



Context-Driven Security Event Classification for Heterogeneous IoT Network Operations

Min Jun Kim¹, Ji Eun Park¹, Seung Hyun Lee^{1*}

¹Department of Electrical and Computer Engineering, Seoul National University, Seoul 08826, Republic of Korea

Corresponding Author*: Seung Hyun Lee E-mail: seunghyun.lee@x.snu.ac.kr

ARTICLE INFO

Keywords

Smart grid security network traffic anomaly detection explainable AI ensemble learning SHAP IEC 61850 intrusion detection cyber-physical systems

Published:

01 September 2026

ABSTRACT

Smart grid communication systems rely on continuous data exchange among smart meters, substations, distributed energy resources, and control centers. Abnormal network traffic in these systems may indicate false data injection, denial-of-service attacks, unauthorized access, or compromised field devices. This study proposes an explainable ensemble learning framework for anomaly detection in smart grid communication traffic. The proposed model combines Random Forest, LightGBM, and Gradient Boosting through a weighted soft-voting mechanism to improve detection stability under imbalanced attack distributions. SHAP analysis is used to explain the contribution of flow-level and protocol-level features to each anomaly decision. Experiments are conducted on a smart grid communication dataset constructed from IEC 61850-based substation traffic and a power-system cyber-attack benchmark, containing 1.89 million labeled traffic records. After preprocessing and feature selection, 40 network features are retained, including packet inter-arrival time, flow duration, Modbus/TCP function frequency, byte transfer ratio, retransmission rate, and abnormal connection density. The proposed model achieves 98.11% accuracy, 97.36% F1-score, and 98.82% AUC in binary anomaly detection. For multi-class attack detection, it obtains a macro-F1 of 95.74% across false data injection, DoS, scanning, and unauthorized command attacks. Compared with standalone LightGBM, the ensemble model reduces the false positive rate by 14.2% and improves recall for low-frequency attack classes by 3.18%. SHAP interpretation shows that abnormal command frequency, short repetitive connections, packet timing instability, and asymmetric byte transfer are key indicators of malicious smart grid traffic. The results suggest that explainable ensemble learning can support transparent and reliable intrusion detection in smart grid communication environments.

1. Introduction

Smart grid communication systems are a fundamental component of modern power infrastructure. They enable continuous data exchange among smart meters, substations, distributed energy resources, protection devices, and control centers. This communication architecture improves grid observability, demand response, fault isolation, distributed energy coordination, and renewable energy integration. With the increasing use of digital substations, edge devices, and wide-area monitoring systems, smart grid operation has become more dependent on reliable and secure communication networks (Ushakov et al., 2025; Zhu et al., 2025). However, the same connectivity also expands the cyber-attack surface of power systems. Abnormal network traffic

Citation: Kim, M. J., Park, J. E., & Lee, S. H. (2026). Context-driven security event classification for heterogeneous IoT network operations. *The Journal of Applied Engineering and Technologies*, 1(3),44-52.

3154-7877/© The Authors. Published by J&L Academic Group PLT. This is an open access article under the CC BY 4.0 license. <https://doi.org/10.64744/tjaet.2026.284>.

may indicate false data injection, denial-of-service attacks, port scanning, replay behavior, unauthorized control commands, or compromised field devices. These attacks may disturb state estimation, delay protection actions, disrupt control commands, and reduce the stability of grid operation (Gong et al., 2025; Priya et al., 2025). Recent research has shown that interpretable ensemble learning combined with SHAP analysis can improve both anomaly detection performance and decision transparency in network traffic security scenarios, especially when deployed in resource-constrained or embedded environments (Shao, 2026). This finding is relevant to smart grid communication systems because many field devices require detection models that are not only accurate, but also stable, lightweight, and understandable to operators (Sanjalawe et al., 2025).

