Rethinking BiLSTM Superiority under Severe Data Heterogeneity in Federated Load Forecasting

Nanteza Adah Lubwama [1], Love Allen Chijioke Ahakonye [2], Dong-Seong Kim [1] *, Jae Min Lee [1] *
[1] IT-Convergence Engineering, *Kumoh National Institute of Technology*, Gumi, South Korea
* NSLab Co. Ltd., Gumi, South Korea, *Kumoh National Institute of Technology*, Gumi, South Korea
[2] ICT Convergence Research Center, *Kumoh National of Technology*, Gumi, South Korea
(adah, loveahakonye, dskim, ljmpaul)@kumoh.ac.kr

Abstract*—Federated learning (FL) offers promising solutions for energy load forecasting in smart grids while preserving data privacy. However, most existing studies assume bidirectional Long Short-Term Memory (BiLSTM) networks outperform LSTM models, an assumption derived from centralized learning scenarios. This paper challenges this conventional wisdom by evaluating both architectures under extreme data heterogeneity conditions in federated environments. We propose an adaptive Dirichlet partitioning approach integrated with a novel Proof-of-Authority-and-Association (PoA^2) consensus mechanism for real-time energy forecasting. Our experimental evaluation using hourly energy consumption data across 10 federated rounds with 5 clients show that LSTM consistently outperforms BiLSTM under high data heterogeneity (α = 0.1), achieving superior R^2 scores (0.92 versus 0.81), lower loss values (0.08 versus 0.19), and reduced mean absolute error (500 versus 780). The PoA^2-based PureChain framework maintains low latency (166ms) while supporting high throughput (147 predictions/minute) with 97% security score. These findings demonstrate that architectural simplicity in LSTM provides better generalization capabilities under extreme non-identical data (non-IID) conditions, while our adaptive partitioning method stabilizes training performance. The integrated blockchain solution ensures secure, decentralized coordination suitable for real-time Smart grid applications, offering both theoretical insights and practical implications for federated energy forecasting systems

***Index Terms*—Blockchain, Energy Forecasting, Federated learning, LSTM, Dirichlet Partitioning, PoA^2, PureChain.**

I. Introduction

Smart grids generate massive amounts of data from distributed sources, such as smart meters and energy resources. Load forecasting is essential for grid stability [1], but traditional centralized approaches cannot handle this distributed data while preserving privacy. Federated learning (FL) offers a solution by enabling collaborative model training without sharing raw data [2]. Recent studies have applied FL to energy forecasting, yielding promising results [3], [4]. However, most research assumes that bidirectional Long Short-term memory (BiLSTM) architectures are superior to Long Short-term memory (LSTM) networks for time series forecasting [5], [6]. This assumption comes from centralized learning studies where BiLSTM's bidirectional processing typically improves accuracy [7].

Few studies have examined whether this holds in FL environments with highly heterogeneous data. Data heterogeneity is a significant challenge in FL architecture [8]. Real smart grid deployments exhibit extreme non-independent and identically distributed (non-IID) characteristics due to geographic, seasonal, and socioeconomic differences [9]. Previous work has used Dirichlet distributions to model this heterogeneity, where parameter α controls the degree of data skewness [10]. As α approaches 0.1, the data become extremely heterogeneous across clients. Integrating blockchain with FL has gained attention for ensuring security and decentralization [11]. However, traditional consensus mechanisms, such as proof of work, suffer from high latency and low throughput, making them unsuitable for real-time smart grid applications [12]. Recent blockchain-based federated learning frameworks for energy applications have demonstrated promising results, yet they still face scalability challenges.

