Deceptive Synthetic Updates: Stealth Free-Rider Attack on Model Aggregation in Federated Learning

Youngjoon Lee
KAIST
School of Electrical Engineering
Daejeon, Republic of Korea
yjlee22@kaist.ac.kr

Jinu Gong
Hansung University
Department of Applied AI
Seoul, Republic of Korea
jinugong@hansung.kr

Joonhyuk Kang
KAIST
School of Electrical Engineering
Daejeon, Republic of Korea
jkang@kaist.ac.kr

Abstract

Federated Learning (FL) allows multiple clients to collaboratively train shared models without exchanging raw data, thereby preserving privacy. However, FL systems are vulnerable to malicious participants known as free-riders who exploit the collaborative nature without providing genuine data contributions. To expose this critical security threat, we introduce a novel stealth free-rider attack that leverages pre-trained forecasting models to generate highly realistic synthetic time-series data. Our approach enables malicious clients to deceive FL systems while obtaining benefits from fair participants' contributions, thereby undermining the integrity of federated networks. Numerical results on EEG-based sleep stage classification demonstrate that our attack maintains comparable performance with free-rider ratios up to 70% while causing catastrophic degradation when all clients are free-riders.

CCS Concepts

• **Computing methodologies** → **Machine learning**.

Keywords

Federated Learning, Free-rider Attack, Synthetic Data, Pre-trained Forecasting Models

ACM Reference Format:
Youngjoon Lee, Jinu Gong, and Joonhyuk Kang. 2025. Deceptive Synthetic Updates: Stealth Free-Rider Attack on Model Aggregation in Federated Learning. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM '25), November 10–14, 2025, Seoul, Republic of Korea.* ACM, Seoul, Republic of Korea, 5 pages. https://doi.org/10.1145/3746252.3760949

1 Introduction

Healthcare AI has revolutionized diagnostic and prognostic tasks by leveraging large-scale patient data [3]. However, traditional centralized approaches face significant challenges due to stringent privacy regulations and the sensitive nature of patient information [11]. Federated Learning (FL) addresses these issues by enabling institutions to collaboratively train shared models without exchanging raw data [9, 13]. Thus, FL improves model generalization through diverse clinical datasets and has been widely adopted in medical imaging and electronic health records analysis [7].

CIKM '25, Seoul, Republic of Korea

ACM ISBN 979-8-4007-2040-6/2025/11
https://doi.org/10.1145/3746252.3760949

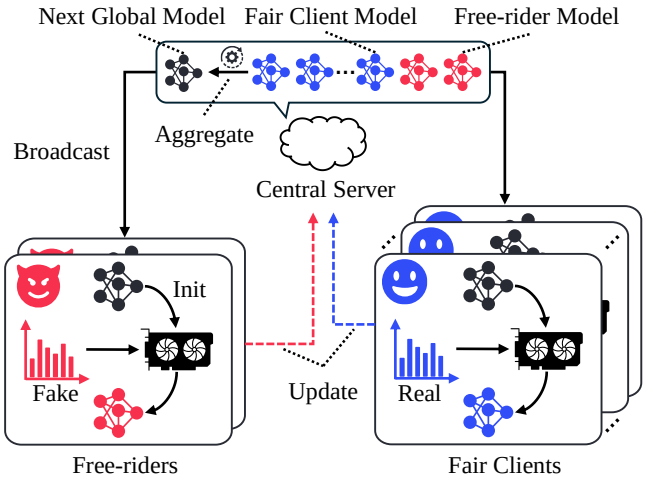

Figure 1: Illustration of the free-rider attack in FL. Fair clients (blue) train on real data and send updates to the server. In contrast, free-riders (red) generate synthetic data and train on these synthetic datasets to send updates based on highly realistic synthetic datasets.