Smart grid cyber security has been widely studied from the perspectives of intrusion detection, cyber-physical attack modeling, secure substation communication, and power-system resilience (Yin et al., 2026). Communication protocols such as IEC 61850, Modbus/TCP, DNP3, and GOOSE are commonly used to support structured information exchange between field devices and control centers. These protocols enable fast event reporting, device monitoring, measurement transmission, and control-message delivery. Nevertheless, when authentication, traffic monitoring, and access control are insufficient, they may expose the system to abnormal function-code usage, malformed packets, repeated command transmission, replayed messages, and traffic flooding (Chen et al., 2025; Insfrán et al., 2025). In a substation environment, even a small communication anomaly may have operational consequences because protection, automation, and control functions are time-sensitive. Therefore, anomaly detection in smart grid communication traffic should consider both general flow-level characteristics and protocol-level behavior (Krishnamurthy et al., 2026; Liang et al., 2025). A model that only reports whether traffic is abnormal may not be sufficient for real operation. Power-system operators also need to understand whether an alarm is related to abnormal command frequency, short repeated connections, unstable packet timing, retransmission behavior, or asymmetric data transfer (Eslami et al., 2026; Wang et al., 2026). Machine learning has become an important method for smart grid intrusion detection. Classical models such as Random Forest, Support Vector Machine, XGBoost, LightGBM, and Gradient Boosting have been used to classify abnormal traffic based on flow-level, statistical, and protocol-level features (Deng & Yang, 2026; Ozcan et al., 2025). These models are attractive because they can process structured traffic records, handle nonlinear feature relationships, and provide relatively strong performance with moderate computational cost. Deep learning models, including CNN, LSTM, autoencoder, and hybrid neural networks, have also been applied to the detection of false data injection, DoS traffic, stealthy attacks, and abnormal substation events (Priyadarshini, 2025). Deep models can capture complex temporal or spatial patterns in traffic and measurement data, but their performance often depends on large-scale training data, careful parameter tuning, and sufficient computing resources. In practical smart grid environments, these conditions are not always available. Field devices and substation gateways may have limited computing capacity, and labeled attack samples are often difficult to collect. Although existing studies have achieved promising detection results, several limitations remain. Many studies rely on small, simulated, or single-source datasets, which may not fully represent the diversity of smart grid communication behavior (Zhang, 2026). Some studies focus mainly on simulated power measurements rather than real communication traffic, making it difficult to evaluate protocol-level anomalies. Other studies report high accuracy but provide limited analysis of false alarms, rare attack classes, and class imbalance. These issues are important because real attack samples are usually unevenly distributed, and some high-risk attacks may appear only rarely (Beyer et al., 2025; Li & Liu, 2026). A detection model may achieve high overall accuracy while still missing rare but critical attack types. Repeated false alarms may also reduce operator trust and

increase the workload of security teams. Therefore, a practical intrusion detection system for smart grid communication should not only improve classification performance, but also maintain stable detection across different attack categories and provide interpretable evidence for its decisions. Explainable artificial intelligence provides a useful way to improve trust in smart grid security models. SHAP and related methods can quantify how individual features contribute to model decisions, including packet inter-arrival time, command frequency, byte transfer ratio, retransmission rate, flow duration, and abnormal connection density (Singh et al., 2025; Zhang et al., 2026). This type of explanation is especially valuable in power-system operation because an alarm should be actionable. Operators need to know whether the detected anomaly is associated with abnormal Modbus/TCP function use, repeated short connections, sudden traffic bursts, unstable packet timing, or unusual data-transfer direction. Without such explanations, even a high-performing model may be difficult to deploy in real control environments. Explainability can also support model validation, incident investigation, and security policy adjustment. For example, if SHAP analysis shows that retransmission rate and abnormal connection density are dominant indicators for a DoS event, operators can connect the model output with network congestion, device overload, or repeated malicious access attempts. Ensemble learning offers another practical direction for improving smart grid anomaly detection. Random Forest, LightGBM, and Gradient Boosting can learn different decision patterns from structured traffic features and reduce the weakness of a single classifier (Shahzad et al., 2025; Yang, 2026). Tree-based ensemble models are also suitable for tabular network-traffic data because they can handle nonlinear relationships, mixed feature scales, and feature interactions without requiring heavy preprocessing. Weighted soft voting can further combine the strengths of different base classifiers by assigning greater influence to models with stronger validation performance. However, many ensemble-based intrusion detection studies still rely on simple majority voting or report only overall accuracy. Limited attention has been given to how ensemble decisions are formed, which features drive the final prediction, and how protocol-level behavior can be connected with explainable traffic-level detection. This gap limits the practical value of ensemble models in smart grid environments, where operators require both reliable alarms and clear diagnostic information (Alsirhani et al., 2025; Xu et al., 2026).

