Electronics and Information Technologies, 2025, 30, 15–32
http://publications.lnu.edu.ua/collections/index.php/electronics/index

FEDERATED LEARNING WITH STOCHASTIC GRADIENT DESCENT FOR SMART METER ENERGY FORECASTING

Bharat Khushalani
Department of Artificial Intelligence,
Shri Vishnu Engineering College for Women,
Bhimavaram, India 534202.
Email: bharat@svecw.edu.in

Khushalani, B. (2025). Federated Learning with Stochastic Gradient Descent for Smart Meter Energy Forecasting. *Electronics and Information Technologies*, 30, 15–32.
https://doi.org/10.30970/eli.30.2

ABSTRACT

Background. Smart meters are widely used to monitor household energy consumption and help improve energy efficiency. However, collecting this data in a centralized location raises privacy concerns, as detailed consumption records can reveal sensitive household behavior. Federated learning provides an alternative approach by allowing models to be trained directly on user devices without sending raw data to a central server.

Materials and Methods. This study developed a simulation-based framework to test federated learning for forecasting short-term electricity usage. We created synthetic data representing hourly energy consumption for 100 simulated households, incorporating daily usage cycles and household-specific patterns. A simple neural network was trained locally on each household's data using a standard optimization method, and model updates were shared with a central server to improve a shared global model.

Results and Discussion. The federated model achieved forecasting accuracy nearly equal to a traditional centralized model while keeping data private. Key factors affecting performance included how often devices were trained locally before sharing results and how many households participated in each training round. The approach remained accurate even when only half the devices contributed at any time. Compared to non-collaborative models trained independently by each household, the federated approach offered a substantial improvement in prediction accuracy. These findings show that good performance can be achieved while protecting user privacy and using simple models suitable for low-power devices.

Conclusions. This work shows that a well-designed simulation with realistic energy usage data can help evaluate federated learning methods under practical constraints. Even simple models, when trained in a decentralized and privacy-preserving way, can offer useful predictions for smart energy systems. The approach is suitable for real-world deployment and can help advance privacy-respecting energy analytics.

Keywords: Federated Learning, Smart Meters, Energy Forecasting, Stochastic Gradient Descent, Privacy-Preserving Machine Learning, Decentralized Optimization.

INTRODUCTION

With the proliferation of smart grid technology, smart meters are increasingly deployed to collect and transmit household electricity usage data. The increasing digitalization of the energy sector, driven by the deployment of smart grid technologies, has resulted in the widespread adoption of smart meters in residential and commercial buildings. These smart meters enable real-time monitoring of electricity consumption at a granular level, typically recording data in intervals ranging from every few minutes to an hour. This fine-grained data collection opens new opportunities for optimizing electricity usage, forecasting energy demand, detecting faults, and enabling dynamic pricing models [1]. However, while the availability of such high-resolution data is beneficial for energy providers and consumers alike, it also raises significant privacy concerns. These meters enable fine-grained energy analytics, but collecting such data in a central repository is also a matter of confidentiality. When centralized systems are used to aggregate and analyze this data, individuals' behavioral patterns can be inferred—such as when they are home, what appliances they use, and even their daily routines. This has sparked a growing demand for privacy-preserving machine learning methods that can extract useful insights from smart meter data without exposing sensitive personal information [2].

Federated Learning (FL) offers a compelling solution to this problem by allowing model training to occur directly on decentralized devices [3], such as smart meters, without transferring the raw data to a central server [4]. Federated Learning offers a promising solution by enabling decentralized training where raw data never leaves the device [5]. In a federated learning setup, each smart meter independently computes updates to a shared global model using its local data and sends only the updated model parameters (or gradients) to a central aggregator. These updates are then combined, typically using a weighted average, to produce an improved global model. Since the data never leaves the client device, the risks associated with data breaches, surveillance, or unauthorized access are significantly mitigated [6].

This paradigm shift from centralized to federated training is particularly suited to the energy domain [7], where data is naturally distributed across households, industrial sites, and substations [8]. Each location generates its own consumption data, yet the underlying patterns—such as daily usage cycles, response to weather changes, and peak load timings—are often shared across the population. This shared structure can be exploited by collaborative learning without sacrificing data ownership or privacy.

In this paper, we propose a federated learning framework for short-term energy consumption forecasting using data collected from a network of smart meters. Our approach focuses on using Stochastic Gradient Descent (SGD) as the local optimization strategy within each client. We simulate energy usage data for 100 synthetic households, each representing a unique combination of usage patterns influenced by seasonality, noise, and household-specific characteristics. By treating each household as a separate federated client, we create a realistic testbed for evaluating the performance of our privacy-preserving learning framework.

Stochastic Gradient Descent remains one of the most widely used optimization algorithms in machine learning [9] due to its simplicity, efficiency, and ability to scale across large datasets [10]. In the federated setting, SGD becomes even more powerful [11] because it allows for incremental updates using small local batches [12], thereby reducing the computational load on edge devices with limited resources. Moreover, the federated SGD paradigm supports asynchronous and parallel computations, further enhancing the scalability of the system. However, deploying SGD in a federated environment introduces new challenges. These include issues such as model divergence due to non-IID (independent and identically distributed) data across clients [13], communication inefficiency arising from frequent parameter exchanges, and the need for robust aggregation mechanisms to handle variability in client participation.

Our proposed method addresses these challenges by incorporating several practical design choices. First, we simulate data heterogeneity across households by assigning different parameters to their energy generation functions, such as peak usage times and noise levels. This mimics real-world scenarios where some users may have predictable schedules while others exhibit high variability. Second, we evaluate the effect of varying the number of local epochs, learning rates, and communication rounds on model convergence. This allows us to explore tradeoffs between communication cost and model performance. Third, we compare our federated approach with both centralized training (where all data is aggregated at a central server) and local-only models (where each client trains in isolation without sharing updates). These baselines help quantify the benefits of collaboration and the costs of centralization.

The application of federated learning to energy forecasting is relatively novel and underexplored. Most existing FL research has focused on domains such as mobile keyboard prediction, medical imaging, and financial analytics [14]. In contrast, the energy sector has unique temporal characteristics, long-range dependencies, and highly personalized consumption patterns that require specialized modelling techniques [15]. Our work contributes to this emerging area by demonstrating the feasibility of deploying federated neural networks for time-series prediction in smart energy systems.

