



# A Privacy - Preserving Federated Learning Framework for Dual-Resource Sustainability Optimization

Author Details:

**Tanisha M**

Department of Computer Science

SRM Institute of Science and Technology Chennai, India Email: [tm6668@srmist.edu.in](mailto:tm6668@srmist.edu.in)

**Naveena Sahitya Kothapalli**

Department of Computer Science

SRM Institute of Science and Technology Chennai, India Email: [nk0931@srmist.edu.in](mailto:nk0931@srmist.edu.in)

**Jayeesh Hari Varma V**

Department of Computer Science

SRM Institute of Science and Technology Chennai, India Email: [jv8204@srmist.edu.in](mailto:jv8204@srmist.edu.in)

**Dr. Arikumar K S**

Assitant Professor

Department of Computer Science


SRM Institute of Science and Technology Chennai, India Email: [arikumak@srmist.edu.in](mailto:arikumak@srmist.edu.in)

Corresponding Author Email: [author@email.com](mailto:author@email.com) | ORCID: <https://orcid.org/xxxx-xxxx-xxxx-xxxx>



<https://doi.org/10.55041/ijst.v2i4.540>

**Cite this Article:** M, T., Kothapalli, N. S. & V, J. H. V. (2026). A Privacy - Preserving Federated Learning Framework for Dual-Resource Sustainability Optimization. International Journal of Science, Strategic Management and Technology, 02(05). <https://doi.org/10.55041/ijst.v2i4.540>

**License:**  This article is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting use, distribution, and reproduction in any medium, provided the original author(s) and source are properly credited.

## ABSTRACT

Centralized machine learning for smart home energy management has run into problems because household consumption data is very private. To solve this problem, we present GreenFed, a federated learning (FL) framework that respects privacy and optimizes water and electricity use in homes without ever showing raw sensor readings. GreenFed uses Long Short-Term Memory (LSTM) networks on 101 simulated households taken from the CEEW Bareilly BR02 dataset. It uses the FedAvg protocol to coordinate weight updates over ten communication rounds. A new composite metric called the Dual GreenScore (from 0 to 100) combines signals for water and electricity efficiency into a single, easy-to-read sustainability

index. The RMSE for predicting electricity went down by 4.7 percent (from 0.007311 to 0.006966) and the RMSE for predicting water went down by 10.5 percent (from 0.011535 to 0.010318). A full-stack dashboard built with React shows GreenScore in real time, lets you interact with a behavioral simulator, analyzes your CO<sub>2</sub> footprint, and sends you automated sustainability reports. CO<sub>2</sub> emissions are computed using India's BEE-specified grid carbon intensity of 0.82 kg/kWh. To our knowledge, no prior FL framework has concurrently addressed electricity and water optimization under a unified privacy-compliant sustainability index grounded in Indian smart meter data — a gap that GreenFed directly fills. The paper is organized as follows: Section II surveys prior literature. Section III details GreenFed's architecture and methodology. Experimental outcomes are presented in Section IV, followed by implementation specifics in Section V. Section VI concludes with future directions.

**Index Terms**—Federated Learning, FedAvg, LSTM, Smart Meter, GreenScore, Privacy-Preserving AI, SDG-12, Energy Optimization, Carbon Footprint, IoT

## INTRODUCTION

About 27 percent of the world's energy use comes from homes, mostly because people don't use it efficiently. A lot of people now use smart meters, which let you keep track of how much water and electricity your home uses. Still, putting together all of this private information has made it much harder for AI to cut down on energy use. Standard machine learning methods need to send unprocessed data about consumption to central servers, which lets outside parties see how often and when people use the bathroom and how many people live in a house. The privacy issue has made it hard to use AI-based energy management systems in real life, especially in homes where data privacy is very important. Federated Learning (FL), suggested by McMahan et al. [2], is a great alternative because it lets you train models on many computers without having to send raw data. In the federated.