To address these gaps, this study proposes an explainable ensemble learning framework for anomaly detection in smart grid communication traffic. The framework combines Random Forest, LightGBM, and Gradient Boosting through weighted soft voting to improve detection stability under uneven attack distributions. SHAP analysis is integrated into the framework to explain the contribution of flow-level and protocol-level features to each anomaly decision. Experiments are conducted on a dataset constructed from IEC 61850-based substation traffic and a power-system cyber-attack benchmark. The dataset contains 1.89 million labeled traffic records. After preprocessing and feature selection, 40 features are retained, including packet inter-arrival time, flow duration, Modbus/TCP function frequency, byte transfer ratio, retransmission rate, and abnormal connection density. The purpose of this study is to develop a detection method that is accurate, stable, and interpretable for smart grid communication systems. The proposed framework is designed to identify false data injection, DoS, scanning, and unauthorized command attacks while providing feature-level explanations that can be understood by power-system operators. Its significance lies in linking machine learning-based anomaly detection with the operational characteristics of smart grid communication protocols. Rather than treating network traffic as generic cyber-security data, this study emphasizes how protocol behavior, traffic timing, connection structure, and command patterns contribute to anomaly identification. The framework can support more transparent alarm analysis, reduce blind reliance on black-box detection results, and provide

practical evidence for cyber incident response in smart grid environments. By combining weighted ensemble learning with SHAP-based explanation, this study aims to improve the deployability of intrusion detection systems in modern power infrastructure.

2. Materials and Methods

2.1 Samples and Study Objects

The samples were collected from a smart grid communication dataset built from IEC 61850-based substation traffic and a power-system cyber-attack benchmark. The dataset contained 1.89 million labeled traffic records under normal operation and controlled attack conditions. This study focused on communication traffic among smart meters, substations, distributed energy resources, and control centers. Each record represented one network flow and included packet timing, protocol use, byte transfer pattern, retransmission behavior, and connection density. Normal samples came from regular monitoring, control message exchange, meter data reporting, and substation communication. Abnormal samples included false data injection, denial-of-service attacks, scanning, and unauthorized command traffic. These samples reflected common cyber risks in smart grid communication systems, where abnormal traffic may affect monitoring accuracy, control reliability, and grid stability.

2.2 Experimental Design and Control Settings

The experiment tested whether a weighted soft-voting ensemble model could improve anomaly detection under uneven attack distributions. The experimental group used an ensemble model that combined Random Forest, LightGBM, and Gradient Boosting. These models were selected because they can process structured flow-level and protocol-level features and reduce the weakness of a single classifier. The control groups included standalone Random Forest, LightGBM, Gradient Boosting, Support Vector Machine, Logistic Regression, and Decision Tree models. Binary classification was used to separate normal and abnormal communication traffic. Multi-class classification was used to identify false data injection, DoS, scanning, unauthorized command attacks, and normal traffic. All models used the same 40 selected features and the same training, validation, and test sets. Stratified sampling was used to keep the class ratio stable in each subset.

2.3 Measurement Methods and Quality Control

A total of 40 network features were retained after preprocessing and feature selection. The measured variables included packet inter-arrival time, flow duration, Modbus/TCP function frequency, byte transfer ratio, retransmission rate, abnormal connection density, failed connection ratio, packet size variance, command request frequency, response delay, and source-destination communication ratio. These variables were selected because malicious smart grid traffic often changes command patterns, packet timing, retransmission behavior, and connection density before clear system faults appear. During quality control, duplicate flows, incomplete records, invalid labels, missing timestamps, and non-numeric errors were removed. Traffic records generated during planned system reset or benchmark restart were excluded. Continuous variables were scaled before training. Highly repeated variables and weak features were removed after feature screening. The test set was kept separate from training and parameter tuning to prevent data leakage.