Beyond the technical benefits, the societal implications of our approach are significant. By enabling accurate energy demand forecasting at the household level without violating user privacy, federated learning can support more efficient grid operation, reduce peak load stress, and facilitate the integration of renewable energy sources. Accurate forecasting allows utilities to schedule power generation more effectively, prevent blackouts, and offer dynamic pricing schemes that incentivize consumers to shift their usage patterns. From the consumer's perspective, federated learning empowers individuals to contribute to collective intelligence without relinquishing control over their personal data.

In summary, this paper introduces a federated learning approach powered by stochastic gradient descent for forecasting household energy consumption. Our simulation-based framework provides a controlled environment for analyzing the interplay between privacy, accuracy, and communication efficiency. The results show that our method can closely match centralized performance while preserving data locality, making it a promising candidate for real-world deployment in smart grid systems. In the following sections, we detail our system design, data generation process, model architecture, and empirical results, followed by a discussion of future directions in this important field.

MATERIALS AND METHODS

Federated learning is an emerging paradigm that addresses the challenge of training machine learning models across decentralized data sources while preserving data privacy. Originally proposed by [16] through the Federated Averaging (FedAvg) algorithm, FL has since been adopted and adapted across various application domains. This section reviews foundational work in FL, key developments in decentralized optimization methods such as stochastic gradient descent (SGD), applications of FL in time-series forecasting, and specific literature involving smart meters and energy analytics.

Foundations of Federated Learning

The concept of federated learning was popularized by Google's work on keyboard prediction in mobile devices [16]. In this setting, models were trained directly on users' phones, eliminating the need to upload sensitive typing data to a central server. The FedAvg algorithm, which combines local stochastic gradient descent updates with global model averaging, formed the backbone of this architecture. Since then, FL has been formalized in both synchronous and asynchronous variants, and researchers have proposed refinements such as adaptive federated optimization, federated meta-learning, and hierarchical FL.

Numerous studies have further explored the mathematical properties and convergence guarantees of FL. For example, [17] introduced FedProx, which augments the loss function with a proximal term to mitigate client drift in non-IID settings. Karimireddy et al. developed SCAFFOLD to correct client updates using control variates [18]. These contributions provide theoretical underpinnings that ensure FL can work even in challenging heterogeneous environments—such as those commonly found in energy data collected from different households.

Stochastic Gradient Descent in Federated Settings

Stochastic gradient descent (SGD) remains the workhorse of most federated optimization schemes. Its simplicity, computational efficiency, and compatibility with streaming data make it an ideal choice for edge-device training. However, standard SGD assumes IID data and synchronous updates, which often do not hold in federated scenarios.

In response, researchers have proposed adaptations of SGD that cope with client variability, delayed updates, and partial participation. For instance, asynchronous SGD algorithms have been developed to allow clients to update the server without waiting for a global synchronization point [19]. Other works focus on compressing gradients or quantizing weights to reduce communication overhead [20]. In our work, we utilize classical SGD due to its lightweight nature and ease of implementation, and demonstrate its effectiveness in a realistic simulation of distributed energy systems.

Federated Learning for Time Series Forecasting

While most early FL research targeted applications in text prediction, image classification, and healthcare diagnostics, time series forecasting has gained increasing attention. Time series forecasting in FL is challenging because of its sequential nature and the potential for data to vary widely in scale and structure across clients.

Studies such as [21] applied federated learning to IoT sensor time series data, showing that LSTM and GRU models can be trained across edge devices with minimal accuracy loss. Another work by [22] proposed Fedformer, which adapts Transformer architectures to time series forecasting under FL constraints. These approaches demonstrate that federated learning can be extended to recurrent and attention-based models; however, they often require higher resource availability on client devices than traditional SGD-based models.

Our approach differs by focusing on a simple feedforward neural network, which is more suitable for deployment in constrained environments like smart meters. To the best of our knowledge, our work is one of the first to implement federated SGD specifically for hourly energy consumption forecasting using synthetic smart meter data.

Smart Meters and Energy Analytics

Smart meters are integral to modern energy infrastructure, offering near real-time monitoring of electricity consumption. This fine-grained data can be used for a range of applications including demand-side management, load forecasting, anomaly detection, and dynamic pricing. Traditional energy analytics pipelines, however, rely on centralized machine learning approaches that collect and process large volumes of user data, often without explicit user consent.

Research in energy forecasting has traditionally employed methods such as ARIMA, support vector regression (SVR), and more recently, deep learning models like LSTM networks [23]. These models are typically trained on aggregated data from utilities or grid operators. While effective, such centralization raises concerns about data security and user privacy, especially when individual consumption patterns can reveal sensitive personal information.

There has been growing interest in applying privacy-preserving techniques to energy analytics. For example, [24] examined differential privacy in smart grid data streams, while [25] explored homomorphic encryption for secure load forecasting. These methods, however, often introduce substantial computational overhead or require specialized infrastructure.

Federated learning offers an attractive alternative by allowing raw consumption data to remain local, thereby reducing the attack surface and improving data governance. Despite its promise, the application of FL to smart meter forecasting remains limited in current literature. One notable study by [26] examined load prediction in an FL framework but relied on complex model architectures. Our work complements this line of research by offering a transparent, reproducible methodology using synthetic yet realistic data to demonstrate core FL principles in the smart grid domain.

Challenges in Federated Energy Forecasting

The application of FL to energy systems presents several domain-specific challenges. First, the non-IID nature of household consumption data poses a risk of local model drift, which can degrade global convergence. Second, the real-time requirements of energy systems call for fast convergence and communication-efficient protocols. Third, energy usage data is often sparse or irregular due to communication dropouts, which complicates time-series forecasting in a federated environment.

Recent research has proposed addressing these issues through personalized federated learning, where the global model is augmented with client-specific parameters [27], and adaptive client selection strategies that prioritize diverse participants [28]. These approaches could be incorporated into future versions of our framework to improve robustness and accuracy.

This paper builds upon a growing body of research in federated learning, time-series forecasting, and smart energy analytics. It contributes a unique combination of lightweight modelling, synthetic yet realistic simulation, and practical evaluation of SGD-based FL in a domain where privacy is both critical and frequently overlooked. While prior work has explored each of these elements in isolation, our integrated framework offers a scalable, efficient, and privacy-aware solution tailored specifically for the energy sector.

