# Federated Learning-Based State Estimation for Integrated Transmission–Distribution Networks

Anass Akaouche
*Intelligent Electrical Power Grids*
*Delft University of Technology*
Delft, The Netherlands
aakaouche@tudelft.nl

Jonathan Aviles-Cedeno
*Intelligent Electrical Power Grids*
*Delft University of Technology*
Delft, The Netherlands
j.a.avilescedeno@tudelft.nl

Jose Rueda-Torres
*Intelligent Electrical Power Grids*
*Delft University of Technology*
Delft, The Netherlands
j.l.ruedatorres@tudelft.nl

***Abstract*—The growing interdependence between transmission and distribution networks, combined with increasing data privacy constraints, necessitates state-estimation methods that facilitate multi-operator coordination without requiring centralized data sharing. This paper presents a federated learning (FL) framework for neural-network-based state estimation in integrated transmission–distribution systems. Each operator trains a local multilayer perceptron (MLP) using only its own measurements, while a TSO-level server aggregates model updates to obtain a global estimator. The approach is demonstrated on a steady-state model of the Dutch transmission system coupled with three reduced-order distribution feeders representing distinct DSO regions. Three FL algorithms are compared in terms of accuracy and convergence: FedAvg, FedProx, and FedAdam. Results show that FL achieves competitive estimation performance relative to a centralized benchmark while preserving data ownership, with FedProx providing the fastest convergence and FedAdam obtaining the most accurate predictions. The study highlights the potential of FL as a foundation for privacy-preserving coordination in future digitalized and multi-operator power systems.**

***Index Terms*—State estimation, transmission–distribution systems, TSO–DSO coordination, power system digitalization, federated learning, neural networks, digital twin.**

## I. Introduction

The growing penetration of converter-interfaced generation, prosumers, and flexible demand is increasing the interdependence between transmission and distribution grids. Thus, having the proper capabilities to estimate system states across the entire power network, represented by voltage magnitudes and angles, becomes especially relevant.

In the Netherlands, as in many countries, TSOs and DSOs run state estimation (SE) independently and exchange information only at their interconnection points. This separation overlooks the influence of DER-rich distribution grids on the injections, voltages, and power flows observed at the transmission level.

This work develops a federated learning (FL) framework for state estimation in combined transmission–distribution (T–D) systems, where each operator trains a local neural estimator without sharing raw data. The study evaluates the performance and convergence of FL-based state estimation under the heterogeneous measurement structure and operating conditions characteristic of multi-operator T–D grids. To this end, we extend a previously developed steady-state model of the Dutch transmission grid [1] by appending three reduced distribution feeders with renewable generation and aggregated loads, forming a synthetic yet representative TSO–DSO system.

The main contributions of this work are:

- A complete FL framework for multi-operator state estimation in an integrated T–D context.
- Comparative analysis of three FL algorithms: FedAvg, FedProx, and FedAdam.
- Validation on a Dutch-grid-based case study, indicating that FL-based state estimation attains voltage estimation accuracy suitable for power-system analysis while preserving operator data locality.

The remainder of this paper is organized as follows. Section II discusses the limitations of conventional SE and motivates the use of an FL neural-network-based approach. Section III details the system modeling and the generation of the dataset used for training and validation. Section IV introduces the proposed FL framework for multi-operator SE. Section V presents and analyzes the numerical results, while Section VI summarizes the main findings and outlines directions for future work.

## II. Background and Motivation

SE in power systems is traditionally performed by TSOs using a centralized Weighted Least Squares (WLS) formulation. Although WLS does not require full measurement coverage, it assumes an observable transmission network and represents each distribution grid as a static net active/reactive power injection at the T–D interface. DSOs operate their own SE independently, so TSOs cannot monitor internal distribution feeder behavior, including DER variability, voltage-dependent loads, and local control actions. This separation does not hinder transmission-level SE, but it limits the accuracy of operational studies that rely on boundary sensitivities or predict how the T–D exchange varies under contingencies or redispatch actions. Existing distributed or hierarchical SE

The research work shown in this paper has received funding from TenneT TSO B.V within the research project on Adaptive fast active power control for stabilization of multi-converter dynamics in offshore electrical energy-hydrogen hubs - FUTURESYSTEM. It reflects only the authors' views, and the aforesaid organization is not responsible for any use that may be made of the paper's content.

979-8-3315-8138-1

schemes enhance boundary consistency through limited data exchange; however, they remain sensitive to heterogeneous measurements and rely on iterative physics-based solvers.

A coordinated TSO–DSO estimator can address these limitations by learning how internal distribution conditions influence the T–D exchange. A unified model reduces reliance on static equivalents or pseudo-measurements, improves voltage-security assessments through better representation of reactive power responses and reverse flows, and enables more coherent congestion and flexibility management across voltage levels.