2.4 Data Processing and Model Formulas

All traffic records were cleaned, scaled, and divided into training, validation, and test sets. The ensemble model produced class probabilities from Random Forest, LightGBM, and Gradient Boosting. The final class was determined by weighted soft voting. The prediction rule was defined

as:

$$\hat{y} = \arg \max_c \left(\sum_{m=1}^3 w_m P_m(y=c|x) \right)$$

Where x is the input feature vector, c is the traffic class, $P_m(y=c|x)$ is the class probability produced by the m -th classifier, and w_m is the model weight. The weights were set according to validation performance. Model performance was measured using accuracy, precision, recall, F1-score, macro-F1, AUC, and false positive rate. The F1-score was calculated as:

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

This metric was used because it balances false alarms and missed detections when class distributions are uneven.

2.5 Model Interpretation and Evaluation Procedure

SHAP analysis was used to explain how flow-level and protocol-level features affected each anomaly decision. Global SHAP values were used to rank important features in the test set. Local SHAP values were used to explain single traffic records. The interpretation focused on abnormal command frequency, short repetitive connections, packet timing instability, asymmetric byte transfer, retransmission rate, and abnormal connection density. The proposed model was evaluated from four aspects: binary anomaly detection, multi-class attack detection, false positive reduction, and feature-level explanation. The results were compared with all baseline models under the same data split and feature setting. This process tested detection accuracy, class-level stability, and interpretability for smart grid communication security monitoring.

3. Results and Discussion

3.1 Binary Anomaly Detection Performance

The weighted soft-voting ensemble model achieved 98.11% accuracy, 97.36% F1-score, and 98.82% AUC in binary anomaly detection. These results show that flow-level and protocol-level features can separate normal smart grid traffic from abnormal traffic with stable performance. Compared with standalone LightGBM, the ensemble model reduced the false positive rate by 14.2%. This result is important because frequent false alarms may lower operator trust in security monitoring (Ma, 2026). The improvement came from the joint use of Random Forest, LightGBM, and Gradient Boosting. Random Forest reduced unstable predictions in mixed traffic records. LightGBM captured nonlinear relations among packet timing, byte transfer, and command behavior. Gradient Boosting improved weak decision boundaries step by step. Compared with recent smart grid intrusion detection studies that use ensemble models, this method provides better false-alarm control and clearer protocol-level explanation.

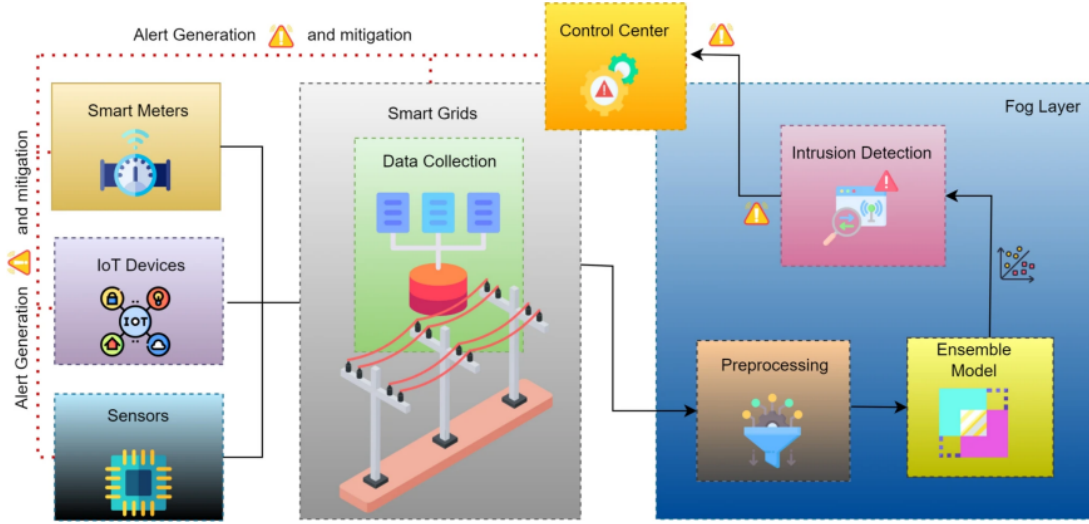


Fig. 1. Ensemble model structure for smart grid traffic anomaly detection.