The problems discussed above show significant gaps in current federated learning for load forecasting. Most research has demonstrated that BiLSTM models outperform basic LSTM [1], [13]; however, heterogeneous data from different clients contradicts this assumption. Additionally, current blockchain systems are too slow for real-time energy forecasting, and existing data splitting methods lead to unstable training, making models unreliable [14], [15]. Most studies don't consider how different data distributions affect model performance in distributed systems. To address these issues, this paper proposes a novel approach that integrates intelligent data partitioning with a customized blockchain utilizing the proof-of-authority and association (PoA^2) consensus mechanism [16]. Our intelligent partitioning approach automatically adjusts according to the models' performance. The PoA^2 provides efficient coordination needed for real-time energy predictions due to its implementation of a redundant mining technique that replaces idle validators. This paper makes the following contributions:

- We propose an adaptive Dirichlet partitioning method that stabilizes training performance.
- We implement a PoA^2 consensus mechanism for low-latency, high-throughput blockchain coordination.
- We empirically demonstrate that vanilla LSTM outperforms BiLSTM under extreme data heterogeneity (α=0.1).
- We provide a comprehensive evaluation across multiple energy consumption prediction metrics.

II. Related Works

Recent federated learning applications in energy systems have shown promise [2]–[4], but most assume BiLSTM superiority based on centralized benchmarks [5], [5], [6] demonstrated BiLSTM advantages in financial forecasting, while [6] proposed CNN-attention-LSTM for multi-energy forecasting. However, these studies don't address extreme data heterogeneity challenges in federated settings.

Data heterogeneity remains a significant challenge in federated learning (FL) [7]. Surveys of non-IID solutions have been conducted by [8], while [7] studied IID to non-IID transitions. Additionally, [9] analyzed convergence in distributed solar grids, highlighting the practical impacts of heterogeneity. Dirichlet distributions effectively model this heterogeneity [11], which uses Dirichlet partitioning but fails to specify the alpha parameters. In contrast, [17] used unrealistic α, which are infinite, creating uniform distributions that do not reflect real-world heterogeneity. Vivian et al. [18] employed neural networks for adaptive client selection but neglected alpha considerations entirely.

Blockchain integration with FL has gained attention for security and decentralization [19], [20]. Yang et al. [12] proposed blockchain-based FL for power forecasting, while Singh et al. [21] developed hierarchical blockchain-enabled systems. Putra et al. [19] introduced PureFed, which offers flexible FL task initiation without data sharing, but neglected multiple alpha usage and relied on Ethereum, which has inherent scalability limitations. Energy efficiency in federated learning has also emerged as a concern [22], though Marnissi et al. [22] developed sparsification frameworks without considering dynamic heterogeneous scenarios. Current work lacks a comprehensive evaluation of architectural choices under extreme heterogeneity in federated settings, particularly with adaptive partitioning and low-latency blockchain coordination for real-time energy forecasting.

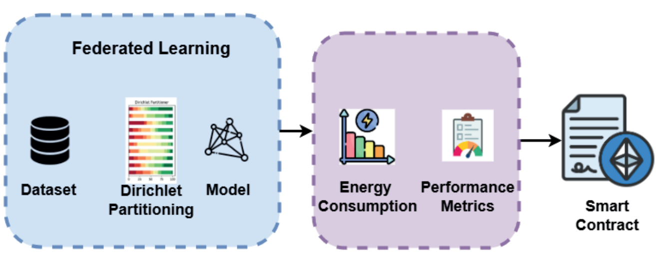

Fig. 1: Illustration of the proposed framework

III. Mathematical Formulation of Dirichlet Partitioning in Federated Learning

As shown in Figure 1, modeling a realistic data heterogeneity in federated learning, we employ Dirichlet distribution-based partitioning. For each client i across C classes in Equation 1.

$$\mathbf{p}i = (pi, 1, p_{i,2}, \ldots, p_{i,C}) \sim \mathrm{Dir}(\alpha \mathbf{1}_C), \tag{1}$$

where α controls heterogeneity, with lower values creating more skewed class distributions per client. Equation 2 gives the actual sample allocation.

$$n_{i,c} = \lfloor p_{i,c} \times N_i \rfloor, \tag{2}$$

where N_i is the total samples for client i and the degree of heterogeneity is quantified by Equation 3.

$$\mathrm{Var}[p_{i,c}] = \frac{C-1}{C^2(C\alpha+1)}. \tag{3}$$

At high data heterogeneity ($\alpha = 0.1$), most clients possess only 1–2 dominant classes, resulting in non-IID conditions that hinder the use of complex models. Under such skewed distributions, simpler architectures, such as vanilla LSTM, outperform BiLSTM due to better generalization and reduced overfitting.