Machine-learning (ML) estimators offer a fast alternative by learning nonlinear mappings between measurements and states. Neural and ensemble models can approach WLS accuracy with lower inference costs and robustness to noise [2], and physics-informed or graph-based architectures embed structural constraints to improve generalization [3]. Centralized ML, however, requires the aggregation of operator data, which raises concerns regarding privacy, ownership, and scalability.

FL enables TSOs and DSOs to jointly train a shared estimator while keeping measurements local to each operator. This structure aligns with the operational separation of system operators, distributes computation, and avoids centralized data aggregation as the number of participants increases. Prior work has applied FL to SE [4] and anomaly detection [5] while preserving data confidentiality, and personalized or hierarchical variants improve adaptation to heterogeneous networks [6]. However, existing studies often rely on small benchmark networks or simplified settings and do not evaluate FL on realistic transmission and distribution scales within a unified synthetic model. This gap motivates the present study, which assesses FL-based SE on an integrated T–D test system that reflects the size, heterogeneity, and privacy constraints expected in future digitalized grids.

## III. System Modeling and Data Generation

### A. Transmission–Distribution Test System

This study builds upon the Dutch steady-state transmission model developed in [1], which represents the 380, 220, 150, and 110 kV backbone of the national power system. This work accounts for distribution-level effects by attaching, at the lowest transmission voltage levels, aggregated models of solar, wind, and equivalent loads representing downstream power flows. Demand and renewable capacities were taken from public technical reports across multiple operating scenarios.

In the present research, three synthetic distribution networks are connected at representative 150 kV busbars located in Noord-Holland (DN1), Gelderland–Flevoland–Utrecht (DN2), and Brabant (DN3), as shown in Fig. 1. Critical demand values and nominal renewable capacities at these interfaces are redistributed within the associated distribution networks. The original aggregated blocks are kept as disabled elements, and new elements are instantiated inside the synthetic distribution networks to preserve comparability.

These areas collectively account for a significant portion of national demand and renewable capacity. The single-line diagram of the distribution networks is shown in Fig. 2.

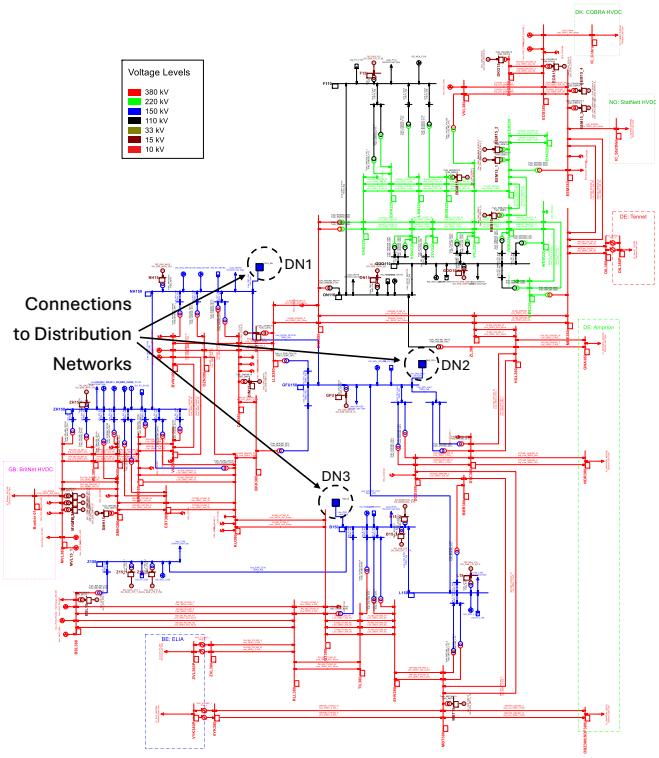



Fig. 1: Single-line diagram of the updated T–D system.

Each distribution network is connected through a 150/33 kV transformer, followed by a 33/10 kV stage, which supplies two feeders that serve industrial and residential consumers. Aggregated 150 kV loads from the original model are disabled and represented at 10 kV within each distribution network, maintaining regional active-power totals with a 60%/40% industrial/residential split.

Onshore wind capacity originally connected at the selected 150 kV nodes is partially relocated to 33 kV within each corresponding distribution network, reflecting the empirical mix of high- and medium-voltage connections of onshore wind generators. Utility-scale photovoltaic (PV) generation, previously represented at 150 kV, is de-aggregated to the distribution level and split between 33 kV (for large plants) and 10 kV (for small plants and rooftop installations), using regional deployment fractions derived from national statistics. Although small-scale rooftop PV operates at a low voltage (approximately 0.4–0.7 kV), it is aggregated at 10 kV here to limit network depth while preserving its dominant impact on feeder voltages and local power balance.