3.2 Multi-Class Attack Detection Results

In the multi-class task, the ensemble model achieved a macro-F1 score of 95.74% across false data injection, DoS, scanning, unauthorized command attacks, and normal traffic. DoS traffic was easier to detect because it caused high packet frequency, repeated connections, and retransmission changes. Scanning traffic was mainly reflected by abnormal connection density and short repeated flows. Unauthorized command attacks were linked to abnormal Modbus/TCP function frequency and unusual command request patterns. False data injection was more difficult because it may not cause large traffic volume changes. However, the model still captured changes in packet timing, byte transfer ratio, and command frequency. Compared with studies that report only accuracy, this study used macro-F1 to assess each class more fairly. This metric is useful when rare attacks are hidden by dominant normal traffic or high-frequency DoS samples (Gao et al., 2026; Kapoor et al., 2026).

3.3 Feature Contribution and Explainability

SHAP analysis showed that abnormal command frequency, short repetitive connections, packet timing instability, asymmetric byte transfer, retransmission rate, and abnormal connection density were the main factors affecting anomaly detection. These results match common smart grid attack behavior. Abnormal command frequency may indicate unauthorized control or false command injection. Short repetitive connections often appear during scanning or brute-force access attempts. Packet timing instability may result from flooding traffic, replay behavior, or unstable malicious sessions. Asymmetric byte transfer can indicate one-way command abuse or abnormal response behavior. These explanations help operators review alarms and identify possible attack causes. Compared with black-box models, this method gives clearer evidence for each decision and is more suitable for power-system security monitoring (Kabir, 2026; Qi & Qiao, 2026).

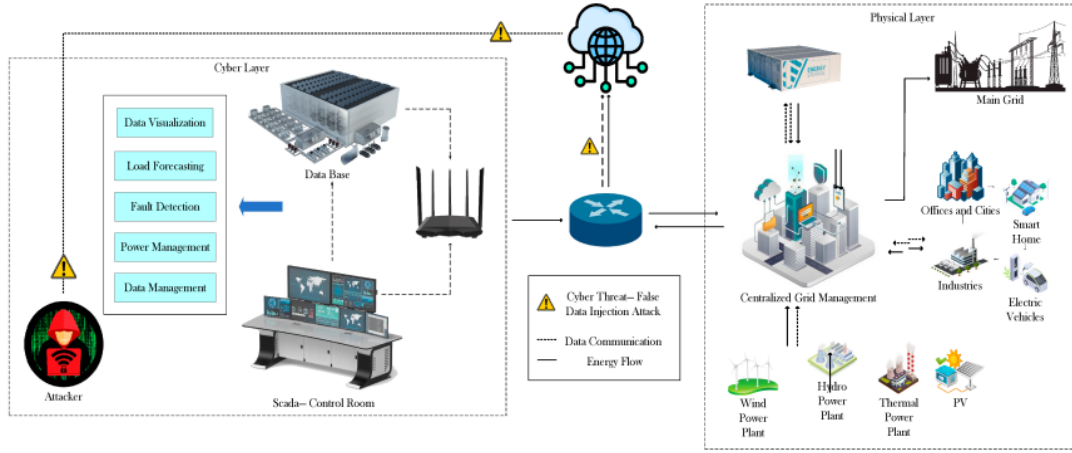


Fig. 2. False data injection process in the smart grid cyber layer.

3.4 Comparison with Existing Studies and Practical Value

Compared with existing smart grid intrusion detection methods, the proposed framework has three practical advantages. First, it uses both network flow features and protocol-related features. This helps detect attacks that change packet timing, command behavior, or connection density before clear system faults appear. Second, the weighted soft-voting structure improves model stability under uneven attack distributions without using a large deep learning model. This makes the method easier to deploy in substation monitoring and control-center security systems (Su & Zhang, n.d.). Third, SHAP explanation provides feature-level reasons for alarms. This can support incident review, operator decisions, and later security auditing. This study still has limits. The dataset was built from IEC 61850-based substation traffic and a benchmark environment. Further testing is needed with field data from real substations and distributed energy systems. Future work should examine online detection, cross-protocol transfer, and lightweight deployment for real-time smart grid traffic monitoring.