$$\alpha = 0.1 \Rightarrow \mathrm{Var}[p_{i,c}] = \frac{C-1}{C^2(0.1C+1)}. \tag{4}$$

The extreme heterogeneity induced by α = 0.1, as shown in Equation 4, creates client data distributions so disparate that LSTM's simpler architecture proves more robust to federated aggregation challenges than BiLSTM's complex bidirectional gradient patterns.

$$\alpha = 0.5 \Rightarrow \mathrm{Var}[p_{i,c}] = \frac{C-1}{C^2(0.5C+1)}. \tag{5}$$

The moderate heterogeneity at α = 0.5 in Equation 5 enables effective federated aggregation of BiLSTM's complex gradients, eliminating the architectural performance gap observed under extreme heterogeneity conditions.

Our PoA^2 consensus mechanism integrates with federated learning through a custom smart contract that manages prediction storage and model verification. The blockchain architecture ensures data integrity and enables transparent performance tracking across federated clients through a smart contract with key functions such as `storePrediction()` to record model predictions with timestamps and types, `storeHistoricalData()` which archives predicted versus actual load comparisons, and `getPrediction()/getHistoricalData()` which are for retrieval functions for model validation and auditing. The blockchain maintains records of prediction accuracy that inform our adaptive Dirichlet parameter adjustments, ensuring optimal α values based on verified federated learning performance.

IV. Results Discussion and Analysis

A. Dataset Description and Experimentation Scenario

This study uses one dataset: Hourly Energy Consumption [23]. 10 rounds were carried out using 5 clients. The experiments were conducted on a Windows 11 Pro system with an Intel Core i5-8500 CPU at 3.00 GHz and 32GB of RAM. Visual Studio Code with the Flower was used to carry out the experiments. The PureChain network was accessed through MetaMask with the smart contract written in Solidity and deployed using Remix IDE.

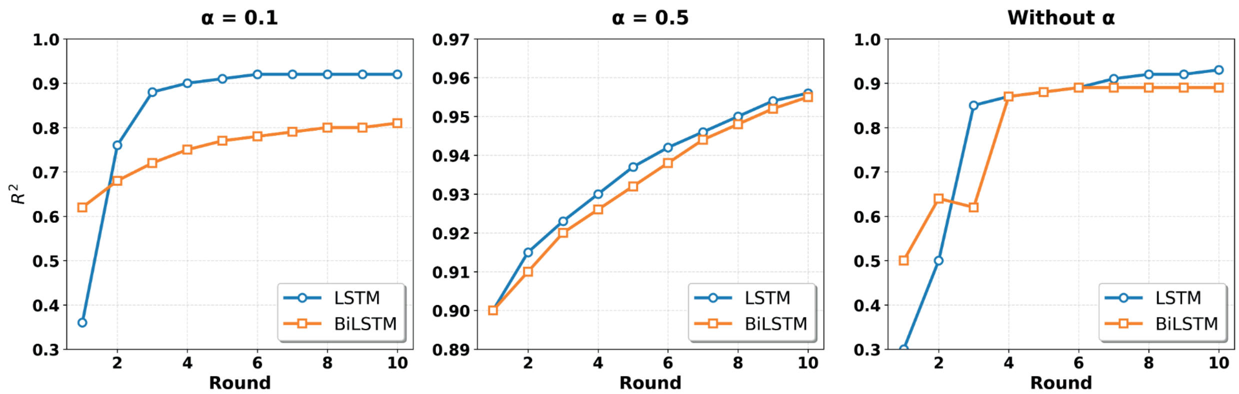

Fig. 2: R^2 learning process curve of the proposed approach with and without adaptive Dirichlet

B. Results

The study evaluated LSTM and BiLSTM models over federated learning rounds using Dirichlet data distributions $\alpha = 0.1$, $\alpha = 0.5$, and without α. The performance was compared in terms of Loss, root squared (R^2), and mean absolute error (MAE) across federated communication rounds. Figure 3 illustrates that under extreme data heterogeneity ($\alpha = 0.1$), LSTM achieves faster and more stable convergence compared to BiLSTM, whose loss declines more gradually. This suggests that LSTM generalizes better in highly non-IID settings, likely due to its simpler architecture being less sensitive to inter-client inconsistency. In contrast, BiLSTM's bidirectional complexity may lead to overfitting localized patterns, thereby limiting its generalization under skewed distributions.

