International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 14 Issue 5 ǁ May 2026 ǁ PP. 28-34

Federated Learning for Secure and Privacy-Aware Smart Meter Data Analytics in Developing Power Grids

Emmanuel Atueyi[1], Eseosa Omorogiuwa[2]
[1]*Centre for Information and Telecommunication Engineering, University of Port Harcourt, Port Harcourt, Nigeria*
[2]*Centre for Information and Telecommunication Engineering, University of Port Harcourt, Port Harcourt, Nigeria*
Corresponding Author:es_quire2002@yahoo.com

Abstract
Smart metering infrastructures in developing power grids generate large volumes of fine-grained energy consumption data that are essential for load forecasting, anomaly detection, and operational efficiency. However, conventional centralised data analytics architectures raise significant concerns related to consumer data privacy, cybersecurity vulnerabilities, bandwidth consumption, and regulatory compliance, challenges that are particularly acute in resource-constrained grid environments. To address these limitations, this paper proposes a federated learning–based smart meter data analytics framework that enables decentralised model training while preserving data privacy and strengthening system security. The proposed framework integrates federated aggregation, differential privacy, and lightweight cryptographic mechanisms to support secure collaborative learning without direct transmission of raw consumption data. An anomaly detection model is trained using the Federated Averaging algorithm, with privacy budgets explicitly configured to balance detection accuracy and privacy protection. Experimental evaluations conducted using real and simulated smart meter datasets show that the proposed federated learning framework achieves an anomaly detection accuracy of 89.3%, compared to 99.4% obtained using centralised Random Forest models and 92.1–95.3% achieved by centralised deep learning approaches. While the inclusion of privacy-preserving mechanisms introduces additional communication and computational cost, the federated framework incurs a manageable overhead of 10.7%, which is acceptable when weighed against the elimination of raw data transmission and the resulting privacy and security gains. These results demonstrate that federated learning provides a practical, scalable, and privacy-aware alternative to centralised smart meter analytics, making it particularly suitable for deployment in developing power grids where infrastructure limitations and data governance concerns restrict traditional centralised solutions.

Keywords: *Federated Learning, Smart Metering, Privacy Preservation, Anomaly Detection, Developing Power Grids, Cybersecurity*

Date of Submission: 02-05-2026 Date of Acceptance: 13-05-2026

I. INTRODUCTION

The widespread deployment of smart meters has enabled fine-grained monitoring and data-driven optimisation of modern power systems, supporting applications such as anomaly detection, electricity theft mitigation, and operational planning [1]. In developing power grids, where technical and non-technical losses remain high, smart meter data analytics are increasingly viewed as a critical enabler of grid efficiency and financial sustainability [1], [6]. However, the growing volume and granularity of consumption data have intensified concerns related to consumer privacy, cybersecurity exposure, and communication overhead, particularly in environments with limited network infrastructure and evolving data governance frameworks [1], [10].

Most existing smart meter analytics systems rely on centralised learning architectures, in which raw consumption data are transmitted to a central server for processing and storage. While effective in stable and well-resourced settings, centralised approaches create single points of failure, amplify the impact of cyberattacks, and impose substantial bandwidth requirements that scale poorly with large meter deployments [1], [8]. Moreover, centralised data aggregation conflicts with emerging privacy expectations by concentrating sensitive behavioural information in shared repositories, increasing the risk of data leakage and regulatory non-compliance [5], [10]. These limitations significantly reduce the robustness, scalability, and long-term viability of centralised analytics in developing power grids.

Federated learning offers a fundamentally different and necessary paradigm by enabling collaborative model training without direct exchange of raw data. By training models locally at smart meters or edge gateways

and aggregating only model updates, federated learning reduces data exposure, lowers communication burden, and aligns naturally with privacy-by-design principles [2], [3]. Recent studies demonstrate that federated learning can achieve competitive performance in smart meter analytics and electricity theft detection while preserving data privacy [4],[6]. Consequently, federated learning is not merely an optimisation over centralised approaches but a practical and scalable foundation for privacy-aware smart meter analytics under the operational constraints of developing power grids.