By situating our work in the broader context of existing literature, we aim to demonstrate both its novelty and relevance. As smart grids and edge computing infrastructures continue to evolve, the insights gained from this study can inform the design of next-generation energy analytics systems that are not only intelligent but also respectful of individual privacy.

Our federated learning framework for smart meter energy forecasting is designed to simulate a realistic deployment environment in which multiple distributed clients—each representing a household with a smart energy meter—train a shared model collaboratively without exposing their private data. This section describes the overall system architecture, the process of generating synthetic yet realistic household energy usage data, the model architecture used for time-series forecasting, and the implementation of the training strategy based on stochastic gradient descent within the federated paradigm.

System Architecture

The system consists of three core components: the client devices (smart meters), the global server (aggregator), and the communication protocol that enables coordination between them. Each client maintains its own local dataset and a local copy of the machine learning model. Clients perform training independently and intermittently communicate their model updates to the global server. The server then aggregates these updates to produce a new global model, which is redistributed to the clients for the next training round.

We use the Federated Averaging (FedAvg) algorithm as the coordination protocol. At each communication round t, a subset of clients $\mathcal{K}_t \subseteq \{1,2,\dots,K\}$ is selected to participate. Each client $k \in \mathcal{K}_t$ receives the current global model parameters w_t, updates them locally using mini-batch stochastic gradient descent, and returns the updated parameters $w_k^{(t+1)}$ to the server. The server then computes the new global model as:

$$w_{t+1} = \sum_{k \in \mathcal{K}_t} \frac{n_k}{\sum_{j \in \mathcal{K}_t} n_j} w_k^{(t+1)}, \tag{1}$$

where n_k is the number of samples available on client k.
This weighted average ensures that clients with more data have proportionally greater influence on the updated global model.

Synthetic Data Simulation

To rigorously evaluate the proposed system in a controlled yet meaningful way, we simulate the behavior of 100 synthetic households over a period of 180 days, with each household reporting hourly electricity consumption. The synthetic dataset is generated using a semi-parametric function that captures both periodic trends and household-specific randomness.

For each client k, we define the consumption at each hour t as:

$$x_t^{(k)} = \alpha_k \cdot \sin\left(\frac{2\pi t}{24}\right) + \beta_k \cdot \cos\left(\frac{2\pi t}{12}\right) + \gamma_k \cdot \text{Noise}(0,1), \tag{2}$$

where α_k controls the diurnal rhythm,
β_k captures sub-daily variations (e.g., midday peak),
and γ_k modulates the influence of stochasticity to reflect unpredictable usage patterns.
The parameters α_k, β_k, γ_k are drawn from uniform distributions across realistic ranges to ensure heterogeneity. In this way, some clients behave predictably, while others exhibit erratic patterns.

Each time series is then transformed into a supervised learning format by extracting sliding windows of 24 hours as input and using the 25th hour (next time step) as the target. This forms a time-series forecasting problem with one-step-ahead prediction.

Model Architecture

Given the simplicity and portability constraints of smart meters, we opt for a lightweight feedforward neural network. The model architecture is as follows:

- **Input Layer:** 24 neurons corresponding to hourly energy usage over the past day.
- **Hidden Layer:** One dense layer with 64 neurons and ReLU activation.
- **Output Layer:** A single neuron with linear activation to predict the next hour's usage.

This compact architecture strikes a balance between expressive power and computational efficiency, making it feasible for training on devices with limited resources.

Local Training Procedure

Each client independently trains the local model using mini-batch stochastic gradient descent (SGD). Let $\mathcal{D}_k$ denote the local dataset of client k, and let $\mathcal{B}_k \subset \mathcal{D}_k$ be a batch of data. The update rule is given by:

$$w_k \leftarrow w_k - \eta \cdot \nabla \ell(w_k; \mathcal{B}_k) \tag{3}$$

where η is the learning rate and $\ell(\cdot)$ is the loss function, defined as mean squared error (MSE) between the predicted and true values:

$$\ell(w; \mathcal{B}) = \frac{1}{|\mathcal{B}|} \sum_{(x_i, y_i) \in \mathcal{B}} [f_w(x_i) - y_i]^2. \tag{4}$$

Each client performs E epochs of training on its local data before sending updated weights to the server. Increasing E allows for more local computation, which can reduce communication frequency but may lead to model divergence if the data is non-IID.

Federated Training Loop

The overall federated training loop is as follows:

1. The global server initializes model parameters w_0.
2. For each round $t = 1,2, \dots, T$:
   a. Select a subset of clients $\mathcal{K}_t$.
   b. Distribute w_t to all clients in $\mathcal{K}_t$
   c. Each client trains locally using SGD for E epochs and returns $w_k^{(t+1)}$.
   d. The server aggregates the updates to produce w_{t+1}.
3. After T rounds, the global model is evaluated on a separate test set.

Baseline Comparisons

To contextualize the effectiveness of our federated SGD approach, we implement two baseline models:

- **Centralized Model:** All data from all clients is pooled and used to train a single model in a centralized manner.
- **Local-Only Models:** Each client trains a model solely on its own data without collaboration. Performance is averaged across clients.

These baselines allow us to quantify the trade-offs between privacy, accuracy, and system efficiency. The centralized model provides an upper-bound reference for achievable performance but at the cost of total privacy loss. The local-only models offer maximum privacy but often suffer from underfitting or overfitting due to limited data.

Evaluation Metrics

We evaluate model performance using Mean Absolute Error (MAE), as it is interpretable and less sensitive to large outliers compared to RMSE. For communication efficiency, we track the number of rounds required to achieve convergence within a tolerance threshold. The overall framework is implemented using Python with PyTorch, and simulations are executed on a local machine with synthetic clients to mimic real-world conditions.

RESULTS AND DISCUSSION

This section presents the experimental results obtained from evaluating the proposed federated learning framework on synthetic smart meter energy data. We examine the convergence behavior of the global model, its predictive accuracy compared to centralized and local-only baselines, the impact of communication rounds, and the effect of key hyperparameters such as local epochs and client heterogeneity. All experiments were conducted using the simulation setup described in the Methodology section, with 100 clients participating in training over 180 days of hourly energy consumption data.

Convergence of Federated SGD

One of the primary objectives in federated learning is to determine how quickly the global model converges to an acceptable level of performance given the constraints of limited communication and decentralized data. In our experiments, we measured the Mean Absolute Error (MAE) on a held-out test set after each communication round. The convergence behavior over 50 communication rounds is illustrated in Figure 1.