Despite these advantages, FL systems face emerging security threats from malicious participants known as free-riders who exploit the collaborative nature of federated networks [6]. Specifically, free-riders attempt to benefit from other participants' data contributions without providing genuine local data for federated training [18]. As illustrated in Figure 1, these malicious clients generate synthetic data and send fake updates to the central server while fair clients upload fair updates based on real medical data. Therefore, these free-riders can successfully deceive current FL systems while obtaining benefits from fair clients' contributions [12]. Moreover, such free-rider attacks discourage fair clients from participating in FL networks and cause significant intellectual property damage.

In this work, we formalize a novel and stealthy free-rider threat model in time-series–based FL that leverages representative pre-trained forecasting models to generate highly realistic synthetic data. Rather than relying on widely studied image synthesis methods such as GANs or diffusion models, we focus on free-rider attacks that generate time-series synthetic data without requiring any additional training. Moreover, we conduct comprehensive experiments on an EEG-based sleep stage dataset using state-of-the-art FL techniques, and we find that even with free-rider ratios ranging from 0.1 to 0.7, performance differences compared to the all-fair-client setting remain minimal.

The main contributions of this paper are as follows:

- We formalize a novel free-rider threat model that leverages pre-trained forecasting models to generate realistic synthetic time-series data without additional training.
- We demonstrate that our proposed attack successfully deceives state-of-the-art FL methods while maintaining comparable performance across varying free-rider ratios (0.1-0.7).
- In addition, we validate that when all clients are free-riders, the FL system experiences a dramatic performance degradation, confirming the dependency on fair client contributions.

2 Proposed Stealth Free-Rider Attack

We consider a standard FL setting with N participating clients, including fair clients and free-riders, and a central server collaboratively training a global model θ_g while preserving data locality. Each client $i \in \{1, \ldots, N\}$ owns a private time-series dataset d^i_{real}, such as EEG signals, reflecting non-IID label distributions simulated via a Dirichlet prior [10].

In each communication round r, the server broadcasts θ^r_g to all clients. Then it samples a subset $\mathcal{S}_r$ of $K \ll N$ clients; when only fair clients are selected, each updates its local model via

$$\theta^r_k \leftarrow \text{ClientOpt}(\theta^r_g, d^k_{real}), \tag{1}$$

where `ClientOpt` denotes the FL-specific local optimizer (e.g., SGD for FedAvg). The server then aggregates these fair updates to form the next global model:

$$\theta^{r+1}_g \leftarrow \text{ServerOpt}(\{\theta^r_k\}_{k\in\mathcal{S}_r}), \tag{2}$$

where `ServerOpt` denotes the aggregation rule (e.g., weighted averaging of client updates).

During a stealth free-rider attack—when $\mathcal{S}_r$ happens to include malicious clients—the free-rider $f \in \mathcal{S}_r$ avoids using any real data and instead generates realistic synthetic time-series data d^f_{fake} via pre-trained forecasters without further fine-tuning. Specifically, the free-rider employs a chunked forecasting approach so that its entire local dataset can be replaced with synthetic time-series data. Given an original time-series sequence $\mathbf{x}_{real} \in \mathbb{R}^T$ of length T (e.g., $T = 3000$ for EEG signals) from client f's private dataset d^f_{real}, we first apply z-score normalization to obtain:

$$\tilde{\mathbf{x}}_{real} = \frac{\mathbf{x}_{real} - \mu_{\mathbf{x}}}{\sigma_{\mathbf{x}}}, \tag{3}$$

where $\mu_{\mathbf{x}}$ and $\sigma_{\mathbf{x}}$ are the sample mean and standard deviation. Next, the free-rider partitions $\tilde{\mathbf{x}}_{real}$ into non-overlapping chunks of size C, thereby obtaining:

$$\mathbf{c}_m = \tilde{\mathbf{x}}_{real}\{mC : (m+1)C\} \quad \text{for } m \in 0, 1, \ldots, \lfloor T/C \rfloor - 1 \tag{4}$$

where $\lfloor \cdot \rfloor$ denotes the floor operation. For each chunk, we apply a pre-trained forecasting model $\mathcal{F}_\phi$ without any fine-tuning to generate synthetic predictions via zero-shot synthesis:

$$\hat{\mathbf{c}}_m = \mathcal{F}_\phi(\mathbf{c}_m; \delta = C) \tag{5}$$

where δ represents the prediction horizon length equal to the chunk size. The synthetic sequence is reconstructed by concatenating all predicted chunks:

$$\mathbf{x}_{synth} = \text{Concat}(\hat{\mathbf{c}}_0, \hat{\mathbf{c}}_1, \ldots, \hat{\mathbf{c}}_{\lfloor T/C \rfloor - 1})[:T] \tag{6}$$

where the final truncation ensures the synthetic sequence matches the original length T. Moreover, to maintain statistical consistency, the free-rider applies z-score normalization to the reconstructed sequence:

$$\mathbf{x}_{fake} = \frac{\mathbf{x}_{synth} - \mu_{synth}}{\sigma_{synth}} \tag{7}$$

where μ_{synth} and σ_{synth} are computed from $\mathbf{x}_{synth}$ itself. As a result, each free-rider f completely replaces its original dataset d^f_{real} with the synthetic version:

$$d^f_{fake} = \{(\mathbf{x}^{(j)}_{fake}, y^{(j)})\}_{j=1}^{|d^f_{real}|}, \tag{8}$$

where labels $y^{(j)}$ are preserved from the original data. The quality of the synthetic time-series data critically determines the effectiveness and stealthiness of our proposed free-rider attack. To ensure high-quality synthetic data generation, we employ representative pre-trained forecasters, including TimesFM (200M), TimesFM (500M) [5], Chronos (Bolt Base), and Chronos (Bolt Small) [2]. These models leverage different architectural designs and parameter scales to capture diverse temporal patterns in time-series data. Note that the synthetic dataset d^f_{fake} for each free-rider client f remains fixed throughout the federated training process. Furthermore, during any federated training round in which the free-rider is selected, it performs local optimization solely on its synthetic dataset:

$$\theta^r_f \leftarrow \text{ClientOpt}(\theta^r_g, d^f_{fake}) \tag{9}$$

This enables free-riders to participate in FL without processing any real private data, while their deceptive synthetic updates are aggregated through Equation (2) alongside fair client contributions. The federated network repeats the process for a maximum of R communication rounds until the desired performance is achieved. Overall procedure is described in Algorithm 1.

Algorithm 1: Proposed Stealth Free-Rider Attack in FL

```
\* Deceptive synthesis for every free-rider f ∈ F *\
foreach f ∈ F do
    d^f_fake ← OfflineSynth(d^f_real, C, F_φ)
end
\* Central server start FL *\
Initialize global model θ^0_g
for r ← 0 to R − 1 do
    server broadcasts θ^r_g; sample S_r with |S_r| = K
    for k ∈ S_r do in parallel
        if k ∈ F then
            \* Deceptive Synthetic Update *\
            θ^r_u ← ClientOpt(θ^r_g, d^k_fake)
        else
            \* Fair Update *\
            θ^r_u ← ClientOpt(θ^r_g, d^k_real)
        end
    end
    θ^{r+1}_g ← ServerOpt({θ^r_k}_{k∈S_r})
end
return θ^R_g
```