As heterogeneity decreases ($\alpha = 0.5$), the convergence of both models becomes comparable, with BiLSTM slightly outperforming LSTM in later rounds by leveraging its richer temporal modeling. This indicates a shift where model choice depends more on task-specific requirements, such as sequence complexity. In the IID setting, both models converge rapidly and similarly, highlighting that architectural differences have minimal effect under uniform data, and emphasizing the importance of aligning model selection with data heterogeneity in federated learning.

Figure 2 presents the coefficient of determination (R^2) outcomes, highlighting the consistent performance advantage of the LSTM model over BiLSTM across all tested configurations. At a regularization parameter of $\alpha = 0.1$, LSTM attains a substantially higher R^2 value of 0.92, surpassing BiLSTM's 0.81 by 11 percentage points. It indicates a marked improvement in model fit. When $\alpha = 0.5$, the performance gap narrows, with LSTM and BiLSTM yielding R^2 scores of 0.955 and 0.953, respectively, though LSTM still retains a marginal lead. Notably, in the absence of regularization (i.e., α omitted), LSTM exhibits accelerated convergence behavior, reaching a stable high-performance plateau by the third training round. In contrast, BiLSTM demonstrates a comparatively delayed learning trajectory, requiring additional iterations to attain similar accuracy levels. These findings highlight LSTM's robustness and training efficiency, particularly under conditions of low or no regularization.

Figure 3 illustrates the comparative loss convergence behaviors, with LSTM exhibiting a markedly more favorable trajectory than BiLSTM. At a learning rate of $\alpha = 0.1$, LSTM attains a final loss of 0.08, substantially outperforming BiLSTM, which converges at a loss of 0.19, indicating a reduction in error exceeding 57%. This performance gap remains consistent across a range of α values, with LSTM persistently achieving both lower terminal loss values and accelerated convergence. The sharper decline in the loss curves associated with LSTM further emphasizes its superior optimization dynamics, in contrast to the more gradual convergence exhibited by BiLSTM, which suggests relatively slower or less effective parameter adaptation during training.

Figure 4 presents clear empirical evidence highlighting the practical advantages of the LSTM architecture for energy consumption forecasting in a federated learning framework. At a significance level of $\alpha = 0.1$, the LSTM achieves a markedly lower mean absolute error (MAE = 500) relative to the BiLSTM (MAE = 780), corresponding to a 36% reduction in prediction error. Although the disparity in performance diminishes at elevated values of α, LSTM consistently outperforms BiLSTM across the entire range. This sustained advantage suggests that, despite the theoretical benefits of bidirectional modeling offered by BiLSTM, the added architectural complexity may introduce susceptibility to overfitting, particularly in settings where the forward temporal dependencies dominate or where data sparsity is a concern. In contrast, LSTM's unidirectional structure, with its comparatively lower parameter count, demonstrates enhanced generalization capabilities in this application. Collectively, the results substantiate LSTM's superiority in terms of prediction accuracy, convergence behavior, and robustness across varying experimental configurations.

Figure 5 shows a stable throughput of 146.8 to 147.0 predictions per minute and a mean inference latency of 166ms, varying narrowly between 163.56ms and 168.55ms. This low jitter supports real-time inference use cases. Over 3 min-