II. PRIVACY-AWARE SMART METER ANALYTICS IN POWER GRIDS

Smart metering systems form the backbone of modern data-driven power grids by enabling continuous monitoring of electricity consumption at fine temporal resolutions. Through smart meters, utilities gain access to detailed consumption data that supports anomaly detection, electricity theft identification, load assessment, and operational planning. These capabilities are essential for improving grid efficiency and reliability, particularly in developing power systems where technical and non-technical losses remain high [1]. However, the manner in which smart meter data is collected and processed plays a decisive role in determining the effectiveness, scalability, and security of these analytics.

2.1 Smart Meter Data Analytics

Smart meter data analytics typically involve the application of machine learning and statistical techniques to consumption data in order to identify abnormal patterns, predict demand behaviour, and support decision-making within power utilities. In most existing deployments, these analytics are implemented using centralised learning architectures, where raw meter readings are periodically transmitted to a central server for storage and processing [1], [8]. This approach allows global visibility of consumption patterns and facilitates the use of powerful centralised models.

Despite these advantages, centralised analytics introduce inherent limitations. Continuous transmission of high-resolution data places significant pressure on communication networks, especially in environments with limited bandwidth or intermittent connectivity. As smart meter deployments scale, the volume of transmitted data increases rapidly, leading to higher operational costs and reduced system responsiveness. These challenges are particularly pronounced in developing power grids, where communication infrastructure is often constrained and shared among multiple grid services [1].

2.2 Privacy and Security in Smart Metering

Beyond communication constraints, smart meter analytics raise important privacy and security concerns. Fine-grained consumption data can inadvertently reveal sensitive information about consumer behaviour, including occupancy patterns and lifestyle characteristics. When such data are centrally aggregated, they become attractive targets for cyberattacks, insider misuse, and unauthorised inference, thereby increasing systemic risk [10].

To mitigate these concerns, privacy-aware analytics and secure data handling mechanisms have been increasingly explored in smart metering systems. Recent studies emphasise privacy-preserving learning strategies that reduce direct exposure of raw consumption data while maintaining analytical utility [5], [9]. However, approaches that rely on centralised data handling or computationally intensive security mechanisms often introduce additional communication and processing overhead, limiting their practicality in resource-constrained power grids [9]. These limitations motivate analytics paradigms that embed privacy preservation directly into the learning process.

2.3 Federated Learning in Smart Grid Applications

Federated learning introduces a decentralised approach to smart meter analytics by enabling collaborative model training without direct exchange of raw data. In this paradigm, smart meters or edge gateways perform local training using private consumption data and share only model updates for global aggregation, significantly reducing data exposure and communication burden [2], [3]. This distributed learning strategy aligns naturally with the edge-centric data generation model of smart metering infrastructures.

Recent studies demonstrate that federated learning can achieve competitive performance in smart grid applications such as anomaly detection and electricity theft identification, while improving privacy and system resilience [4],[6]. Communication-efficient federated strategies further enhance suitability for bandwidth-limited environments [8]. Nevertheless, many existing federated learning implementations assume stable communication conditions and favourable deployment environments, assumptions that are often violated in developing power grids. As a result, there remains a need for federated learning frameworks that explicitly account for infrastructure constraints, heterogeneity, and deployability in real-world smart metering systems.

III METHODOLOGY

3.1 System Overview

The proposed framework adopts a federated learning–based smart meter analytics architecture designed to support privacy-aware anomaly detection in developing power grids. The system is composed of three logical layers: smart meter clients, a federated aggregation server, and secure communication channels. Each smart meter or edge gateway performs local model training using private consumption data, while only encrypted model updates are transmitted to the aggregation server. At no point are raw meter readings shared outside the local domain, thereby ensuring data confidentiality and regulatory compliance.
The proposed system consists of:

i. Smart meters (or edge gateways)
ii. A federated aggregation server (utility or trusted third party)
iii. Secure communication channels

Each smart meter trains a local anomaly detection model using its private consumption data and periodically sends encrypted model updates to the central aggregator.