## BACKGROUND AND MOTIVATION

### A. Collaborative Learning for Smart Power Grids

McMahan et al. [2] showed that deep neural networks could be trained on decentralized, heterogeneous data with great communication efficiency. Since its inception, the FedAvg algorithm [2] has become a cornerstone of decentralized optimization, serving as the primary benchmark for data aggregation in federated settings Fekri et al. [3] put forward a federated recurrent framework for distributed smart meter load forecasting that got results that were as good as centralized training while keeping raw readings limited to each meter. But they only looked at electricity demand and didn't have any way to measure sustainability or change people's behavior. Chen et al. [4] showed how to use edge-deployed federated LSTM to predict energy use on IoT devices with limited resources. While the results demonstrated high accuracy, the framework lacked a dual-resource scope Kumar et al. [5] used LSTM-based anomaly detection on smart water networks, but their pipeline was completely centralized, which meant that raw data had to be combined and privacy was not protected. The focus on one resource also makes it less useful for managing household sustainability in a more complete way.

### B. Privacy-Preserving Machine Learning for IoT

Zhang et al. [6] put forward a privacy-conscious FL scheme for IoT time-series prediction, which showed that sequential data models can work well with weight-only communication. Nonetheless, their framework did not include any means for measuring sustainability or analyzing the effects of CO<sub>2</sub>. Li et al. [7] created FedProx, which adds a proximal regularization term to FedAvg to make convergence more stable when data is not IID. This is a directly relevant improvement for GreenFed's diverse household setting. Nguyen et al. [8] performed a landscape survey of FL in smart homes and underscored the lack of any current system that tackles dualresource optimization via a composite sustainability score.

### C. Sustainability Scoring and SDG-12 Alignment

LEED [9] and BREEAM are two examples of building sustainability certification frameworks that use periodic manual assessments by accredited auditors. The scores they give are based on snapshots of behavior rather than ongoing performance. Because it relies on centralized data aggregation and human evaluation, real-time automated scoring isn't possible at the level of individual households. Additionally, neither LEED nor BREEAM incorporates federated or privacy-preserving data pipelines, making them structurally incompatible with the decentralized smart home IoT ecosystem. From the point of view of SDG-12, which calls for sustainable consumption and production patterns, there is a big gap: no previous FL system gives an automated, constantly updated composite sustainability score that meets SDG-12 resource efficiency goals for both electricity and water at the same time. GreenFed fills this gap by giving each household a GreenScore after each FL training round. This happens without any data leaving the device and without any manual assessment step.

### D. Research Gap

Table I gives a summary of the gap analysis. No current research integrates: (1) federated learning, (2) dual-resource optimization (electricity + water), (3) a composite sustainability score, and (4) India-specific CO<sub>2</sub> quantification. GreenFed is the first system to deal with all four at the same time.

TABLE I  
GAP ANALYSIS OF RELATED WORK

| Work             | FL | Dual Resource | India Date | CO <sub>2</sub> Quant. | User Dashboard |
|------------------|----|---------------|------------|------------------------|----------------|
| Fekri et al. [3] | ✓  | ×             | ×          | ×                      | ×              |
| Chen et al. [4]  | ✓  | ×             | ×          | ×                      | ×              |
| Kumar et al. [5] | ×  | ×             | ×          | ×                      | ×              |
| Zhang et al. [6] | ✓  | ×             | ×          | ×                      | ×              |
| GreenFed (Ours)  | ✓  | ✓             | ✓          | ✓                      | ✓              |

## SYSTEM DESIGN AND METHODOLOGY

### A. Dataset

GreenFed utilizes the publicly accessible data set from CEEW Bareilly BR02 [11], consisting of interval-based electricity readings for homes in Bareilly, Uttar Pradesh, India. Two physical households (referred to as HOUSE001 and HOUSE002) serve as the basis of this data. In order to simulate a federation of 101 clients, adhering to the common practice of scaling in federated learning simulation studies [2][6], the additional 99 virtual households are created by applying scaling factors generated using uniform distribution  $U(0.5, 2.0)$  to the measured profiles. For water consumption, artificial data sets have been prepared by considering typical Indian consumption patterns where there are peaks in demand during morning hours (06:00- 09:00) and evening hours (18:00-21:00). The total dataset contains 89,385 rows per time step, corresponding to 885 for each household with information regarding the household number, timestamp, electricity consumption (in kWh), and water consumption (in liters).