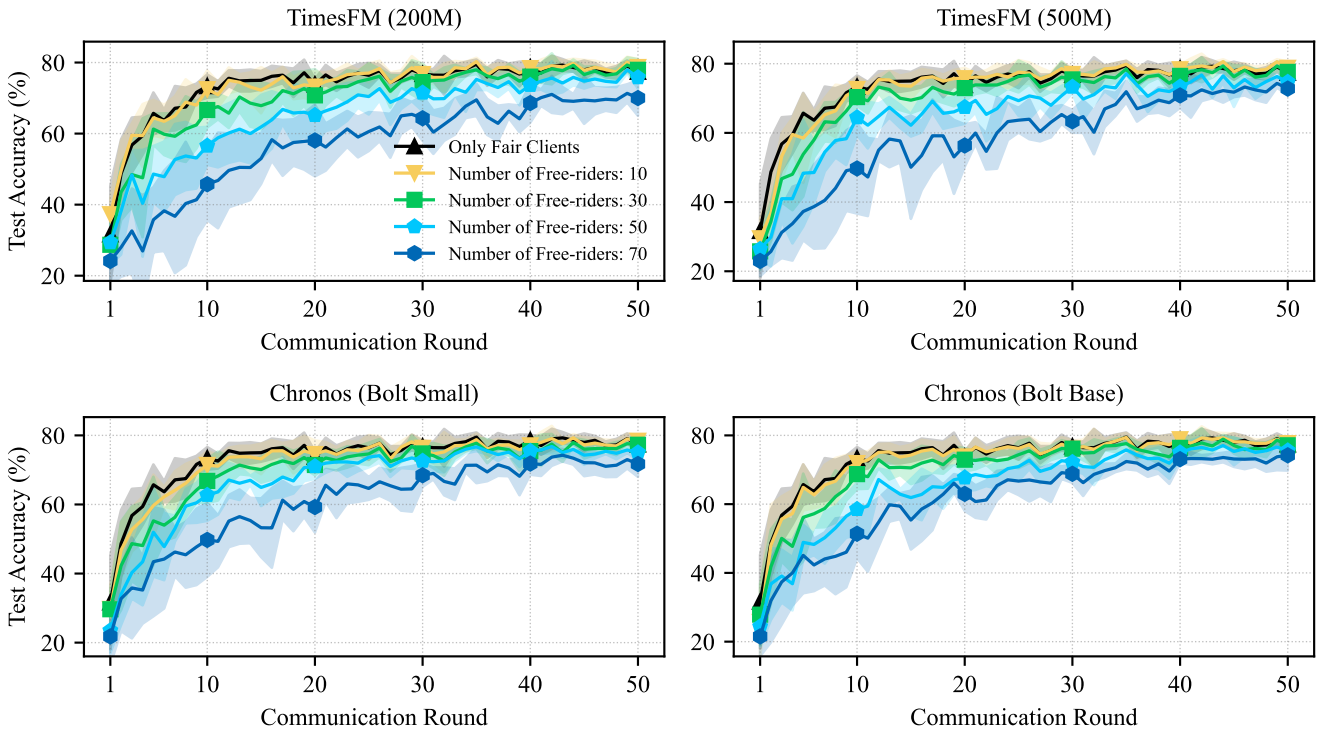

Figure 2: Performance comparison under varying free-rider ratios in FL. Each subplot shows test accuracy over communication rounds for TimesFM (200M/500M), Chronos (Bolt Small/Bolt Base) respectively. The curves represent different scenarios: fair clients only (baseline) and systems with 10, 30, 50, and 70 free-riders out of 100 total clients.

3 Experiment and Results

3.1 Experiment Setting

To evaluate the efficacy of our stealth free-rider attack, we conducted experiments using the EEG-based sleep stage classification dataset [8] and the ConvNeXt V2 model [17]. Specifically, the dataset was partitioned across $N = 100$ clients using a Dirichlet distribution with parameter $\alpha = 1.0$, thereby creating non-IID conditions that reflect realistic medical data heterogeneity across institutions. Moreover, in each communication round, we randomly sample $K = 10$ clients to participate in local updates. We then tested six recent FL methods: FedAvg [13], FedDyn [1], FedSAM [14], FedGamma [4], FedSMOO [15], and FedSpeed [16]. In addition, we utilized four representative pre-trained forecasters—TimesFM (200M/500M), and Chronos (Bolt Small/Bolt Base)—for synthetic data generation. In particular, we fix the chunk size at $C = 64$, which corresponds to the maximum prediction length supported by Chronos models, and systematically vary the free-rider ratio from 0.1 to 0.7 to evaluate attack effectiveness under diverse threat scenarios. Furthermore, all experiments are run with five different random seeds, and we report the mean and standard deviation. Other hyperparameter settings and detailed procedures are available in our open repository[1].

