International Conference on Intelligent Computing and Explainable AI (ICICEA) – April 2026

RESEARCH ARTICLE OPEN ACCESS

Explainable Federated Learning Framework for Privacy-Preserving Smart Grid Energy Prediction

Tharani V **, Madhumitha A**,

**UG Students- Computer Science and Engineering, MRK Institute of Technology, Cuddalore, Tamilnadu

Abstract

Smart grid environments continuously generate detailed energy consumption data through smart meters and connected IoT devices. While conventional centralized machine learning approaches can achieve strong prediction performance, they require transferring user data to a central server, which raises privacy concerns. In this work, we present an Explainable Federated Learning (XFL) framework that performs distributed model training while keeping user data local. An LSTM-based model is trained across multiple clients using federated averaging, and SHAP is applied to interpret model predictions. Experiments conducted on a smart meter dataset show that the proposed approach achieves 94.3% prediction accuracy while maintaining data privacy and improving model transparency .

Keywords: Federated Learning,Explaoinable AI,Smart Grid,LSTM, Privacy Preservation

I. Introduction

Smart grid systems combine traditional power infrastructure with digital technologies to enable efficient monitoring and control of energy usage. With the deployment of smart meters, large volumes of time-series data are generated, making predictive analytics an important component of energy management.

However, many existing machine learning solutions rely on centralized data collection, which can expose sensitive user information. Federated Learning (FL) provides an alternative by allowing models to be trained locally on user devices while sharing only model parameters.

Although FL improves privacy, it introduces a new challenge: lack of interpretability. Understanding why a model makes certain predictions is critical for real-world deployment. To address this, we integrate Explainable AI techniques into the federated learning pipeline.

Contributions of this work:

1. Design of a privacy-preserving federated learning framework for energy prediction

2. Integration of LSTM networks for time-series forecasting

3. Application of SHAP for interpreting model decisions

4. Demonstration of improved accuracy and explainability compared to baseline models

---

II. Literature Review

Early approaches to energy forecasting relied on statistical models such as ARIMA, which are limited in capturing nonlinear patterns. More recent studies have adopted deep learning methods, particularly LSTM networks, to improve prediction performance.

Federated Learning was introduced as a distributed learning paradigm that avoids direct data sharing. However, most FL-based studies focus primarily on performance and privacy, without addressing model interpretability.

On the other hand, Explainable AI techniques such as SHAP have been widely used to interpret complex machine learning models. These methods provide insights into feature importance but are typically applied in centralized settings.

In this work, we combine these directions by integrating explainability into a federated learning framework for smart grid applications.

III. Proposed Methodology

A. Model Overview

The proposed system combines three key components: local LSTM models, federated aggregation, and SHAP-based explanation. Each client independently trains a model using its local data, ensuring that raw data never leaves the device.

B. Federated Learning Process

The training process follows an iterative communication cycle:

1. Each client initializes and trains a local LSTM model

2. Model parameters are transmitted to the central server

3. The server aggregates parameters using a weighted averaging scheme

4. The updated global model is shared back with clients

Federated Averaging Formula:

□_□□□□□□ = □ (□_□ / □) * □_□

Where:

- W_k represents the parameters of client k

- n_k is the number of samples at client k

- n is the total number of samples across all clients

C. Explainability using SHAP

To improve interpretability, SHAP values are computed for model predictions. This helps identify the contribution of each feature, such as temperature, time of day, and appliance usage, to the final prediction.

IV. System Architecture

The architecture consists of multiple smart meters connected to local devices, each performing independent training. A central server coordinates model aggregation without accessing raw data. An additional explainability module analyzes predictions to provide meaningful insights.

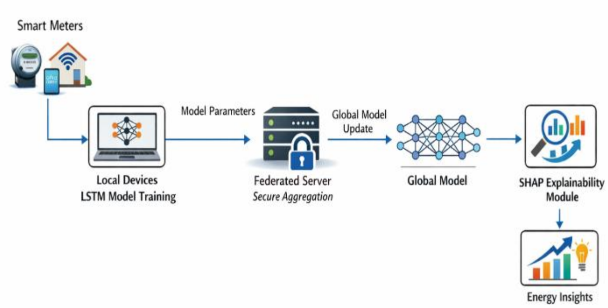

Fig. 1: Explainable Federated Learning Architecture for Smart Grid Energy Prediction

V. Experimental Setup

The proposed model was implemented using Python with TensorFlow and the SHAP library. Experiments were conducted on a smart meter dataset containing approximately 10,000 records.

The dataset includes features such as energy consumption, temperature, time intervals, and appliance usage patterns. The model was trained over multiple communication rounds to ensure convergence.

VI. Results and Discussion

Model	Accuracy (%)	Privacy	Explainability
Random Forest	88.5	Low	Medium
LSTM	91.2	Low	Low
Federated Learning	92.8	High	Medium
Proposed XFL	94.3	High	High

The experimental results indicate that the proposed approach achieves the highest accuracy among all evaluated models. Unlike centralized models, the federated approach preserves user privacy. Additionally, SHAP-based explanations provide clear insights into the factors influencing predictions, making the model more trustworthy.

VII. Conclusion

This work presents an explainable federated learning framework for energy consumption prediction in smart grids. By combining distributed training with model interpretability, the proposed system addresses both privacy and transparency challenges. Future work can focus on real-time deployment and scaling the framework to larger datasets.

References

[1] B. McMahan et al., "Communication-Efficient Learning of Deep Networks from Decentralized Data," AISTATS, 2017.

[2] S. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions," NIPS, 2017.

[3] T. Li et al., "Federated Learning: Challenges, Methods and Future Directions," IEEE, 2020.