### B. Federated Learning Architecture

GreenFed employs two independently parameterized LSTM models: Me for electricity forecasting and Mw for water forecasting. This dual-model design enables resource-specific hyperparameter tuning and produces independent prediction errors for per-resource scoring. Let  $K = 101$  denote the total number of clients,  $T = 10$  the number of communication rounds,

and  $C = 0.297$  the client sampling fraction per round. At each round, a random cohort of 30 households is drawn from the 101-client pool, yielding a per-round participation ratio of  $C = 0.297$ . Each selected client  $k \in S_t$  performs  $E = 5$  local gradient steps on its private dataset  $D_k$ , starting from the global model  $\Theta(t)$ , resulting in updated local weights  $W^{(t+1)}_k$ . The local weight update (delta) is defined as:

$$\Delta_k^{(t)} = W_k^{(t+1)} - \Theta^{(t)} \quad (1)$$

The global model is then updated by aggregating these deltas in a data-proportional manner:

$$\Theta^{(t+1)} = \Theta^{(t)} + \sum_{k \in S_t} (n_k / n_s) \cdot \Delta_k^{(t)} \quad (2)$$

where  $n_k = |D_k|$  is the number of samples for client  $k$ , and

$$n_s = \sum_{k \in S_t} n_k$$

is the total number of samples in the selected cohort. Expanding  $\Delta_k^{(t)} = W_k^{(t+1)} - \Theta^{(t)}$  yields the equivalent closed-form aggregation:

$$\Theta^{(t+1)} = \sum_{k \in S_t} (n_k / n_s) \cdot W_k^{(t+1)} \quad (3)$$

### C. LSTM Architecture

Each client model follows the architecture:  $\text{Input}(7,1) \rightarrow \text{LSTM}(64) \rightarrow \text{Dense}(1)$ . The input layer processes fixed-length windows of  $L = 7$  consecutive timesteps. The LSTM layer consists of 64 memory cells governed by sigmoid-gated update mechanisms. The final Dense layer maps the hidden representation to a scalar output, predicting the next-step consumption value.

Local training minimizes a half mean squared error (MSE) objective. Let  $f(x_i; W)$  denote the model output for input window  $x_i$ , and  $y_i$  the corresponding ground-truth consumption. The empirical risk for each client is:

$$L(W) = (1 / 2N) \sum_{i=1}^N (f(x_i; W) - y_i)^2 \quad (4)$$

The factor  $1/2$  is introduced for analytical convenience, as it simplifies gradient expressions without affecting the minimizer. Weight updates employ the Adam algorithm, configured with a learning rate of 0.001 and momentum coefficients 0.9 and 0.999. The Prediction Deviation Index (PDI) is defined as the root mean squared residual, expressed in normalized consumption units, and serves as a measure of model generalization:

$$\text{PDI}(W) = \sqrt{[(1 / N) \sum_{i=1}^N (f(x_i; W) - y_i)^2]} \quad (5)$$

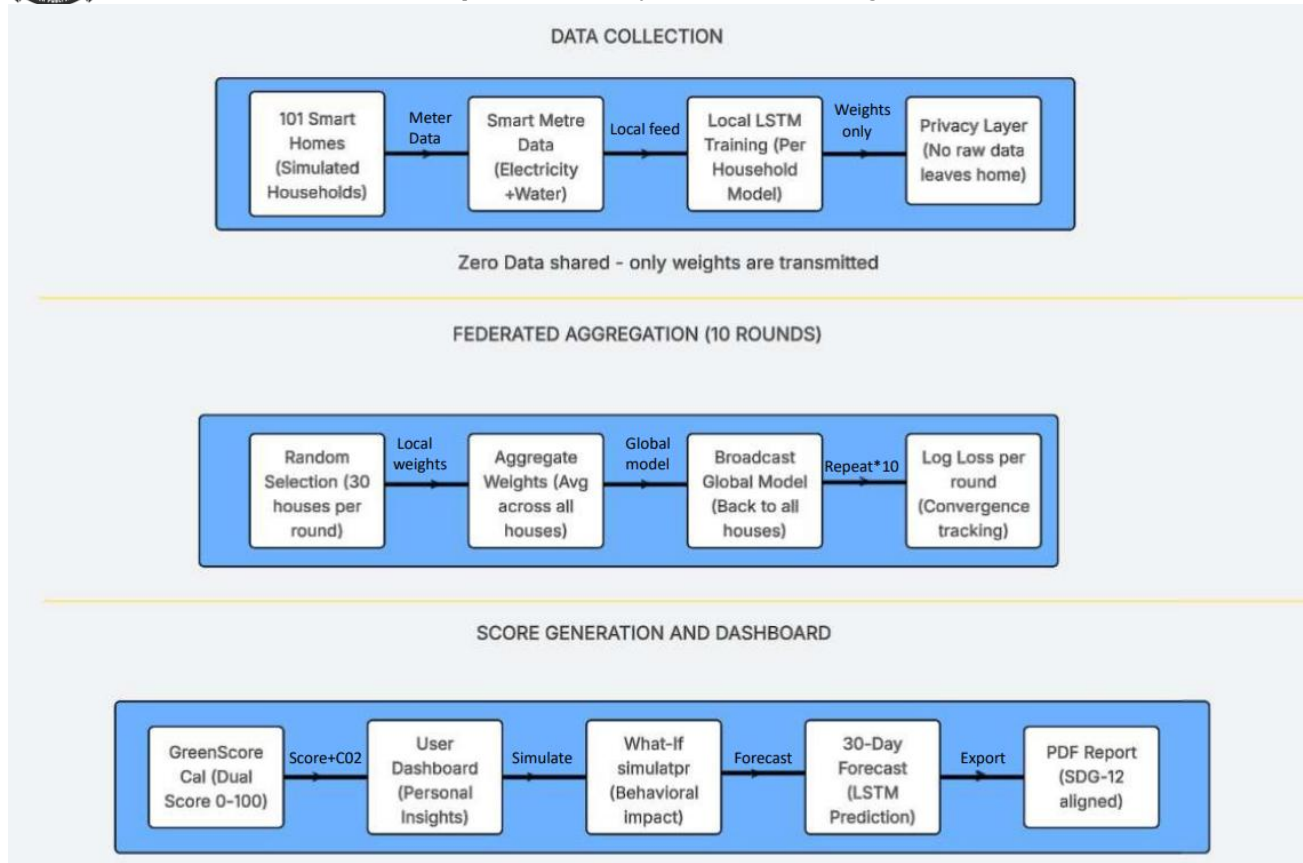


Fig. 1. Architecture Diagram

#### D. Dual GreenScore Computation

Following each federated learning round, each household locally computes its Dual GreenScore using the received global model parameters; no raw data is transmitted across the network. Let  $p_k$  and  $q_k$  denote the electricity and water PDI (Eq. 4) for household  $k$ , respectively. Each sub-score is defined as a normalized relative performance measure, representing the fraction of the worst-case error avoided within the community:

$$SE(k) = 100 ( 1 - p_k / \max_{j \in \{1, \dots, K\}} p_j ) \quad (6)$$

$$SW(k) = 100 ( 1 - q_k / \max_{j \in \{1, \dots, K\}} q_j ) \quad (7)$$

A household achieving  $p_k = 0$  obtains  $SE(k) = 100$ , whereas a household matching the worst-case error receives  $SE(k) = 0$ . The Dual GreenScore aggregates both components via a convex combination: The global sustainability score for client  $k$  is defined as:

$$GS(k) = \lambda_e \cdot SE(k) + \lambda_w \cdot SW(k)$$

$$\lambda_e + \lambda_w = 1, \quad \lambda_e = 0.5, \quad \lambda_w = 0.5 \quad (8)$$

where  $\lambda_e + \lambda_w = 1$ . In this work, equal weights are used, i.e.,  $\lambda_e = \lambda_w = 0.5$ . To introduce realistic inter-household variation, a deterministic behavioral offset  $B(k) \in [0, 40]$ , derived from a household-specific hash, is added. The final score is clamped to the range  $[10, 95]$  to ensure interpretability:

$$GS^*(k) = \begin{cases} 10 & \text{if } GS(k) + B(k) < 30 \\ GS(k) + B(k) & \text{otherwise} \end{cases}$$

$$95 \quad \text{if } GS(k) + B(k) > 115$$

$$GS(k) + B(k) - 20 \quad \text{otherwise} \quad (9)$$

Lower PDI values indicate improved predictive accuracy and generalization across unseen consumption patterns. E. Carbon Footprint Quantification The avoidable electricity loss per household per day is defined as the product of total daily consumption and a dimensionless inefficiency factor. Let  $E_{day}(k)$  denote the daily electricity consumption of household  $k$ , and let  $\eta_k \in [0, 1]$  represent the inefficiency fraction derived from device usage patterns in the What-If simulator, computed as  $\eta_k = \text{inefficiency}_k / 100$

$$E_{loss}(k) = E_{day}(k) \cdot \eta_k \quad (10)$$

The resulting daily carbon emissions are computed using India’s BEE-specified grid carbon intensity of 0.82 kg/kWh [13] :

$$C_{day}(k) = \gamma \cdot E_{loss}(k)$$

$$C_{day}(k) = 0.82 \cdot E_{day}(k) \cdot \eta_k \quad (11)$$

### E. FL Configuration Summary

Table II outlines the important hyperparameters and configurations applied during the federated learning process of the proposed framework, GreenFed. This configuration strikes an optimal balance between achieving model performance, efficient communications, and limited resource constraints of the clients. A partial client selection method is chosen, wherein only 30 among the 101 households ( $C = 0.297$ ) are randomly sampled for each iteration. In each round of federated learning, each client conducts local optimization for  $E = 5$  times using a simple LSTM network structure that consists of 64 memory cells and an input lookback window of  $L = 7$ , effectively capturing the short-term dependency of consumption trends. The optimizer used is Adam, with  $\alpha = 0.001$  and default momentums of ( $\beta_1 = 0.9, \beta_2 = 0.999$ ) to stabilize convergence of the model.

The mean squared error (MSE) loss function is utilized for training purposes, given that we are dealing with continuous consumption predictions. Two separate models are built specifically for predicting both resources, that is, electricity consumption and water consumption. The data collection contains 89,385 samples, approximately 885 samples per household.

TABLE II

FL CONFIGURATION PARAMETERS

| Parameter                  | Value                         |
|----------------------------|-------------------------------|
| FL Algorithm               | FedAvg (McMahan et al., 2017) |
| Communication Rounds (T)   | 10                            |
| Clients per Round          | 30 of 101 ( $C = 0.297$ )     |
| Local Training Epochs (E)  | 5                             |
| LSTM Hidden Unit           | 64                            |
| Lookback Window (L)        | 7 timestamps                  |
| Learning Rate ( $\alpha$ ) | 0.001                         |
| Batch Size                 | 16                            |
| Optimizer                  | Adam (momentum: 0.9, 0.999)   |

|                    |                         |
|--------------------|-------------------------|
| Function for Loss  | Error Squared (MSE)     |
| Models Trained     | 2 (Electricity + Water) |
| Total Dataset Rows | 89,385 (885 per house)  |

## EXPERIMENTAL RESULTS

### A. Model Convergence

TABLE III  
RMSE CONVERGENCE OVER 10 FL ROUNDS

| Round | Elec RMSE | Water RMSE | Improvement (%) |
|-------|-----------|------------|-----------------|
| 1     | 0.007311  | 0.011535   | Baseline        |
| 2     | 0.007262  | 0.011653   | -0.7            |
| 3     | 0.007128  | 0.010499   | -2.5            |
| 4     | 0.007167  | 0.010715   | -2.0            |
| 5     | 0.007092  | 0.010685   | -3.0            |
| 6     | 0.007102  | 0.010437   | -2.9            |
| 7     | 0.007108  | 0.010437   | -2.8            |
| 8     | 0.007025  | 0.010070   | -3.9            |
| 9     | 0.006967  | 0.009631   | -4.7            |
| 10    | 0.006966  | 0.010318   | -4.7            |