3.2 Numerical Results

3.2.1 Impact of Proposed Stealth Free-rider Attack. To investigate the effect of our proposed stealth free-rider attack, we evaluated four pre-trained forecasting models across free-rider ratios from 0.1 to 0.7 using vanilla FL. As shown in Fig. 2, our results demonstrate stable performance degradation patterns, with minimal impact at lower free-rider ratios. Specifically, the synthetic data successfully deceives the FL system while maintaining convergence characteristics. Furthermore, the attack shows gradual performance decline as free-rider ratio increases, thereby indicating effective masking of malicious behavior. Notably, all four models achieve identical baseline performance of 77.38% ± 1.68% with fair clients only.

Among the forecasters, Chronos (Bolt Base) exhibits the highest resilience with only 4.1% relative performance drop at 70% free-rider ratio. In contrast, TimesFM (200M) shows the greatest vulnerability, experiencing 9.5% relative performance degradation under the same conditions. Moreover, at 30% free-rider ratio, Chronos models show slight performance improvements (0.1-0.3%), while TimesFM models exhibit minimal degradation (-0.3% to -0.7%). Interestingly, systems with lower free-rider ratios (10-30%) converge faster, reaching stable performance in 14-16 rounds compared to 17 rounds in the fair-only baseline. However, higher ratios (50-70%) significantly slow convergence, requiring up to 39 rounds for stable performance.

[1]https://github.com/yjlee22/free-rider

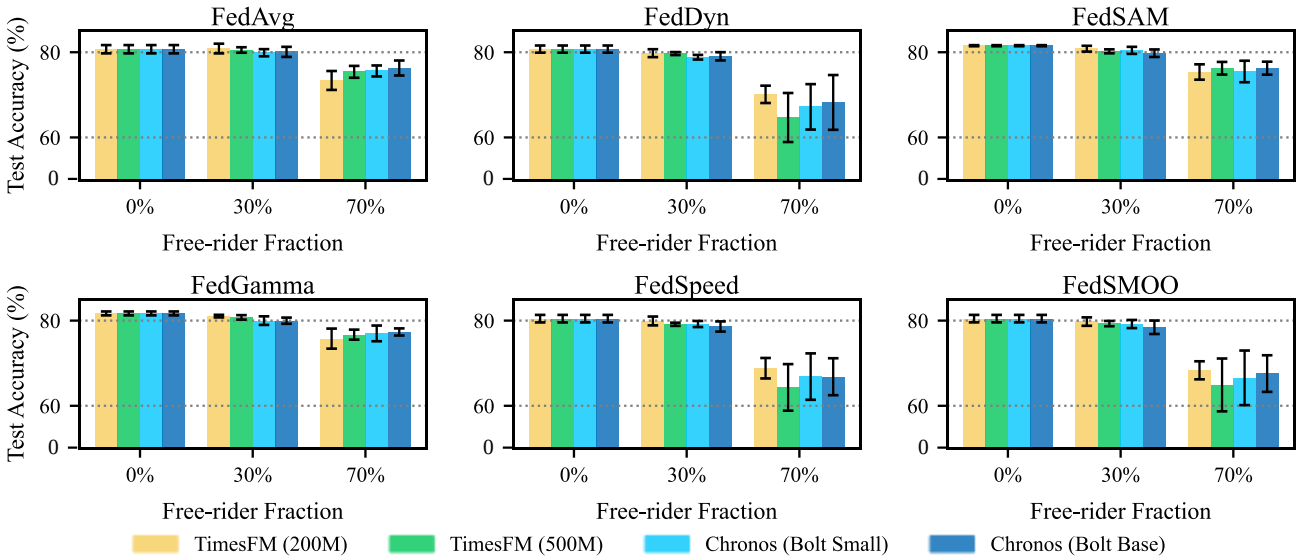

Figure 3: Effectiveness of the proposed stealth free-rider attack across six state-of-the-art FL methods. Each subplot shows test accuracy under three scenarios: fair clients only (0%), 30% free-riders, and 70% free-riders.

Table 1: Performance comparison between fair clients only and all free-riders scenarios.