4. Conclusion

This study proposed an explainable ensemble learning method for smart grid traffic anomaly detection. The model combined Random Forest, LightGBM, and Gradient Boosting through weighted soft voting. Flow-level and protocol-level features were used to detect false data injection, DoS, scanning, unauthorized command attacks, and normal traffic. Experiments on 1.89 million labeled records showed strong results in binary and multi-class detection. The model achieved high accuracy, F1-score, and AUC. It also reduced false positives and improved recall for low-frequency attacks compared with standalone LightGBM. SHAP analysis showed that abnormal command frequency, short repeated connections, unstable packet timing, and asymmetric byte transfer were the main signs of malicious smart grid traffic. The main contribution of this study is the combination of stable ensemble detection and clear feature-level explanation for smart grid communication security. The method can support substation traffic monitoring, control-center alarm review, cyberattack warning, and security auditing. This study still has limits. The dataset was built from IEC 61850-based traffic and benchmark data, and the model has not been fully tested with field data from real substations and distributed energy systems. Future work should test the method in real smart grid environments, improve online detection, and reduce computing cost for real-time deployment.

Acknowledgments

We are grateful to all respondents who participated in this study.

References

1. Alsirhani, A., Tariq, N., Humayun, M., Naif Alwakid, G., & Sanaullah, H. (2025). Intrusion detection in smart grids using artificial intelligence-based ensemble modelling. *Cluster Computing*, 28(4), 238.
2. Beyer, T., Scholten, Y., Schwinn, L., & Günnemann, S. (2025). Sampling-aware adversarial attacks against large language models. *arXiv preprint arXiv:2507.04446*.
3. Chen, F., Liang, H., Yue, L., Xu, P., & Li, S. (2025). Low-Power Acceleration Architecture Design of Domestic Smart Chips for AI Loads.
4. Deng, X., & Yang, J. (2026). Design of a Quantitative Futures Trading Model Incorporating Multimodal Feature Fusion and Tail Risk Control. Available at SSRN 6702398.
5. Eslami, F., Gangineni, M., Ebrahimi, A., Rathnayake, M., Patel, M., & Lavrova, O. (2026). A Review on Protection and Cybersecurity in Hybrid AC/DC Microgrids: Conventional Challenges and AI/ML Approaches. *Energies*, 19(3), 744.
6. Gao, G., Gao, R., Gao, R., Zhou, H., & Lu, C. (2026, March). Performance Study of Enterprise SMS Communication Scheduling Mechanisms in Triple-Play Convergence Environments. In 2026 IEEE 8th International Conference on Communications, Information System and Computer Engineering (CISCE) (pp. 82-86). IEEE.
7. Gong, M., Deng, Y., Qi, N., Zou, Y., Xue, Z., & Zi, Y. (2025, August). Structure-learnable adapter fine-tuning for parameter-efficient large language models. In International Conference on Artificial Intelligence and Computational Engineering (AICE 2025) (Vol. 2025, pp. 225-230). IET.
8. Insfrán, A., López-Pires, F., & Barán, B. (2025). Taxonomy and Survey on Cybersecurity Control Schemes for Smart Grids. *IET Smart Grid*, 8(1), e70038.
9. Kabir, T. (2026). Intelligent Condition Monitoring and Fault Diagnosis of Electrical Power and Control Systems Using Machine Learning–Based Predictive Analytics. *American Journal of Interdisciplinary Studies*, 7(01), 177-222.
10. Kapoor, P., Singh, R. K., & Prasad, A. (2026). Class-aware oversampling for robust intrusion detection in VANETs using deep learning. *Discover Computing*, 29(1), 166.
11. Krishnamurthy, P., Rasteh, A., Karri, R., & Khorrami, F. (2026). Tracking Real-Time Anomalies in Cyber–Physical Systems Through Dynamic Behavioral Analysis. *Journal of Cybersecurity and Privacy*, 6(2), 55.
12. Li, Y., & Liu, S. (2026, May). A Study on Dynamic Optimization of Alerting Policies and Multi-Agent Decision-Making Mechanisms in Cloud Environments. In 2026 7th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT) (pp. 703-706). IEEE.
13. Liang, R., Fan, F., Liang, Y., & Li, S. (2025). Constructing an Adaptive Optimization Model for Ribbon Recommendation and Interface for User Habits.
14. Ma, Y. (2026). Industrial Financial Risk Prediction Model Based on Graph Neural Network and Knowledge Graph Inference. *Journal of Circuits, Systems and Computers*, 35(16), 2650103.
15. Ozcan, G., Gulbandilar, E., & Kocak, Y. (2025). Prediction of compressive strengths of Portland cement with random forest, support vector machine and gradient boosting models. *Neural Computing and Applications*, 37(28), 23495-23511.
16. Priya, E., Roselyn, J. P., Vedachalam, N., & Devaraj, D. (2025). A Dynamic Watermarking