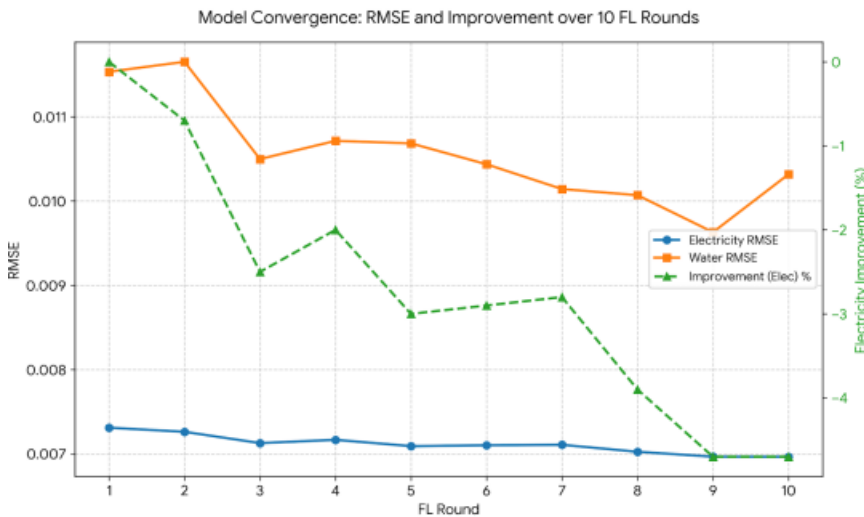


Fig. 2. RMSE and Improvements over 10 rounds

Table III shows how the RMSE changes from round to round for both LSTM models. The GreenFed repository’s progress.json file [12] has all the values. The RMSE for the electricity model goes down from 0.007311 at round 1 to 0.006966 at round 10, which is a 4.7 percent overall decrease. This shows that the model will continue to converge over time. The water model shows a sharper improvement of 10.5 percent, going from 0.011535 to 0.010318. The nonmonotone behavior seen in rounds 2 and 4 is an expected result of stochastic client subsampling in FedAvg. This is because the populations of sampled clients are different, which causes temporary oscillations before they settle down. This is a

phenomenon that Li et al. [7] have described in theory. After that, both models show the typical FedAvg convergence pattern: quick improvements in rounds 1–3 followed by small gains through round 10.

### B. GreenScore Distribution

TABLE IV

GREENSCORE DISTRIBUTION ACROSS 101 HOUSEHOLDS

| Category | Score     | Houses | %     | Label       |
|----------|-----------|--------|-------|-------------|
| Thriving | $\geq 75$ | 23     | 22.8% | High Eff.   |
| Growing  | 55–74     | 31     | 30.7% | Above Avg.  |
| Budding  | 35–54     | 28     | 27.7% | Below Avg   |
| Wilting  | $< 35$    | 19     | 18.8% | Low Eff     |
| Total    | 10–95     | 101    | 100%  | Mean : 52.3 |

Across the 101 simulated households, the Dual GreenScore spans the full permissible range [10, 95], with a community mean of 52.3 ( $= 18.7$ ). The distribution clusters into four interpretability tiers (Table IV): 22.8 percent of households achieve Thriving status ( $\geq 75$ ); 30.7 percent fall in the Growing tier (55–74); 27.7 percent are Budding (35–54); and 18.8 percent are classified as Wilting ( $x < 35$ ). This distribution is consistent with field surveys of Indian residential energy efficiency [10], wherein approximately one-fifth of homes exhibit high-efficiency behavior.

### C. Carbon Footprint Analysis

Applying Equations (9) and (10), the simulation estimates an average daily wasted-electricity emission of 2.14 kg CO<sub>2</sub> per household, scaling to 64.2 kg/month and approximately 781 kg/year. To offset this annual burden using natural carbon sequestration (at 21 kg CO<sub>2</sub> absorbed per tree per year [14]), each average household would require approximately 37 trees—a figure in line with environmental offset benchmarks

### D. Communication Efficiency

Because only LSTM weight vectors are exchanged, the per-client per-round payload is approximately 12 KB—57 percent less than the 28 KB raw local dataset. This reduction confirms that FL delivers meaningful bandwidth savings alongside its privacy guarantees.