### B. Scenario Creation and Dataset Generation

A dataset of steady-state operating points is generated in DIgSILENT PowerFactory by varying net demand and renewable injections across the integrated T–D system. All inverter-based resources operate at a maximum power factor between 0.9 and 0.95 and follow a voltage-$Q$ droop control scheme. The network is simulated under fully renewable operating conditions, with the reactive effects of lines, transformers,

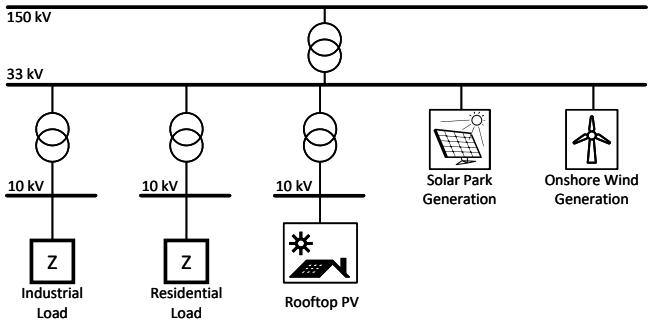



Fig. 2: Single-line diagram of the distribution networks.

and shunt elements fully represented. Each operating point is solved through an AC power flow, and the resulting bus voltages and branch flows constitute the data used in the FL experiments. Transmission- and distribution-level measurements remain partitioned to emulate the data boundaries between system operators.

Operating points are generated through structured parameter sweeps around stable base-case values. Total national demand is divided into 25 bins between 12 GW and 26 GW. Within each bin, renewable generation is scaled to match the aggregated demand plus transmission losses. Allocations of PV, wind, and load are perturbed by small random variations to obtain diverse yet feasible operating points. Solar generation spans $[0,\ 0.80]$ p.u., and wind generation spans $[0,\ 1.00]$ p.u., representing a wide range of Dutch renewable conditions.

To emulate measurement uncertainty, bounded uniform noise of $\pm 2\%$ is added to all input measurements, consistent with SCADA accuracy classes that specify maximum percentage errors rather than a specific probability distribution.

Each operator accesses only its local measurements and predictions, consistent with the FL framework. As shown in Fig. 3, the input features include bus voltage magnitudes, generator and load active and reactive powers, and tie-line power exchanges, while the neural network predicts the real and imaginary voltage components of all buses in both the transmission and distribution systems.

A total of 30 000 feasible operating points are generated; 10 000 are first used to warm start the neural estimator, reflecting the fact that practical SE systems are continually trained and never begin from an uninitialized model. The remaining 20 000 scenarios are used during the FL training rounds.

## IV. Federated Learning Framework

### A. Formulation and Algorithms

Let $\mathcal{K} = \{1, \ldots, K\}$ denote the set of participating operators (one TSO and $K-1$ DSOs in this case). Each client $k \in \mathcal{K}$ holds a private dataset $D_k = \{(x_i^{(k)}, y_i^{(k)})\}_{i=1}^{n_k}$, and the total number of data points is $n = \sum_k n_k$. Federated learning (FL) seeks to minimize the *weighted empirical risk*:

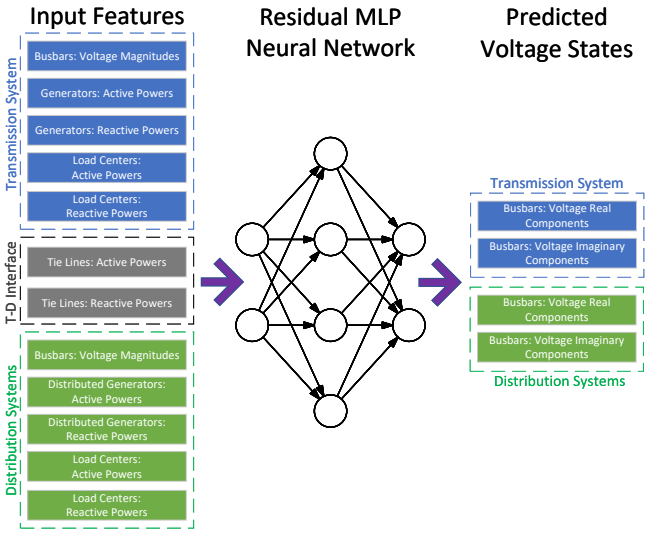



Fig. 3: Composition of input features and predicted voltage states for the residual MLP model.

$$F(w) = \sum_{k=1}^{K} \frac{n_k}{n} F_k(w), \tag{1}$$

with local objectives

$$F_k(w) = \frac{1}{n_k} \sum_{i=1}^{n_k} \ell\Big(f_w(x_i^{(k)}), y_i^{(k)}\Big), \tag{2}$$

where $f_w(\cdot)$ is the neural estimator and $\ell(\cdot)$ is the mean-squared error between predicted and reference voltages.