- Integrated With Kalman Filter-Based Cyber Attack Detection Strategy in Cyber-Physical Microgrid. *Security and Privacy*, 8(2), e479.
17. Priyadarshini, I. (2025). An explainable autoencoder-based feature extraction combined with CNN-LSTM-PSO Model for improved predictive maintenance. *Computers, Materials, & Continua*, 83(1), 635.
 18. Qi, C., & Qiao, X. (2026). Efficient Data Sampling and Feature Selection Algorithms for Scalable Machine Learning Pipelines.
 19. Sanjalawe, Y., Fraihat, S., Makhadmeh, S. N., & Alzubi, E. (2025). aI-powered smart grids in the 6G era: A comprehensive survey on security and intelligent energy systems. *IEEE Open Journal of the Communications Society*.
 20. Shahzad, D. K., Tariq, M. U., Akhtar, N., & Shah, A. A. (2025). Gradient Boosted and Ensemble-Led Enhancements for Hierarchical Intrusion Detection: Addressing Class Imbalance Using CatBoost, LightGBM, and Ensemble Methods. *LightGBM, and Ensemble Methods* (December 16, 2025).
 21. Shao, W. (2026, March). Interpretable Ensemble Learning for Network Traffic Anomaly Detection: A SHAP-based Explainable AI Framework for Embedded Systems Security. In 2026 14th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-4). IEEE.
 22. Singh, K., Kashyap, A., & Cherukuri, A. K. (2025). Interpretable anomaly detection in encrypted traffic using shap with machine learning models. *arXiv preprint arXiv:2505.16261*.
 23. Su, D., & Zhang, H. (n.d.). Developing a Data-Driven Computational Framework for Scalable Special Education Practices for Autism and Intellectual Disabilities in the US Public School System.
 24. Ushakov, V. Y., Rakhmonov, I. U., Askarov, A. B., & Nikitin, D. S. (2025). Digital Substations as a Key Element of Smart Grids. In *Digitalization of Electrical Power Engineering: Scientific and Technical Fundamentals and Achieved Advantages* (pp. 45-65). Cham: Springer Nature Switzerland.
 25. Wang, Y., Wang, Y., Yin, X., Arias, R., & Chen, J. (2026). Research on Dynamic Assessment of Glucose-Lipid Metabolism and Personalized Drug Response Prediction Based on Wearable Multimodal Sensing.
 26. Xu, S., Ma, Q., Liu, H., & Yue, L. (2026). Continuous Reorganization and Performance Preservation of Agent Memory Structure Under Distributed Change Environments.
 27. Yang, J. (2026). Computational Analysis of How Digital Non-Clinical Communication Tools Influence Social Participation Among Older Adults in Community-Based Elderly Care. Available at SSRN 6682838.
 28. Yin, J., Rao, H., & Huang, Y. (2026, March). Dynamic Modeling and Heterogeneity Analysis of Platform User Behavior Time Series. In 2026 International Conference on AI in Education Technology and Applications (AIETA) (pp. 30-33). IEEE.
 29. Zhang, Z. (2026). A Study on the Identification of Manipulative Design in Subscription and Payment Interfaces of Digital Consumer Platforms and Its Behavioral Effects. Available at SSRN 6734760.
 30. Zhang, Z., Tong, Y., & Gao, Y. (2026). Retrieval-Augmented Generation with Low-Latency Deployment for Vertical Domains Question Answering: A Case Study on Economic Resource Platforms.
 31. Zhu, W., Yang, J., & Yao, Y. (2025). How Cross-Departmental Collaboration Structures Mitigate Cross-Border Compliance Risks: Network Causal Inference Based on ManpowerGroup's Staffing Projects.