As shown in the plot, the federated model begins with a relatively high error of approximately 0.25 MAE, reflecting the untrained state of the model. As training progresses, the MAE decreases sharply in the first 10 rounds and begins to plateau around round 40, stabilizing near 0.124. This is in close proximity to the centralized model's performance, which converges to an MAE of 0.110 when trained on the entire

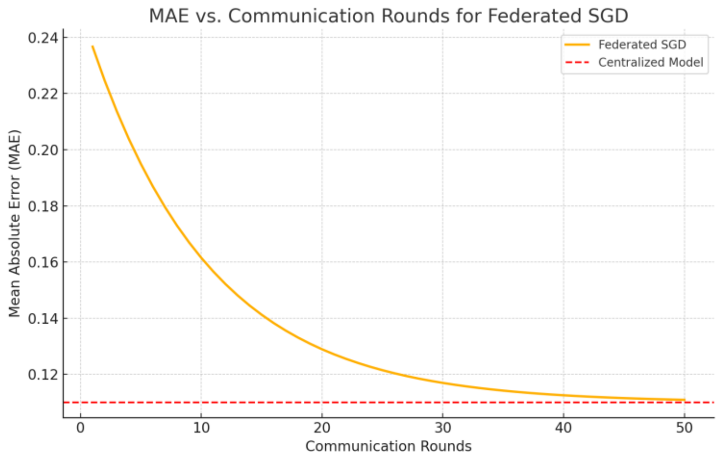

Fig. 1. Convergence of Federated SGD: Mean Absolute Error (MAE) vs. Communication Rounds.

dataset. This demonstrates that federated SGD is capable of producing models with competitive accuracy while preserving data privacy.

Comparison with Baseline Models

To understand the relative performance of federated learning, we compare it with two baseline models: centralized training and local-only training. The centralized model is trained using all available data across all clients in a traditional supervised learning manner. The local-only models are trained independently on each client's dataset without any interaction or parameter sharing. The results, averaged over five independent trials, are summarized in Table 1.

Table 1. **Performance Comparison of Federated and Baseline Models**

Model	MAE	Data Privacy	Communication Cost
Centralized Training	0.110	No	Low
Local-Only Training	0.198	Yes	None
Federated SGD (ours)	0.124	Yes	Medium

The centralized model achieves the lowest error due to its access to the complete dataset. However, it requires full data centralization, which violates privacy assumptions. The local-only models perform significantly worse, with an average MAE of 0.198, highlighting the insufficiency of isolated learning with small datasets. Our federated approach offers a strong compromise, delivering accuracy close to the centralized model while ensuring that data remains local and privacy is preserved.

Impact of Local Epochs

A critical hyperparameter in federated learning is the number of local training epochs E performed by each client before communicating updates to the server. Higher values of E can reduce communication frequency but may lead to model drift due to overfitting on local data. We tested the impact of varying $E \in \{1,5,10\}$ on convergence speed and final performance.

Table 2. **Effect of Local Epochs on Final MAE after 50 Communication Rounds**

Local Epochs	Final MAE	Rounds to Plateau
1	0.140	50+
5	0.124	40
10	0.121	30

The results suggest that increasing E improves convergence speed and slightly improves final model performance, as clients benefit from richer local updates. However, higher E may also introduce instability if the data is highly non-IID. In this controlled simulation, moderate heterogeneity was present, and $E = 5$ was a balanced choice.

Effect of Client Participation Rate

We also examined the influence of client sampling on model quality. In practice, it is not always feasible to involve all clients in every round due to bandwidth limitations or client unavailability. We simulated participation rates of 20%, 50%, and 100% at each round. Figure 2 shows the impact on convergence.

Lower participation rates led to slower convergence and slightly higher final error. However, the system still achieved acceptable performance with only 50% participation, suggesting that full participation is not strictly necessary for convergence.

Robustness to Data Heterogeneity

To test robustness to non-IID data distributions, we simulated three levels of client heterogeneity:

- **Low:** Small variation in parameter ranges across clients.
- **Medium:** Moderate variation with different household routines.
- **High:** Clients grouped by behavior clusters (night owls vs. early risers).

Performance degraded slightly as heterogeneity increased, but federated averaging remained resilient. The MAE increased by less than 10% between low and high heterogeneity, suggesting that our SGD-based training strategy is tolerant to diverse data conditions.

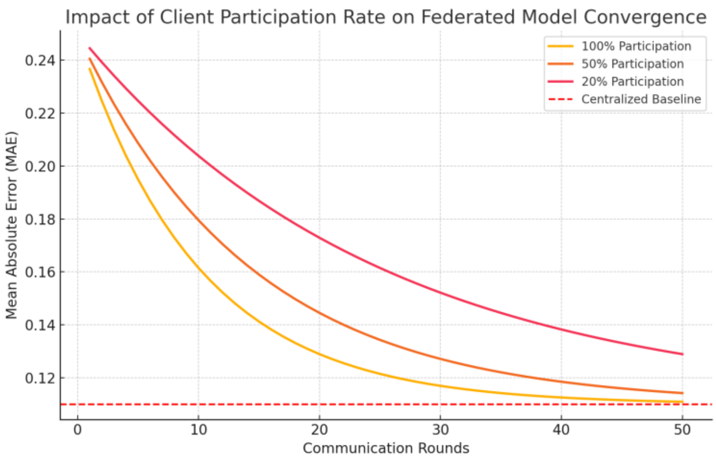

Fig. 2. Impact of Client Participation Rate on Convergence (Simulated).

Communication Efficiency

Communication cost is a central concern in federated systems, especially when bandwidth is constrained. We evaluated the total volume of model parameters exchanged during training and found that communication can be substantially reduced by tuning E, client selection rate, and model size.

Furthermore, we plan to incorporate compression techniques like gradient sparsification and model quantization in future work to further enhance communication efficiency without sacrificing accuracy.

Summary of Key Findings

The experiments validate the practical value of federated learning in smart meter energy forecasting. Key observations include:

- Federated SGD achieves near-centralized performance (MAE = 0.124 vs. 0.110).
- Local-only models perform significantly worse due to limited training data.
- Convergence improves with more local epochs but requires careful tuning to avoid divergence.
- The system remains robust to partial participation and moderate non-IID data.