Method	Forecaster	Test Acc. (%)	Diff (%)
FedAvg	TimesFM (500M)	27.93 ± 2.58	-52.79
	Chronos (Bolt Base)	35.97 ± 2.95	-44.74
FedDyn	TimesFM (500M)	41.03 ± 5.94	-39.71
	Chronos (Bolt Base)	44.63 ± 4.15	-36.12
FedSAM	TimesFM (500M)	37.00 ± 8.92	-44.54
	Chronos (Bolt Base)	38.31 ± 3.19	-43.23
FedGamma	TimesFM (500M)	38.59 ± 8.88	-43.10
	Chronos (Bolt Base)	34.91 ± 7.78	-46.78
FedSpeed	TimesFM (500M)	39.74 ± 5.42	-40.71
	Chronos (Bolt Base)	40.61 ± 2.87	-39.85
FedSMOO	TimesFM (500M)	40.59 ± 6.21	-39.87
	Chronos (Bolt Base)	40.93 ± 3.36	-39.53

3.2.2 Impact of Free-rider Attack Across Different FL Methods. To demonstrate the generalizability of our proposed stealth free-rider attack, we evaluated its effectiveness across six state-of-the-art FL methods. As shown in Fig. 3, all methods show minimal performance degradation at 30% free-rider ratio, with accuracy drops ranging from only 0.5% (FedAvg) to 1.7% (FedDyn). However, at 70% free-rider ratio, significant performance degradation becomes evident across all methods, with accuracy drops ranging from 6.1% to 17.1%. Notably, FedSMOO exhibits the most severe vulnerability with a 16.9% accuracy drop at 70% free-rider ratio, while FedGamma demonstrates the highest resilience with only a 6.1% drop. Nevertheless, the consistent attack effectiveness across diverse FL algorithms confirms the generalizability of our proposed approach.

3.2.3 Ablation Study. To validate the dependency on fair client contributions, we conducted an ablation study comparing scenarios with only fair clients versus all clients being free-riders (100% free-rider ratio). As shown in Table 1, the all-free-rider scenario results in catastrophic performance degradation across all FL methods, with accuracy drops ranging from 36.12% to 52.79%. Specifically, FedAvg exhibits the most severe vulnerability with TimesFM (500M), experiencing a dramatic 52.79% accuracy drop from 80.71% to 27.93%. In contrast, FedDyn demonstrates the highest resilience, showing the smallest degradation of 36.12% with Chronos (Bolt Base), achieving 44.63% accuracy in the all-free-rider setting. Despite these severe degradations, the results indicate that synthetic data can still enable some level of learning, as evidenced by the above-random performance across all configurations.

4 Conclusion

In this work, we propose a novel stealth free-rider attack for FL that leverages pre-trained forecasting models to generate highly realistic synthetic time-series data. We demonstrated that our attack successfully deceives state-of-the-art FL methods across varying free-rider ratios from 0.1 to 0.7 while maintaining comparable performance to fair-client scenarios. Furthermore, our ablation study confirmed that fair client contributions are essential, with all-free-rider scenarios causing catastrophic performance drop. Thus, our work exposes critical FL security vulnerabilities and confirms the necessity of genuine data contributions for robust FL.

Acknowledgements

This research was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP)-ITRC (Information Technology Research Center) grant funded by the Korea government (MSIT) (IITP-2025-RS-2020-II201787).

GenAI Usage Disclosure

We use Claude 4 to assist with code development and the refinement of the manuscript.

References

[1] Durmus Alp Emre Acar, Yue Zhao, Ramon Matas, Matthew Mattina, Paul Whatmough, and Venkatesh Saligrama. May 2021. Federated learning based on dynamic regularization. In *Proc. ICLR*. Vienna, Austria.

[2] Abdul Fatir Ansari, Lorenzo Stella, Ali Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, et al. Nov. 2024. Chronos: Learning the Language of Time Series. *Trans. Mach. Learn. Res.* (Nov. 2024).