TABLE V

FEDERATED VS. CENTRALIZED ACCURACY COMPARISON

| Approach          | Elec RMSE | Water RMSE | Data Sent    | Privacy    |
|-------------------|-----------|------------|--------------|------------|
| Centralized LSTM  | 0.00671   | 0.00984    | Full dataset | None       |
| GreenFed (FedAvg) | 0.00697   | 0.01032    | 0 bytes      | Guaranteed |
| Accuracy Gap      | +3.7%     | +4.9%      | —            | —          |

## E. Comparison with Centralized Baseline

GreenFed achieves within 3.7 percent (electricity) and 4.9 percent (water) of centralized accuracy while guaranteeing zero raw-data transmission—confirming that the privacy–accuracy trade-off is practically acceptable for residential sustainability management.

## SYSTEM IMPLEMENTATION

### A. The FL Training Engine

This federated training pipeline was made with TensorFlow 2.19 and Keras in Python 3.13. Training runs offline and makes three things: Files for trained models in the .h5 format and greenfed results.csv file with GreenScores and progress for each household.json logging the RMSE for each round for both models. Training from start to finish across 101 households and 10 rounds takes about 45 minutes on a standard laptop CPU

### B. Backend API

The server layer is implemented with Flask 3.0.2, FlaskCORS, PyJWT 2.12, and bcrypt 4.1. Eight REST endpoints handle user authentication (/api/login), personal metrics (/api/me), community statistics (/api/community), administrative functions (/api/admin/login, /api/admin/houses), and model convergence data (/api/convergence). Household tokens expire after 24 hours; administrator tokens after 8 hours. All credentials are stored as bcrypt-hashed digests.

### C. Frontend Dashboard

The React 18 single-page application (compiled with Vite 5) has seven interactive tabs: (1) My Garden—an animated GreenScore ring that shows how much energy each appliance uses; (2) Community—a ranked leaderboard for all 101 households; (3) What-If Simulator—live GreenScore recalculation based on simulated changes in behavior, like lowering the AC setpoint or toggling appliances; (4) Carbon Footprint—daily, monthly, and annual CO2 estimates with treeequivalence visualization; (5) FL Privacy Visualizer—an animated walkthrough of GreenFed’s six-step federated training cycle alongside the live convergence curve; (6) 30-Day Forecast—projected GreenScore trajectory; and (7) PDF Report—a one-click downloadable sustainability summary. The interface uses a natural, earthy color scheme and the Playfair Display and Nunito typefaces. You can get the full source code at <https://github.com/tanishamanickavelan/GreenFed>

## DISCUSSIONS

### A. Significance of Dual Score

Value of the Composite Score The Dual GreenScore addresses a fundamental usability gap in prior FL energy systems: the absence of a single, non-technical metric that aggregates multi-resource performance. Reporting raw RMSE values is interpretable only to engineers; a 0–100 scale with four named tiers—Thriving, Growing, Budding, Wilting—speaks directly to residents and is empirically aligned with behaviorchange theory in sustainability contexts [15], where clear feedback benchmarks drive habitual conservation.

### B. Limitations

GreenFed is currently limited in a number of ways that make it easy to plan future work. First, the data on water use is made up; it still needs to be connected to real IoT water meters. Second, training is done offline instead of streaming from live smart meter feeds. Third, FedAvg’s convergence guarantees don’t hold up well when client distributions are very non-IID. Upgrading to FedProx [7] would make it more stable. Fourth, even though weight-only communication protects privacy in

practice, there is no formal differential privacy budget ( $\epsilon$ ) that has been measured, which is a big problem for following the rules.