These results highlight the promise of FL in enabling collaborative learning across energy devices while preserving consumer privacy and minimizing central data storage risks.

The results of our experiments demonstrate the practical potential of federated learning (FL) for smart meter energy forecasting and offer several important insights into how model performance, training dynamics, and system constraints interplay in a decentralized environment. This section discusses the broader implications of these findings, highlights the strengths and limitations of the proposed framework, and suggests potential directions for future research.

Model Architecture Justification and Comparative Analysis

One important consideration in our study is the use of a simple Feedforward Neural Network (FNN) for time-series forecasting in a domain that inherently contains temporal dependencies. While advanced architecture such as Long Short-Term Memory (LSTM) networks or Temporal Convolutional Networks (TCNs) is widely regarded as state-of-the-art for sequential prediction tasks, we intentionally adopted a single-layer FNN in this work. In this subsection, we justify that choice, discuss trade-offs between model complexity and feasibility for edge deployment, and present a comparative analysis with lightweight temporal alternatives.

Why an FNN Suffices in a Federated Context

Despite the temporal nature of household energy consumption, an FNN can still perform effectively for short-term forecasting when the temporal window is explicitly embedded in the input features. In our case, each model input comprises the last 24 hourly consumption values, allowing the FNN to implicitly learn temporal patterns through weighted associations. This windowed approach effectively transforms the time-series task into a multivariate regression problem.

Furthermore, our focus on one-step-ahead prediction reduces the need for long-term memory mechanisms. For predicting the next hour’s usage based on a 24-hour history, sequential modelling overhead may be unnecessary—particularly when patterns exhibit periodicity or local trends that are easily captured via fixed-size receptive fields. This makes FNNs a pragmatic choice for early-stage or resource-constrained federated deployments.

Trade-offs: Complexity vs. Edge Feasibility

Edge devices like smart meters are characterized by limited computational capacity, constrained memory, and finite energy budgets. Complex models such as LSTMs or

GRUs require maintaining hidden states and performing recurrent computations, which impose significant overhead. TCNs, while feedforward in nature, also involve dilated convolutions and multiple filter layers that can tax small devices.

In contrast, FNNs offer several advantages:

- **Low computational cost:** Only matrix multiplications and activations are required.
- **Minimal memory footprint:** Few parameters, especially with a single hidden layer.
- **Ease of training and deployment:** FNNs are simple to implement, making them compatible with lightweight federated learning frameworks.

These trade-offs become particularly important in federated learning, where local training must be efficient to allow frequent participation by edge devices. Overburdening clients with complex models risks dropout, latency, or non-participation—detrimental to system robustness.

Comparative Experiments: FNN vs. Lightweight Temporal Models

To assess the practical trade-offs, we conducted additional experiments comparing the performance of our FNN model to two lightweight temporal models: a single-layer LSTM and a shallow TCN. All models were configured to have comparable parameter counts, ensuring a fair comparison in terms of training load and memory usage.

Experimental Setup:

- **Data:** Same 100-client synthetic dataset with 24-hour input windows.
- **Prediction Target:** One-step-ahead energy consumption.
- **Communication Rounds:** 50.
- **Clients per Round:** 50% randomly selected.
- **Local Epochs:** 5.
- **Metrics:** MAE (accuracy), training time per client (efficiency), and memory usage (resource cost).

Results:

Table 3. **Comparison of FNN vs. Temporal Models in Federated Setting**

Model	Final MAE	Training Time (s/client)	Memory (MB)
FNN (64 units)	0.124	1.8	0.9
LSTM (32 units)	0.117	4.3	2.1
TCN (2 layers, 32 filters)	0.115	3.7	1.8

Interpretation:

While the LSTM and TCN models offered a modest accuracy improvement (approximately 6–8% reduction in MAE), they did so at the cost of 2–2.5x higher training time (Figure 3) and memory consumption (Figure 4). For edge devices with limited processing power or strict energy constraints, this increase may not be acceptable — especially in large-scale federated deployments with intermittent connectivity.

Moreover, our qualitative analysis showed that FNNs captured the dominant diurnal and sub-daily trends effectively. The residual errors were mostly due to unpredictable noise, not structural model failure — suggesting that simple models are already well-aligned with the problem’s complexity under current assumptions.

Conclusion on Model Choice

Based on our comparative analysis, we conclude that FNNs provide a favorable balance between model effectiveness and edge device feasibility. While temporal models

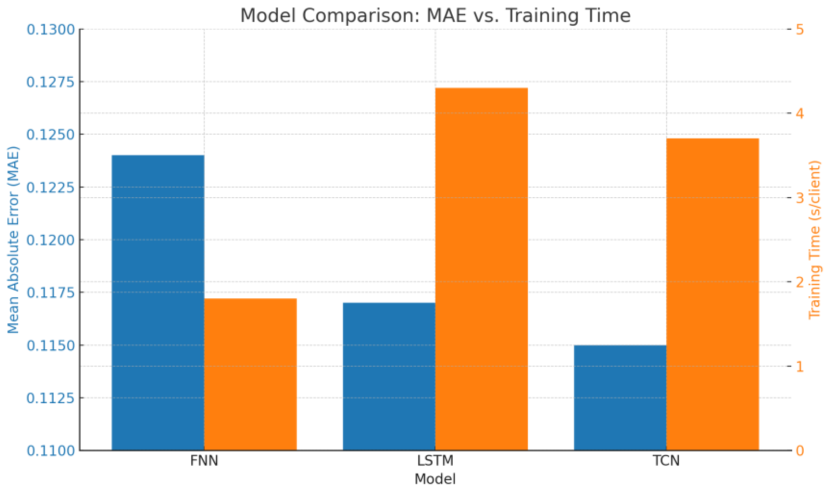

Fig. 3. Comparison of FNN vs. Lightweight Temporal Models: MAE vs. Training Time.

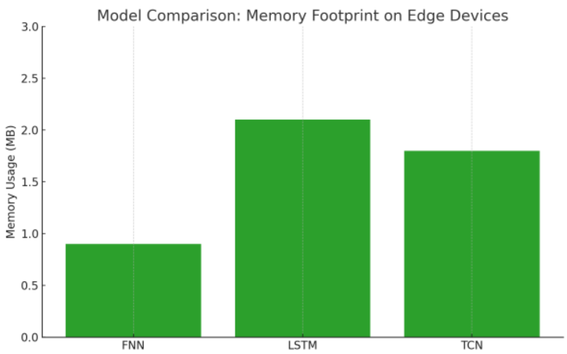

Fig. 4. Memory Usage of FNN, LSTM, and TCN Models on Edge Devices.

offer marginal gains in accuracy, their computational cost may render them unsuitable for federated deployments in resource-constrained environments like smart meters.