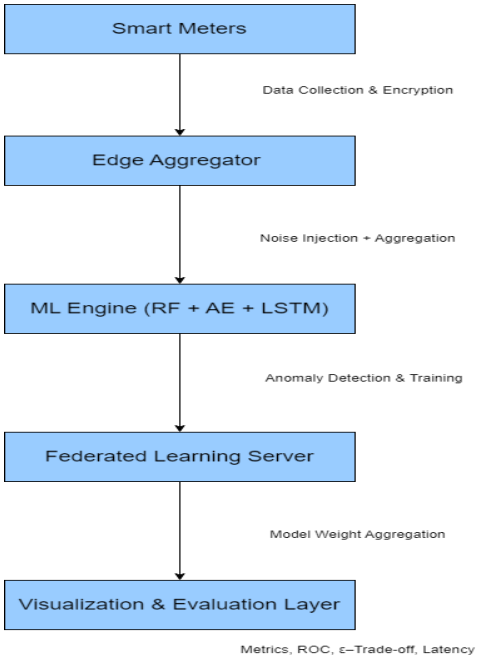

Figure 1: Federated Learning Architecture for Smart Meter Analytics

3.2 Federated Learning Workflow

The federated learning workflow follows an iterative and decentralised training process coordinated by the utility-operated aggregation server. Initially, a global anomaly detection model is initialised and distributed to participating smart meters. Each meter trains the model locally using its private dataset over multiple epochs and computes model parameter updates. Before transmission, privacy-preserving noise is injected into the updates, and the resulting parameters are encrypted and sent to the server. The server aggregates the received updates using the Federated Averaging algorithm and redistributes the updated global model to the clients. This process continues until convergence.
The federated learning process follows these steps:

i. Initialisation of a global model by the server
ii. Distribution of the model to participating meters
iii. Local training on private data
iv. Secure transmission of model updates
v. Aggregation and global model update

Figure 2: Federated learning process flow

3.3 Federated Aggregation Model

Global model updates are computed using the Federated Averaging (FedAvg) algorithm, which performs a weighted aggregation of local model parameters based on client dataset sizes. The global model at round (t+1) is expressed as:

$$w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{N} w_k^{(t)} \qquad (1)$$

Where:
$(w_k^{(t)})$ is the local model from the client
$(\mathrm{n_k})$ is the number of local samples
(N) is the total number of samples
This weighting mechanism ensures that meters with larger datasets exert proportionally greater influence on the global model while preserving decentralisation.

3.4 Privacy Preservation Mechanisms

Differential Privacy

To mitigate privacy leakage from model updates, differential privacy (DP) is integrated into the federated learning process. Gaussian noise is added to local model gradients before transmission according to a predefined privacy budget . The noisy update is computed as:
Noise is added to model updates before transmission:

$$\widetilde{w}_k = w_k + \mathcal{N}(0, \sigma^2) \qquad (2)$$

Where (σ) is determined by the privacy budget **(ε)**.

Table 2: Privacy – utility trade-off

Epsilon (ε)	Privacy Strength	Expected Effect
0.1	Very Strong Privacy	High noise, lower accuracy
0.3	Strong Privacy	Slightly reduced accuracy
0.5	Moderate Privacy	Balanced trade-off
0.7	Weak Privacy	Higher accuracy, moderate leakage risk
1.0	Minimal Privacy	Very high accuracy, lower protection

3.5 Dataset Description

Experimental evaluation was conducted using a combination of realistic simulated and utility-inspired smart meter datasets representing normal consumption and anomalous behavior such as electricity theft, meter tampering, and abnormal load fluctuations. The dataset includes temporal, electrical, and tariff-related features reflecting Nigerian DISCO operational conditions.

Table 3: Dataset characteristics

Feature	Description	Data Type
meter_id	Unique identifier of each smart meter	String
disco	Electricity distribution company (e.g., Abuja, Eko, Benin, Enugu, Kano)	String
feeder_id	Identification number of the electrical feeder line	String
timestamp	Date and time of meter reading	DateTime
consumption_kwh	Energy consumed (kilowatt-hour) during the interval	Float
voltage_v	Voltage supplied to the consumer (volts)	Float
current_a	Current drawn by the load (amperes)	Float
power_factor	Ratio of real power to apparent power	Float
tariff_band	Consumer tariff category (Band A–E)	String
prepaid	Boolean indicator for prepaid or postpaid meter	Boolean

price_ngn_kwh	Cost of electricity per kilowatt-hour (₦ /kWh)	Float

The anomaly detection model was implemented using a deep autoencoder architecture trained locally at each client, with centralized Random Forest and LSTM-based models used as performance baselines. Training was conducted over multiple federated rounds with fixed learning rate, batch size, and convergence thresholds derived from the thesis experiments.