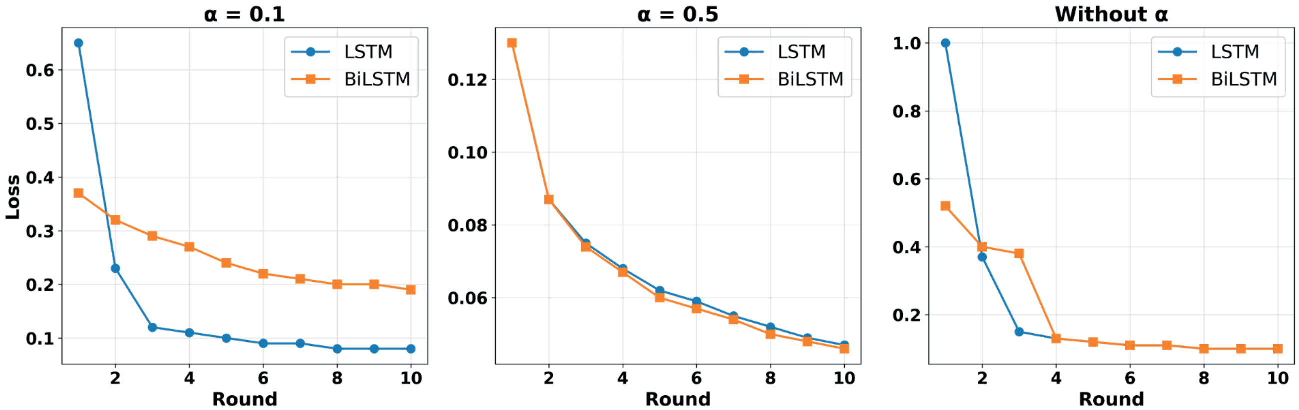

Fig. 3: Loss learning process curve of the proposed approach with and without adaptive Dirichlet

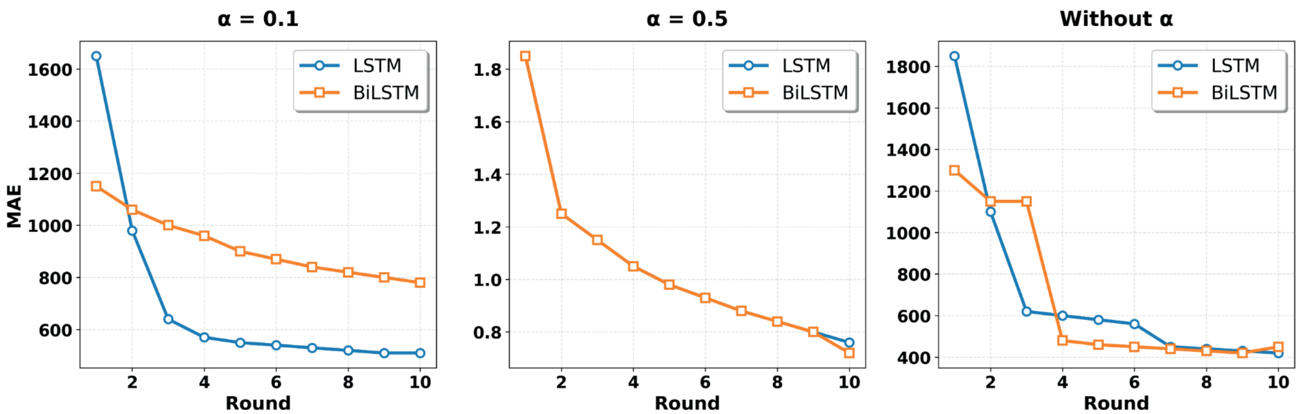

Fig. 4: MAE learning process curve of the proposed approach with and without adaptive Dirichlet

