Hierarchical Federated Learning for Real-Time Anomaly Detection in Rural Electrical Grids: A Privacy-Preserving Edge–Fog–Cloud Architecture

Mekfoule Sidi Moctar El Bechir

Department of Computer Science, Master 2 AI Machine Learning & Data Science

Faculty of Science and Technology, University of Nouakchott Al Aasriya

Supervisor: Dr. El Benany Med Mahmoud

February 10, 2026

Abstract

The expansion of electrical infrastructure in rural Mauritania faces critical challenges: detecting grid anomalies in real time while adhering to data privacy regulations. Traditional centralized monitoring systems require raw sensor data migration to central servers, posing significant privacy risks and bandwidth constraints. This paper proposes a novel Data-to-Code architecture leveraging Edge Computing and Federated Learning for the SOMELEC (Société Mauritanienne d'Électricité) network. By deploying local training nodes at village substations and utilizing a Fog Computing layer (Apache Kafka) for asynchronous model aggregation, we achieve a global anomaly detection accuracy of 94.5% without a single row of private electrical consumption data leaving local premises. Experimental results on 288,000 sensor readings demonstrate 100% data privacy, a 99.9% bandwidth reduction, and real-time detection latency under 2 seconds.

Keywords: Federated Learning, Edge Computing, Anomaly Detection, Smart Grid, Apache Kafka, Distributed Systems, Privacy-Preserving AI, SOMELEC, Random Forest.

1 Introduction

Electrical grid monitoring has traditionally relied on consolidating sensor logs into centralized data warehouses. While effective for historical analysis, this *Code-to-Data* paradigm introduces privacy violations and unacceptable latency for real-time fault detection. In the context of Mauritanian electrical infrastructure, limited rural bandwidth and strict data privacy requirements necessitate a paradigm shift.

We propose a distributed framework where learning is pushed to the Edge. Instead of sharing sensitive household consumption data (voltage, current, power), village substations share only mathematical knowledge in the form of model parameters. This paper validates a three-layer Edge–Fog–Cloud architecture capable of aggregating intelligence from isolated villages into a robust national anomaly detection model.

1.1 Problem Context

The SOMELEC network faces the following challenges:

- Frequent undetected failures (MTTD $>$ 4 hours)
- High maintenance costs (50,000–100,000 MRU per intervention)
- Technical losses ($\sim$25% of generated power)
- Privacy concerns regarding household consumption patterns
- Limited rural bandwidth availability

1.2 Our Contribution

This research contributes:

1. Zero raw data transmission (privacy preservation)
2. 99.9% bandwidth reduction
3. Real-time detection ($<$ 2 seconds)
4. Fault-tolerant distributed design
5. Hierarchical optimization architecture

2 Methodology

2.1 Architectural Design

The system follows a hierarchical Edge–Fog–Cloud design. Edge nodes perform local training and detection, Fog nodes aggregate regional models using Apache Kafka, and the Cloud performs global federated averaging.

2.2 Data Collection and Simulation

Due to the lack of public SOMELEC datasets, a realistic simulator was developed using:

- World Bank electricity consumption data for Mauritania
- IEEE Std 1159-2019 voltage and current ranges
- SOMELEC technical specifications

Table 1: Electrical State Distributions

State	Voltage (V)	Current (A)	Probability
Normal	$N(220, 3)$	$N(15, 1.5)$	85%
Overvoltage	$N(260, 10)$	$N(15, 1.5)$	5%
Undervoltage	$N(180, 10)$	$N(15, 1.5)$	4%
Overload	$N(220, 3)$	$N(35, 5)$	3%
Failure	$N(150, 15)$	$N(5, 2)$	3%

2.3 Local Model Training

A Random Forest classifier was selected for robustness, interpretability, and low computational cost. Feature engineering includes voltage, current, power, and normalized ratios. Only model parameters are shared, ensuring full privacy.

2.4 Federated Averaging

The global model is updated using Federated Averaging:

$$w^{(t+1)} = \sum_{k=1}^{K} \frac{n_k}{n} w_k^{(t+1)}$$

where n_k is the number of samples at village k.

3 Experimental Setup

Experiments were conducted using Docker containers simulating two Edge villages, one Fog server, Kafka broker, and Cloud server. Approximately 288,000 sensor readings were generated per day.

4 Results

4.1 Global Accuracy

The federated model converged rapidly, achieving 94.5% accuracy after 10 rounds.

Table 2: Global Model Accuracy

Round	Accuracy	Villages
1	80.0%	2
5	91.5%	2
10	94.5%	2

4.2 Latency and Bandwidth

Edge-based detection achieved <2 seconds latency. Bandwidth usage was reduced from 43 MB/day to 48 KB/day per village.

5 Discussion

The results demonstrate that federated learning can preserve privacy while maintaining high accuracy and real-time responsiveness. Apache Kafka provided resilience and fault tolerance essential for critical infrastructure.

6 Conclusion

This work demonstrates a privacy-preserving, fault-tolerant, and real-time anomaly detection system for rural electrical grids. The architecture achieves 94.5% accuracy, 99.9% bandwidth reduction, and sub-2-second detection latency, making it suitable for national deployment by SOMELEC.

References

[1] World Bank, *Electric Power Consumption (kWh per capita) – Mauritania*, 2023.

[2] McMahan et al., “Communication-Efficient Learning of Deep Networks from Decentralized Data,” AISTATS, 2017.

[3] Apache Software Foundation, *Apache Kafka Documentation*, 2024.

[4] IEEE Standards Association, IEEE Std 1159-2019.

[5] Breiman, L., “Random Forests,” *Machine Learning*, 2001.