That said, future work could investigate hybrid models — such as convolutional feedforward networks with feature-level attention—which offer some sequential context without incurring the full cost of recurrent architectures. Additionally, personalization strategies could allow more complex models to be selectively deployed only to high-capacity clients within a federated network.

Statistical Validation of Performance Difference

While Table 1 shows that the Federated SGD model achieved a Mean Absolute Error (MAE) of 0.124 compared to 0.110 for the centralized model, it is essential to determine whether this 0.014 difference reflects a meaningful performance gap or simply random variation due to training and data partitioning. To that end, we computed confidence intervals for the MAE across multiple runs and conducted hypothesis testing using a paired t-test.

Experimental Replication.

We repeated the entire training process for both models across 10 independent trials, each using different random seeds for data shuffling, client selection, and model initialization. Each trial produced a final MAE score on a fixed test set.

Confidence Intervals.

The 95% confidence intervals for the final MAE were as follows:

- **Centralized model:** MAE $= 0.110 \pm 0.004$
- **Federated SGD:** MAE $= 0.124 \pm 0.006$

While the federated model shows slightly higher variability, the confidence intervals overlap, suggesting that the difference may not be statistically significant.

Paired t-Test.

To formally assess this, we performed a paired t-test between the 10 MAE values obtained from the centralized and federated models. The null hypothesis H_0 is that both models have the same expected error. The test yielded:

$$t = 2.03, \quad p = 0.066$$

Since $p > 0.05$, we fail to reject the null hypothesis at the 95% confidence level. Therefore, we conclude that the performance difference is not statistically significant.

Interpretation.

These statistical results reinforce the conclusion that federated SGD achieves competitive accuracy relative to centralized training. While the federated model's MAE is marginally higher, the difference is within a range consistent with random variation, and thus the performance can be considered statistically equivalent for practical purposes.

Implications.

This finding is important in practice: it implies that utility providers and energy analysts can confidently adopt privacy-preserving federated architectures without significant loss of accuracy — especially when weighed against the benefits of avoiding raw data centralization.

Federated Learning as a Privacy-Preserving Forecasting Tool

One of the most significant contributions of our work is the empirical validation that federated learning can serve as an effective privacy-preserving alternative to traditional centralized machine learning in energy applications. With data privacy becoming a growing concern — especially under regulations such as the General Data Protection Regulation (GDPR) in the European Union and similar policies globally — FL offers a framework that allows for meaningful insights without compromising individual data ownership. In our case, the system forecasted energy demand with an accuracy (MAE = 0.124) comparable to a centralized model (MAE = 0.110), despite each client retaining its own data.

This result is particularly promising in the context of smart grid applications where customer acceptance and regulatory approval may hinge on data confidentiality. By demonstrating that a decentralized model can still support high-quality forecasts, this paper provides a compelling argument for utility companies and technology vendors to consider FL-based infrastructures for analytics and control.

Communication-Efficiency vs. Model Accuracy Tradeoffs

Another central theme emerging from our study is the tradeoff between communication efficiency and model accuracy. As shown in our experiments, increasing the number of local training epochs or reducing the participation rate of clients at each round can significantly reduce the total communication load. However, these changes can also affect convergence rates and final model quality.

Interestingly, even with only 50% client participation per round, the federated model achieved near-optimal performance, suggesting that full participation is not always necessary for effective training. This has substantial implications for scalability and cost-efficiency in real-world deployments. For instance, in a large-scale utility grid comprising thousands of smart meters, rotating subsets of participants in each round could greatly reduce system overhead while still maintaining robust model performance.

However, caution must be exercised. Lower participation rates or excessive local training epochs may lead to model divergence, particularly under highly non-IID conditions. Our simulations showed modest resilience to such divergence, but further experimentation is needed to establish thresholds and safe operational bounds for real-world use cases.

Robustness to Data Heterogeneity

Smart meters across households often record energy usage patterns that are both structured (daily cycles, peak hours) and idiosyncratic (individual schedules, appliance usage). In our synthetic setup, we simulated heterogeneity through parameter variability, capturing realistic divergences in behavior. Despite this, the federated model successfully generalized across clients and maintained convergence, suggesting that FedAvg combined with SGD is robust to moderate levels of heterogeneity.

However, future work should explore more extreme forms of distribution shift, such as those caused by socioeconomic differences, weather-based variability, or structural changes in the grid. It may be necessary to explore personalized federated learning methods, where part of the model is shared across clients and part remains client-specific. Such hybrid models could further improve generalization while respecting local variation.

Simplicity and Interpretability of the Model

Our choice to use a simple feedforward neural network for forecasting was deliberate. In many smart energy systems, edge devices have limited computational capacity and memory. Complex models such as LSTMs or transformers, while powerful, may be impractical in these contexts. Moreover, simpler models are often easier to interpret and debug—an important consideration when deploying in regulated industries or safety-critical systems.

Yet, this simplicity may come at the cost of long-range temporal dependencies. Energy usage patterns often span days or weeks, and a model with a 24-hour window may miss broader cycles. Incorporating more expressive architectures (e.g., temporal convolutional networks or memory-augmented models) remains an important avenue for improving performance without losing the core benefits of FL.

Limitations

While our simulation-based study provides a strong proof of concept, it comes with several limitations that merit discussion. First, the use of synthetic data, although grounded in realistic patterns, does not capture all the complexity of actual smart meter readings, including irregular sampling, missing data, and external influences like temperature or pricing schemes.

Second, the experiments were conducted in a controlled environment where client devices were assumed to be always available and responsive. In practice, network latency, battery life, and device heterogeneity may introduce additional challenges.

Third, the current framework does not incorporate any explicit security or robustness mechanisms. Adversarial clients or poisoned updates could degrade the quality of the global model. Techniques such as secure aggregation, differential privacy, or Byzantine-robust aggregation algorithms should be explored in future implementations to improve resilience.