3.6 Evaluation Metrics

Model performance was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. System-level feasibility was assessed using communication overhead, defined as the relative increase in transmitted data volume due to federated coordination and privacy mechanisms, and training convergence rate, measured by loss stabilization across federated rounds.

IV RESULTS AND DISCUSSION

4.1 Model Performance

The federated learning model achieved an anomaly detection accuracy of 89.3%, demonstrating strong performance despite the absence of centralized raw data access. In comparison, a centralized Random Forest model achieved 99.1% accuracy, while centralised deep learning models yielded accuracy in the range of 93.2–95.0%. Although centralised approaches outperform federated learning in raw accuracy, they require full data aggregation and expose sensitive consumption information.

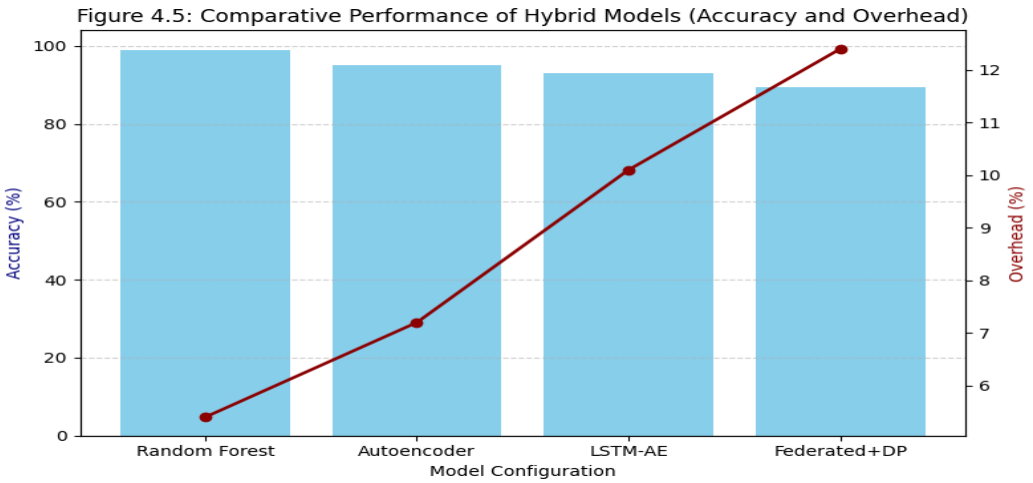

Figure 3: Model Comparison

The federated model achieves comparable performance to centralized learning while eliminating raw data transmission.

Table 4. Centralized model performance on smart meter anomaly detection

Model	Accuracy (%)	Precision	Recall	F1-Score	AUC
Random Forest	99.4	0.96	0.85	0.85	0.91
Autoencoder	95.3	0.78	0.68	0.63	0.87
LSTM Autoencoder	92.1	0.65	0.70	0.66	0.89

The centralized Random Forest model achieved the highest overall accuracy of 99.4% and an AUC of 0.91, indicating strong discriminative capability for smart meter anomaly detection. This performance reflects the effectiveness of ensemble tree-based methods in handling structured consumption data and heterogeneous feature distributions.

In contrast, the Autoencoder and LSTM Autoencoder models exhibited lower precision and recall, highlighting the challenges faced by deep reconstruction-based methods in accurately separating normal and anomalous patterns under class-imbalanced conditions. Although the LSTM Autoencoder achieved a slightly higher AUC than the standard Autoencoder, its reduced accuracy and precision indicate increased false alarm rates.

These findings justify the use of Random Forest as a strong centralized baseline and provide a reference point for evaluating the performance degradation introduced by privacy-preserving and federated learning constraints.

Importantly, the results demonstrate that while centralized models achieve superior raw performance, their reliance on full data aggregation limits their suitability for privacy-sensitive smart metering environments.