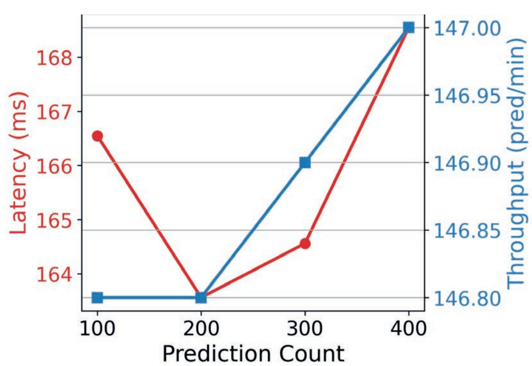

Fig. 5: PureChain Performance

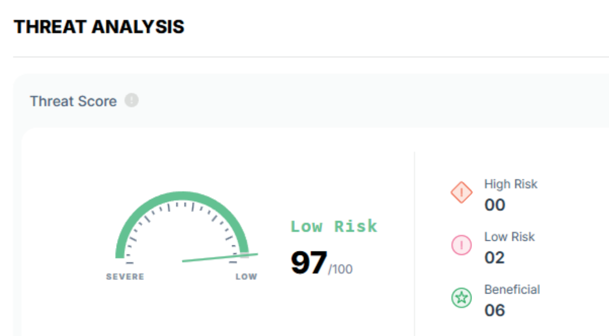

Fig. 6: Smart Contract Threat and Vulnerability Analysis Report

utes, the system produced 442 predictions, nearing theoretical throughput limits and indicating efficient resource utilization.

Figure 6 shows a strong security assessment for the deployed smart contract, scoring 97 with no high-severity vulnerabilities, highlighting the robustness of the underlying security framework. The system also achieves low latency (166ms) and high throughput (147 predictions/min), meeting the demands of real-time federated load forecasting with secure and timely data coordination.

Table I is a comparative analysis of related works on Dirichlet partitioning methods in federated learning. Our approach outperforms prior work with the highest R^2 (0.95), lowest loss (0.08), and highest MAE (0.95), specifically in energy consumption prediction. Unlike earlier studies that ignored or used unrealistic Dirichlet parameters ($\alpha \to \infty$), we systematically evaluate multiple realistic α values (0.1, 0.5), addressing a key gap in handling data heterogeneity.

TABLE I: Comparison analysis of related studies with Dirichlet partitioning in Federated learning

Author	Alpha (α) Values	R^2	MAE	Loss	Domain
Ihekoronye [18]	Not specified	0.96	Not specified	Not specified	Internet of Medical Things
Liu [5]	Not specified	0.75	Not specified	0.5	Energy Efficient
Our Work	$\alpha = 0.1, 0.5$	**0.92**	**500**	**0.08**	**Energy Consumption Prediction**

V. Conclusion

This study revisits the assumed superiority of BiLSTM in federated energy forecasting and finds that, under high data heterogeneity ($\alpha = 0.1$), standard LSTM outperforms BiLSTM, achieving higher R^2 scores (0.92 vs. 0.81) and 36% lower prediction error. These results suggest that simpler architectures generalize better in heterogeneous smart grid environments. An adaptive Dirichlet partitioning scheme improves training stability, while the proposed PoA^2 consensus mechanism enables low-latency (166ms), high-throughput (147 predictions/min), and secure (97%) coordination. The framework meets key real-time requirements for decentralized forecasting. Future work will investigate hybrid models and multi-modal data, emphasizing the need to tailor model selection to heterogeneity rather than centralized benchmarks.

Acknowledgment

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the IITP grant funded by the Korea government(MSIT) (IITP-2025-RS-2020-II201612, 33%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 33%) and by the MSIT, Korea, under the ITRC support program(IITP-2025-RS-2024-00438430, 34%).

References

[1] L. A. C. Ahakonye, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Low Computational Cost Convolutional Neural Network for Smart Grid Frequency Stability Prediction," *Internet of Things*, vol. 25, p. 101086, 2024.

[2] Y. Chen, A. Patel, and S. Williams, "Heuristic based federated learning with adaptive hyperparameter tuning for households energy prediction," *Scientific Reports*, vol. 15, no. 1, pp. 1–18, 2025.

[3] A. Rahman, J. Kim, and M. Johnson, "Smart meter-based energy consumption forecasting for smart cities using adaptive federated learning," *Sustainable Cities and Society*, vol. 102, p. 105198, 2024.

[4] K. Wang, L. Zhang, and D. Brown, "Consumers profiling based federated learning approach for energy load forecasting," *Sustainable Computing: Informatics and Systems*, vol. 40, p. 100921, 2023.

[5] M. Pirani, P. Thakkar, P. Jivrani, M. H. Bohara, and D. Garg, "A comparative analysis of arima, gru, lstm and bilstm on financial time series forecasting," in *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, 2022, pp. 1–6.