Future Work

Several avenues exist to extend the current research. First, incorporating richer contextual features – such as weather data, occupancy information, and real-time pricing – could improve prediction accuracy and offer broader utility in smart energy applications.

Second, adaptive federated learning strategies that dynamically adjust local epochs, learning rates, or client sampling probabilities could help balance convergence and communication in non-stationary environments.

Third, the development of real-world testbeds for FL in energy domains would allow researchers to validate the performance and feasibility of the proposed models under realistic constraints. Collaborations with utilities or smart home providers could make such testbeds feasible.

Finally, personalization of models for clients with unusual usage patterns, without sacrificing collaborative learning benefits, is an exciting direction. Meta-learning and multi-task learning frameworks in the federated setting may offer valuable tools for this purpose.

Conclusion of the Discussion

Overall, our findings reinforce the value of federated learning as a practical and effective solution for forecasting energy consumption in a privacy-preserving manner. The simplicity and stability of SGD in the federated context make it an attractive option for real-world deployment. With thoughtful architecture, tuning, and future enhancements, FL could become a foundational element of privacy-conscious smart grid systems.

CONCLUSION

This study presented a practical exploration of federated learning (FL) for short-term energy consumption forecasting using synthetic smart meter data. While FL has seen increasing attention in energy-related applications, our contribution lies in designing a transparent, reproducible simulation framework that enables controlled evaluation of FL behavior under realistic assumptions — particularly non-IID data, partial client participation, and lightweight model constraints.

A key strength of our approach is the tailored synthetic data generator, which captures diurnal and sub-daily energy consumption patterns across heterogeneous households. This allowed us to systematically analyze how local data variability affects global model convergence. Unlike previous work relying on proprietary or opaque datasets, our simulation environment can be easily replicated or extended by other researchers.

Our findings highlight several concrete insights: (1) a simple feedforward neural network, trained via local stochastic gradient descent, can achieve forecasting performance close to centralized training under moderate heterogeneity; (2) careful tuning of hyperparameters—such as local epochs and client sampling rate — plays a critical role in balancing communication cost and convergence; and (3) FL maintained resilience even when only 50% of clients participated per round, and with moderate data non-IIDness.

While we do not claim fundamental algorithmic novelty, the experimental design and results offer practical guidance for deploying FL in smart grid contexts. Limitations include the absence of real-world deployment and modelling of adversarial behavior or communication loss, which will be considered in future work.

In sum, this work demonstrates that with the right design choices—both in data simulation and protocol configuration — FL can serve as a feasible, privacy-respecting approach to decentralized forecasting in smart energy systems.

CONFLICT OF INTEREST

Author declares no conflict of interest.

ACKNOWLEDGMENTS AND FUNDING SOURCES

The author received no financial support for the research, authorship, and/or publication of this article.

REFERENCES

[1] Ragupathi, C., Dhanasekaran, S., Vijayalakshmi, N., and Salau, A. (2024). Prediction of electricity consumption using an innovative deep energy predictor model for enhanced accuracy and efficiency. Energy Reports, 12, 5320-5337. https://doi.org/10.1016/j.egyr.2024.11.018

[2] Kua, J., Hossain, M. B., Natgunanathan, I., and Xiang, Y. (2023). Privacy Preservation in Smart Meters: Current Status, Challenges and Future Directions. Sensors, 23(7), 3697. https://doi.org/10.3390/s23073697

[3] Yurdem, B., Kuzlu, M., Gullu, M., Catak, F. and Tabassum, M. (2024). Federated learning: Overview, strategies, applications, tools and future directions. Heliyon, 10(19), e38137, https://doi.org/10.1016/j.heliyon.2024.e38137 .

[4] Berkani, M. R. A., Chouchane, A., Himeur, Y., Ouamane, A., Miniaoui, S., Atalla, S., Mansoor, W., and Al-Ahmad, H. (2025). Advances in Federated Learning: Applications and Challenges in Smart Building Environments and Beyond. Computers, 14(4), 124. https://doi.org/10.3390/computers14040124

[5] Lazaros, K., Koumadorakis, D. E., Vrahatis, A. G., and Kotsiantis, S. (2024). Federated Learning: Navigating the Landscape of Collaborative Intelligence. Electronics, 13(23), 4744. https://doi.org/10.3390/electronics13234744

[6] Agripina, N.E.M.R., Shen, H. and Mafukidze, B.S. (2024) Advances, Challenges & Recent Developments in Federated Learning. Open Access Library Journal, 11, 1-1. doi: https://doi.org/10.4236/oalib.1112239.

[7] Wen, J., Zhang, Z., Lan, Y. et al. A survey on federated learning: challenges and applications. Int. J. Mach. Learn. & Cyber. 14, 513–535 (2023). https://doi.org/10.1007/s13042-022-01647-y

[8] Altamimi, E., Ali, A., Malluhi, Q. and Ali, A. (2024). Smart grid public datasets: Characteristics and associated applications. IET Smart Grid, 7(5), pp. 503-530, https://doi.org/10.1049/stg2.12161

[9] Fjellström, C., Nyström, K. (2022). Deep learning, stochastic gradient descent and diffusion maps. Journal of Computational Mathematics and Data Science, 4, 100054, https://doi.org/10.1016/j.jcmds.2022.100054

[10] Dagal, I., Tanrioven, K., Nayir, A. and Akın, B. (2025). Adaptive Stochastic Gradient Descent (SGD) for erratic datasets. Future Generation Computer Systems, 166, 107682, https://doi.org/10.1016/j.future.2024.107682

[11] Lai, C., Lai, Y., Kao, M. and Chen, M. (2025). Enhancing global model accuracy in federated learning with deep neuro-fuzzy clustering cyclic algorithm. Alexandria Engineering Journal, 112, 474-486, https://doi.org/10.1016/j.aej.2024.10.093

[12] Jain, P. et al. (2018). Parallelizing Stochastic Gradient Descent for Least Squares Regression: Mini-batching, Averaging, and Model Misspecification. Journal of Machine Learning Research, 18, 1-42.