[3] Igor Bisio, Caterina Fallani, Chiara Garibotto, Halar Haleem, Fabio Lavagetto, Mehrnaz Hamedani, Angelo Schenone, Andrea Sciarrone, and Matteo Zerbino. Mar. 2025. AI-Enabled Internet of Medical Things: Architectural Framework and Case Studies. *IEEE Internet Things Mag.* 8, 2 (Mar. 2025), 121–128.

[4] Rong Dai, Xun Yang, Yan Sun, Li Shen, Xinmei Tian, Meng Wang, and Yongdong Zhang. Dec. 2024. FedGAMMA: Federated Learning With Global Sharpness-Aware Minimization. *IEEE Trans. Neural Netw. Learn. Syst.* 35, 12 (Dec. 2024), 17479–17492.

[5] Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. Jul. 2024. A decoder-only foundation model for time-series forecasting. In *Proc. ICML*. Vienna, Austria.

[6] Yann Fraboni, Richard Vidal, and Marco Lorenzi. Apr. 2021. Free-rider attacks on model aggregation in federated learning. In *Proc. AISTAT*. Virtual Event.

[7] P. Kairouz and H.B. McMahan. 2021. *Advances and Open Problems in Federated Learning*. Found. Trends Mach. Learn., Vol. 14. 1–210 pages.

[8] B. Kemp, A.H. Zwinderman, B. Tuk, H.A.C. Kamphuisen, and J.J.L. Oberye. Sep. 2000. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. *IEEE Trans. Biomed. Eng.* 47, 9 (Sep. 2000), 1185–1194.

[9] Youngjoon Lee, Jinu Gong, Sun Choi, and Joonhyuk Kang. 2025. Revisit the Stability of Vanilla Federated Learning Under Diverse Conditions. *arXiv preprint arXiv:2502.19849* (2025).

[10] Qinbin Li, Yiqun Diao, Quan Chen, and Bingsheng He. May 2022. Federated learning on non-iid data silos: An experimental study. In *Proc. IEEE ICDE*. Kuala Lumpur, Malaysia.

[11] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. May 2020. Federated learning: Challenges, methods, and future directions. *IEEE Signal Process. Mag.* 37, 3 (May 2020), 50–60.

[12] Yanli Li, Zhongliang Guo, Nan Yang, Huaming Chen, Dong Yuan, and Weiping Ding. May 2025. Threats and defenses in the federated learning life cycle: A comprehensive survey and challenges. *IEEE Trans. Neural Netw. Learn. (Early Access)* (May 2025).

[13] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Apr. 2017. Communication-efficient learning of deep networks from decentralized data. In *Proc. AISTAT*. Fort Lauderdale, United States.

[14] Zhe Qu, Xingyu Li, Rui Duan, Yao Liu, Bo Tang, and Zhuo Lu. Jul. 2022. Generalized federated learning via sharpness aware minimization. In *Proc. ICML*. Baltimore, United States.

[15] Yan Sun, Li Shen, Shixiang Chen, Liang Ding, and Dacheng Tao. Jun. 2023. Dynamic regularized sharpness aware minimization in federated learning: Approaching global consistency and smooth landscape. In *Proc. ICML*. Hawaii, United States.

[16] Yan Sun, Li Shen, Tiansheng Huang, Liang Ding, and Dacheng Tao. May 2023. FedSpeed: Larger Local Interval, Less Communication Round, and Higher Generalization Accuracy. In *Proc. ICLR*. Kigali, Rwanda.

[17] Sanghyun Woo, Shoubhik Debnath, Ronghang Hu, Xinlei Chen, Zhuang Liu, In So Kweon, and Saining Xie. Jun. 2023. Convnext v2: Co-designing and scaling convnets with masked autoencoders. In *Proc. IEEE/CVF CVPR*. Vancouver, Canada.

[18] Mang Ye, Xiuwen Fang, Bo Du, Pong C Yuen, and Dacheng Tao. Oct. 2023. Heterogeneous federated learning: State-of-the-art and research challenges. *ACM Comput. Surv.* 56, 3 (Oct. 2023), 1–44.