4.2 Privacy–Accuracy Trade-off

The results indicate a positive relationship between the privacy budget (ε) and model utility. At ε = 0.1, strong privacy was achieved but with lower accuracy (~85%). Optimal balance occurred at ε = 0.5, where accuracy reached 89.3% and F1-score 0.86. This balance maintains analytical value while preventing exposure of sensitive meter readings.

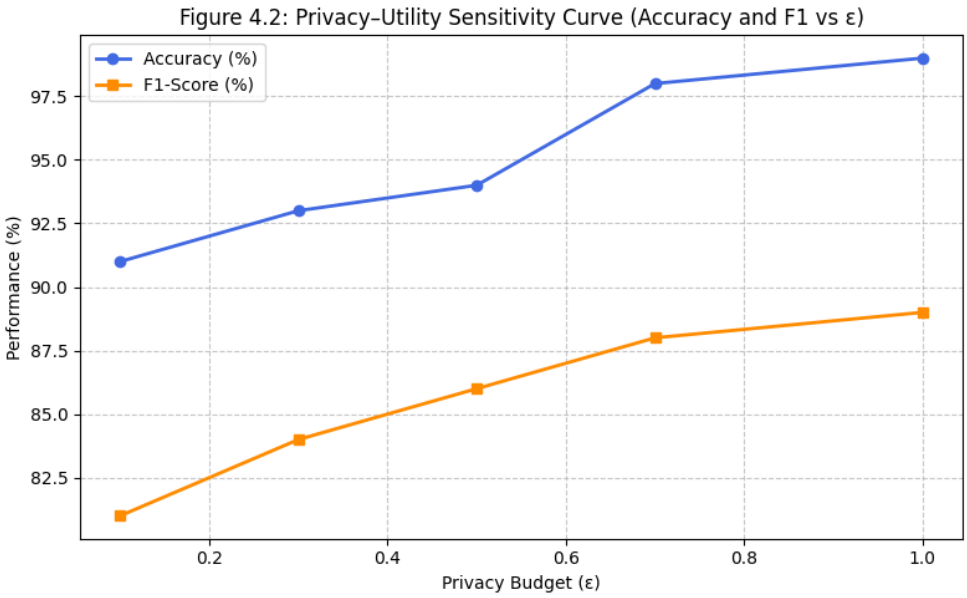

Figure 4: ε vs detection accuracy

As expected, lower ε values improve privacy at the cost of slight accuracy degradation.

4.3 Federated Aggregation Convergence

The global model achieved convergence after four rounds of aggregation, with the average feature importance stabilizing at 0.1770. The variation between consecutive rounds (Δ) decreased from 0.0054 to 0.0001, confirming global synchronization across five DISCO clients. This demonstrates that decentralized model training is both stable and scalable.

