as a senior engineer in the Power Dispatching and Control Center of Guangdong Power Grid Co., Ltd. My current research direction is to engage in network security work in power distribution monitoring systems.

Yaoxin Pan (September 1996–), male, graduated with a master's degree in computer technology from Sun Yat-sen University. After graduation, I worked as an engineer at the Power Dispatching and Control Center of Guangdong Power Grid Co., Ltd. My current research area is network security for power monitoring systems.

Hongjia Fang (October 1994–), male, graduated with a bachelor's degree from South China Normal University. After graduation, I worked as an engineer at China Southern Power Grid Digital Enterprise Technology (Guangdong) Co., Ltd. My current research area is network security for power monitoring systems.