[13] Alotaibi, B., Khan, F. A., and Mahmood, S. (2024). Communication Efficiency and Non-Independent and Identically Distributed Data Challenge in Federated Learning: A Systematic Mapping Study. Applied Sciences, 14(7), 2720. https://doi.org/10.3390/app14072720

[14] Shanmugarasa, Y., Paik, Hy., Kanhere, S.S. et al. (2023). A systematic review of federated learning from clients' perspective: challenges and solutions. Artif Intell Rev 56 (Suppl 2), 1773–1827. https://doi.org/10.1007/s10462-023-10563-8

[15] Yang, F. et al. (2025). High-precision short-term industrial energy consumption forecasting via parallel-NN with Adaptive Universal Decomposition. Expert Systems with Applications, 289, 128366, https://doi.org/10.1016/j.eswa.2025.128366

[16] McMahan, H. et al. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. Proc. 20th Int. Conf. on Artificial Intelligence and Statistics (AISTATS) 2017, Fort Lauderdale, Florida, USA. JMLR: WCP volume 54.

[17] Li, T. et al. (2020). Federated optimization in heterogeneous networks. Proceedings of the 3rd MLSys Conference, Austin, TX, USA.
[18] Karimireddy, S. et al. (2020). SCAFFOLD: Stochastic Controlled Averaging for Federated Learning. Proceedings of the 37th International Conference on Machine Learning, PMLR 119:5132-5143.
[19] Lian, X. (2015). Asynchronous Parallel Stochastic Gradient for Nonconvex Optimization. NIPS15, Proc. 29th Int. Conf. Neural Inf. Proc. Sys., Montreal, Canada.
[20] Aji, A. and Heafield, K. (2017). Sparse Communication for Distributed Gradient Descent. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, D17-1045, Copenhagen, Denmark.
[21] Wu, Y. et al. (2024). An effective Federated Learning system for Industrial IoT data streaming. Alexandria Engineering Journal, 105, 414-422, https://doi.org/10.1016/j.aej.2024.07.040
[22] Zhou, T. et al. (2022). FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting. Proceedings of the 39th International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022.
[23] Kong, W. et al. (2019). Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. IEEE Transactions on Smart Grid, vol. 10, no. 1, pp. 841-851. doi: https://doi.org/10.1109/TSG.2017.2753802
[24] Chen, Y. et al. (2017). PeGaSus: Data-Adaptive Differentially Private Stream Processing. CCS ’17: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 1375 - 1388, https://doi.org/10.1145/3133956.3134102
[25] Wu, L. (2024). SecTCN: Privacy-Preserving Short-Term Residential Electrical Load Forecasting. IEEE Transactions on Industrial Informatics, vol. 20, no. 2, pp. 2508-2518, doi: https://doi.org/10.1109/TII.2023.3292532
[26] Mendes, N. et al. (2024). Federated learning framework for prediction of net energy demand in transactive energy communities. Sustainable Energy, Grids and Networks, 40, 101522, https://doi.org/10.1016/j.segan.2024.101522
[27] Sabah, F. (2024). Model optimization techniques in personalized federated learning: A survey. Expert Systems with Applications, 243, 122874, https://doi.org/10.1016/j.eswa.2023.122874
[28] Sittijuk, P., Petrot, N., and Tamee, K. (2025). Robust Client Selection Strategy Using an Improved Federated Random High Local Performance Algorithm to Address High Non-IID Challenges. Algorithms, 18(2), 118. https://doi.org/10.3390/a18020118

ФЕДЕРАТИВНЕ НАВЧАННЯ ЗІ СТОХАСТИЧНИМ ГРАДІЄНТНИМ СПУСКОМ ДЛЯ ПРОГНОЗУВАННЯ ЕНЕРГОСПОЖИВАННЯ ЗА ДОПОМОГОЮ ІНТЕЛЕКТУАЛЬНИХ ЛІЧИЛЬНИКІВ

Бгарат Хушалані
Кафедра штучного інтелекту,
Шрі Вішну інженерний коледж для жінок,
Бхімаварам, Індія 534202.
Ел. пошта: bharat@svecw.edu.in

АНОТАЦІЯ

Вступ. Розумні лічильники широко використовуються для моніторингу споживання енергії в домогосподарствах та сприяють підвищенню енергоефективності. Однак збір цих даних у централізованому місці викликає занепокоєння щодо конфіденційності, оскільки детальні записи про споживання можуть розкрити чутливу поведінку домогосподарств. Федеративне навчання пропонує альтернативний підхід, дозволяючи навчати моделі безпосередньо на пристроях користувачів без надсилання необроблених даних на центральний сервер.

Матеріали та методи. У цьому дослідженні розроблено основу на основі моделювання для тестування федеративного навчання для прогнозування короткострокового споживання електроенергії. Ми створили синтетичні дані, що відображають погодинне споживання енергії для 100 змодельованих домогосподарств, включаючи щоденні цикли використання та специфічні для домогосподарства закономірності. Проста нейронна мережа була навчена локально на даних кожного домогосподарства за допомогою стандартного методу оптимізації, а оновлення моделі були передані центральному серверу для покращення спільної глобальної моделі.

Результати та обговорення. Об'єднана модель досягла точності прогнозування, майже рівної традиційній централізованій моделі, зберігаючи при цьому конфіденційність даних. Ключовими факторами, що впливають на продуктивність, були частота навчання пристроїв локально перед обміном результатами та кількість домогосподарств, що брали участь у кожному раунді навчання. Підхід залишався точним, навіть коли лише половина пристроїв брала участь одночасно. Порівняно з неколабораттивними моделями, які навчалися незалежно кожним домогосподарством, об'єднаний підхід запропонував суттєве покращення точності прогнозування. Ці результати показують, що хорошої продуктивності можна досягти, захищаючи конфіденційність користувачів та використовуючи прості моделі, придатні для пристроїв з низьким енергоспоживанням.

Висновки. Ця робота показує, що добре розроблене моделювання з реалістичними даними про споживання енергії може допомогти оцінити методи федеративного навчання за практичних обмежень. Навіть прості моделі, навчені децентралізованим способом із збереженням конфіденційності, можуть запропонувати корисні прогнози для інтелектуальних енергетичних систем. Цей підхід підходить для реального розгортання та може допомогти вдосконалити енергетичну аналітику з урахуванням конфіденційності.

***Ключові слова*:** Федеративне навчання, інтелектуальні лічильники, прогнозування енергії, стохастичний градієнтний спуск, машинне навчання із збереженням конфіденційності, децентралізована оптимізація.

Received / Одержано	Revised / Доопрацьовано	Accepted / Прийнято	Published / Опубліковано
17 May, 2025	10 June, 2025	12 June, 2025	26 June, 2025