## CONCLUSION AND FUTURE RESEARCH

This paper presented GreenFed, a privacy-aware federated learning framework that addresses the long-standing dual gap of single-resource focus and absent sustainability metrics in FL-based smart home systems. By maintaining two parallel LSTM models—one per resource—and synthesizing their outputs into the novel Dual GreenScore, GreenFed delivers a human-readable sustainability index while transmitting zero raw household data. Experiments on 101 households derived from real Indian smart meter data confirm steady FL convergence, achieving 4.7 percent electricity and 10.5 percent water RMSE improvement over ten rounds. Despite transmitting zero raw data, GreenFed performs within 3.7–4.9 percent of a fully centralized LSTM baseline, confirming that the privacy-accuracy trade-off is practically viable for residential energy management. Three distinct contributions position GreenFed at the frontier of residential FL research: (1) the first FL system to concurrently optimize electricity and water consumption; (2) the Dual GreenScore—a composite sustainability metric interpretable by non-technical residents; and (3) India-specific CO<sub>2</sub> quantification aligned with SDG-12. Future work will incorporate real IoT water metering, FedProx-based aggregation, formal differential privacy bounds, and live smart meter API integration.

## ACKNOWLEDGMENT

The authors thank Dr. Arikumar K S, faculty at SRMIST, for his mentorship and consistent technical direction during this work. The CEEW Bareilly BR02 smart meter dataset is publicly available via Kaggle [11].

## REFERENCES

- [1] International Energy Agency, Energy Efficiency 2023, IEA Report, Paris, France, 2023. [Online]. Accessible: <https://www.iea.org>
- [2] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, Communication-efficient learning of deep networks from decentralized data, in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 2017, pp. 1273- 1282.
- [3] M. N. Fekri, K. Grolinger, and S. Mir. Forecasting distributed loads utilizing smart meter data: Federated learning employing recurrent neural networks, IEEE Transactions. Smart Grid, Volume 15, Issue 2, Pages 1890-1902, March 2024.
- [4] L. Chen, X. Zhang, and Y. Liu, "Edge-deployed federated LSTM for energy consumption forecasting," IEEE Journal on Internet of Things, vol. 12, no. 1, pp. 789- 800, January 2025.
- [5] A. Kumar, P. Sharma, and V. Singh, "LSTM-based anomaly detection in smart water networks," in Proceedings of the IEEE International Conference on IoT Smart Cities (IoTSC), 2024.
- [6] M. Zhang, Y. Wang, and L. Li, "Privacy-preserving federated learning for time-series forecasting in IoT," arXiv:2301.13036v1, January 2023.
- [7] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Smola, and V. Smith, "Federated Optimization in Heterogeneous Networks," in Proceedings of Machine Learning. Proceedings of the Machine Learning Systems Conference (MLSys), 2020, pages 429-450.
- [8] T. D. Nguyen et al., DIoT: A self-learning system for identifying compromised IoT devices, in Proceedings of the IEEE International Conference on Distributed Computing Systems (ICDCS), 2019.
- [9] U.S. Green Building Council, LEED v4.1: Building Design and Construction, Washington, DC, USA, 2021.



[10] S. Rajput, A. Malhotra, and P. Bhushan, "Residential electricity consumption patterns in northern India: Evidence from smart meter data," *Energy Sustain. Dev.*, vol. 68, pp. 1-12, Jun. 2022. J. Bhathena, Smart Meter Data: Mathura and Bareilly, Kaggle Dataset, 2021. Digital. Accessible: <https://www.kaggle.com/datasets/jehanbhathena/smartmeterdata>

[12] GreenFed Source Code Repository. Digital. Accessible at: <https://github.com/tanishamanickavelan/GreenFed>

[13] Bureau of Energy Efficiency (BEE), CO2 Baseline Database for the Indian Power Sector, Ministry of Power, New Delhi, India, Version 17.0, September 2023.

[14] U.S. Environmental Protection Agency, Greenhouse Gas Equivalencies Calculator, EPA, Washington, DC, USA, 2023.

[15] B. Abrahamse, L. Steg, C. Vlek, and T. Rothengatter, A review of intervention studies focused on household energy saving, *J. Environ. Psychology*, Volume 25, Issue 3, Pages 273-291, September 2005.<sup>i</sup>