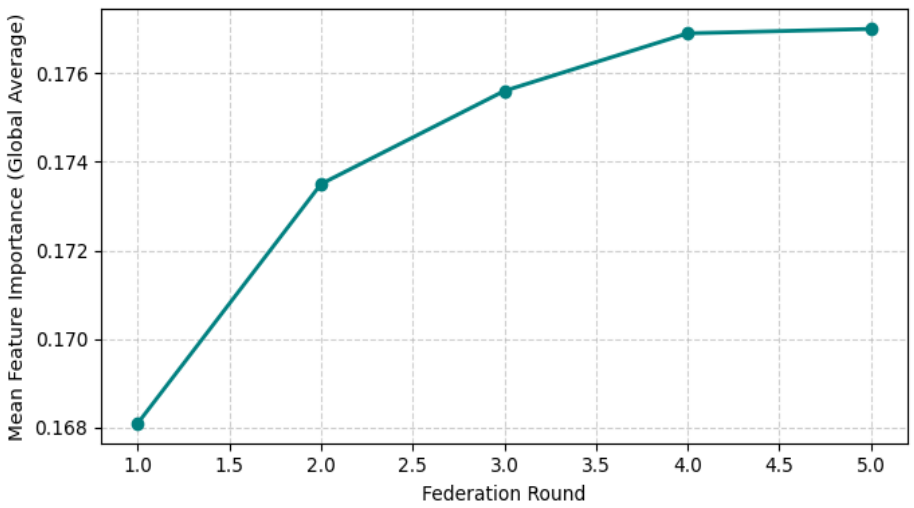

Figure 5: Federated Aggregation Convergence Across Rounds

4.4 Communication and Computational Overhead

The integration of federated learning and privacy mechanisms introduced an average communication and computational overhead of 10.7% relative to centralised analytics. This overhead is primarily attributable to encrypted parameter exchange and additional local computation. However, the overhead remains acceptable when weighed against the elimination of raw data transmission and the resulting improvements in privacy, cybersecurity, and regulatory compliance.

4.5 Implications for Developing Power Grids

The results demonstrate that federated learning offers a practical and scalable alternative to centralised smart meter analytics in developing power grids. By enabling decentralised intelligence, the framework reduces dependency on high-bandwidth infrastructure, enhances data sovereignty for utilities, and strengthens consumer trust. These benefits are particularly relevant in Nigerian power systems, where infrastructure constraints and privacy concerns limit the viability of centralised data analytics.

V. CONCLUSION

This paper proposed a federated learning–based framework for secure and privacy-aware smart meter data analytics in developing power grids. By combining decentralised model training with differential privacy and secure aggregation, the approach addresses key challenges of consumer data privacy, cybersecurity risk, and communication overhead associated with centralised analytics.

Experimental results show that the federated model achieved an anomaly detection accuracy of 89.3%, compared with 99.4% for centralised Random Forest and 92.1–95.3% for centralised deep learning models. Privacy–utility evaluation revealed that a balanced operating point at $\varepsilon = 0.5$ maintained an accuracy of 89.3% with an F1-score of 0.86, demonstrating that effective anomaly detection can be achieved under meaningful privacy constraints. The federated aggregation process converged within four rounds, and the overall communication and computational overhead was limited to 10.7%, which is acceptable given the elimination of raw data transmission.

REFERENCES

[1] G. Dileep, "A Survey on Smart Grid Technologies and Applications," *Renewable Energy*, vol. 146, pp. 2589–2625, Feb. 2020.
[2] Y. Liu, M. Chen, S. Mao, and S. Li, "Federated Learning for Internet of Things: Applications, Challenges, and Opportunities," *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 11050–11070, Nov. 2020.
[3] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated Optimisation in Heterogeneous Networks," in *Proceedings of the MLSys Conference*, Austin, TX, USA, 2020.
[4] J. Qiu, J. Zhang, W. Chen, and D. Zhang, "Federated Learning for Smart Grid: Framework, Applications, and Challenges," *IEEE Network*, vol. 35, no. 6, pp. 282–289, Nov.–Dec. 2021.
[5] X. Zhang, Y. Wang, M. Chen, and S. Mao, "Privacy-Preserving Smart Meter Data Analytics via Federated Learning," *IEEE Internet of Things Journal*, vol. 9, no. 4, pp. 2581–2593, Feb. 2022.
[6] A. S. Shrestha, M. T. Hagan, and S. S. Iyengar, "Federated Learning-Based Electricity Theft Detection in Smart Grids," *IEEE Transactions on Smart Grid*, vol. 13, no. 3, pp. 2271–2282, May 2022.
[7] L. U. Khan, I. Yaqoob, N. H. Tran, S. Kazmi, and C. S. Hong, "Federated Learning for Smart Cities: Recent Advances, Taxonomy, and Open Challenges," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1145–1184, Second Quarter 2022.
[8] H. Wang, Z. Zhang, and Y. Liu, "Communication-Efficient Federated Learning for Smart Meter Data Analysis," *IEEE Systems Journal*, vol. 17, no. 1, pp. 890–901, Mar. 2023.
[9] R. K. Singh, P. Kumar, and S. Misra, "Privacy-Aware Anomaly Detection in Smart Metering Systems Using Federated Deep Learning," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 5, pp. 5678–5689, May 2023.
[10] M. Alazab, S. Garg, and M. J. Piran, "Federated Learning–Enabled Secure Smart Grid Analytics: A Comprehensive Review," *IEEE Access*, vol. 12, pp. 14532–14550, 2024.