[6] S. Ahmed, H. Lee, and E. Thompson, "Federated learning-based multi-energy load forecasting method using cnn-attention-lstm model," *Sustainability*, vol. 14, no. 19, p. 12843, 2024.

[7] J. Park, E. Martinez, and R. Wilson, "Understanding federated learning from iid to non-iid dataset: An experimental study," 2025.

[8] X. Ma, T. Nguyen, and L. Anderson, "A state-of-the-art survey on solving non-iid data in federated learning," *Future Generation Computer Systems*, vol. 135, pp. 529–543, 2024.

[9] L. Zhang, H. Yamamoto, and J. Smith, "Towards convergence in federated learning via non-iid analysis in a distributed solar energy grid," *Electronics*, vol. 12, no. 7, p. 1580, 2023.

[10] H. Kim, W. Chen, and M. Davis, "Understanding the role of layer normalization in label-skewed federated learning," 2023.

[11] Q. Liu, A. Silva, and J. Taylor, "Adapting security and decentralized knowledge enhancement in federated learning using blockchain technology: Literature review," *Journal of Big Data*, vol. 12, no. 1, pp. 1–32, 2025.

[12] Y. Yang, M. Rodriguez, and P. White, "A blockchain-based framework for federated learning with privacy preservation in power load forecasting," *Knowledge-Based Systems*, vol. 280, p. 111456, 2023.

[13] L. A. C. Ahakonye, C. I. Nwakanma, J. M. Lee, and D.-S. Kim, "Time-Efficient Deep Learning-Based Energy Consumption Prediction for Smart Factory," in *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, 2022, pp. 879–882.

[14] J. C. Ferreira, L. B. Elvas, A. L. Martins, and N. Domingues, "Blockchain, iot, and smart grids challenges for energy systems," in *International Conference on Smart Computing and Communication*. Springer, 2024, pp. 65–80.

[15] D. Khan, L. T. Jung, and M. A. Hashmani, "Systematic literature review of challenges in blockchain scalability," *Applied Sciences*, vol. 11, no. 20, p. 9372, 2021.

[16] D.-S. Kim, I. S. Igboanusi, L. A. Chijioke Ahakonye, and G. O. Anyanwu, "Proof-of-Authority-and-Association Consensus Algorithm for IoT Blockchain Networks," in *2025 IEEE International Conference on Consumer Electronics (ICCE)*, 2025, pp. 1–6.

[17] S. Han, C. Zhang, L. Wang, W. Zheng, and X. Wen, "Fedecs: Client selection for optimizing computing energy in federated learning," in *2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2023, pp. 1–6.

[18] V. U. Ihekoronye, J. M. Lee, and D.-S. Kim, "Impact of adaptive client selection on federated learning for iomt ecosystem," in *2024 15th International Conference on Information and Communication Technology Convergence (ICTC)*, 2024, pp. 1603–1608.

[19] M. Adi Paramartha Putra, N. Bogi Aditya Karna, R. Naufal Alief, A. Zainudin, D.-S. Kim, J.-M. Lee, and G. A. Sampedro, "Purefed: An efficient collaborative and trustworthy federated learning framework based on blockchain network," *IEEE Access*, vol. 12, pp. 82 413–82 426, 2024.

[20] L. A. C. Ahakonye, C. I. Nwakanma, and D.-S. Kim, "Tides of Blockchain in IoT Cybersecurity," *Sensors*, vol. 24, no. 10, p. 3111, 2024.

[21] D. Singh, X. Zhou, and R. Jones, "A hierarchical blockchain-enabled distributed federated learning system with model contribution based rewarding," *Digital Communications and Networks*, vol. 10, no. 4, pp. 853–865, 2024.

[22] O. Marnissi, H. El Hammouti, and E. H. Bergou, "Adaptive sparsification and quantization for enhanced energy efficiency in federated learning," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 4307–4321, 2024.

[23] Wikipedia contributors, "American electric power — Wikipedia, the free encyclopedia," 2025, [Online; accessed 12-July-2025]. [Online]. Available: https://en.wikipedia.org/wiki/American_Electric_Power