In this work, we compare three federated optimization strategies: FedAvg, FedProx, and FedAdam.

Each global communication round $t$ proceeds through three stages:

**1) Broadcast:** the TSO sends the global model $w_t$ to all the DSOs.

**2) Local update:** client $k$ initializes its local model as $w_{t,0}^{(k)} = w_t$ and performs $E$ Stochastic Gradient Descent (SGD) steps indexed by $j = 0, \ldots, E-1$.

- *FedAvg* [7] and *FedAdam* [8] use the standard local SGD rule

$$w_{t,j+1}^{(k)} = w_{t,j}^{(k)} - \eta \nabla F_k(w_{t,j}^{(k)}), \tag{3}$$

  which is effective when client datasets are not strongly heterogeneous.
- *FedProx* [9] modifies the client-side objective by adding a proximal term that limits drift from $w_t$:

$$F_k^{\text{prox}}(w; w_t) = F_k(w) + \frac{\mu}{2}\|w - w_t\|^2, \tag{4}$$

  where $\mu \geq 0$ controls the regularization strength. The resulting SGD step becomes

TABLE I: Federated learning hyperparameters for each strategy.

| | FedAvg | FedProx | FedAdam |
|---|---|---|---|
| Warm start | Yes | Yes | Yes |
| Rounds | 300 | 300 | 300 |
| Clients ($K$) | 4 | 4 | 4 |
| Batch size | 128 | 128 | 128 |
| Local epochs | 2 | 2 | 2 |
| Client Learning Rate $\eta$ | $10^{-4}$ | $10^{-4}$ | $10^{-4}$ |
| Client Weight Decay | $10^{-4}$ | $10^{-4}$ | $10^{-4}$ |
| Proximal $\mu$ | – | 0.005 | – |
| Server Learning Rate $\eta_s$ | – | – | $10^{-3}$ |
| Server $(\beta_1, \beta_2)$ | – | – | $(0.9, 0.999)$ |
| Server $\tau$ | – | – | $10^{-9}$ |

$$w_{t,j+1}^{(k)} = w_{t,j}^{(k)} - \eta\Big(\nabla F_k(w_{t,j}^{(k)}) + \mu(w_{t,j}^{(k)} - w_t)\Big). \quad (5)$$

After completing $E$ steps, the client returns $w_{t+1}^{(k)} = w_{t,E}^{(k)}$.

**3) Aggregation:** the TSO updates the global model using one of the following strategies.

- *FedAvg* [7] and *FedProx* [9]: weighted average of client models,

$$w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n}\, w_{t+1}^{(k)}. \quad (6)$$

- *FedAdam* [8]: To accelerate convergence under heterogeneous updates, the TSO applies adaptive moment estimation. Let

$$\Delta_t = \sum_{k=1}^{K} \frac{n_k}{n}\big(w_{t+1}^{(k)} - w_t\big). \quad (7)$$

Then,

$$m_{t+1} = \beta_1 m_t + (1-\beta_1)\Delta_t, \quad (8)$$

$$v_{t+1} = \beta_2 v_t + (1-\beta_2)\Delta_t \odot \Delta_t, \quad (9)$$

$$w_{t+1} = w_t - \eta_s \frac{m_{t+1}}{\sqrt{v_{t+1}} + \epsilon}, \quad (10)$$

where $\beta_1, \beta_2 \in [0, 1)$ are decay factors for the first and second moments, $\epsilon$ ensures numerical stability, and $\eta_s$ is the server learning rate.

### B. Implementation Setup

The FL process reflects the operational hierarchy of the Dutch grid, where the TSO maintains the global model and the DSOs train locally on their private measurements. The implementation uses the open-source *Flower* framework with synchronous communication rounds. The hyperparameter values used for FedAvg, FedProx, and FedAdam are summarized in Table I.

For the federated experiments, the 20 000 available scenarios are split globally into 80% training, 10% validation, and 10% testing. Feature normalization uses the global training statistics (the mean and standard deviation), and each client receives only the standardized features corresponding to its own measurement set. This preserves TSO–DSO data separation while ensuring consistent scaling across all local datasets.

All clients employ the same residual MLP architecture that maps the 243 input features to 270 voltage components. The network uses an input width of 64, two residual blocks of width 64, and a final linear output layer, totalling approximately $5\times10^4$ trainable parameters. The 20 000 scenarios used in the federated rounds provide a sample–to–parameter ratio appropriate for a model of this size, supporting stable training across heterogeneous clients. The hidden layers apply GELU activations, dropout, and Layer Normalization, which help maintain robustness and good gradient flow during learning.

Although Section IV-A presents the theoretical SGD updates, the implementation uses AdamW for local optimization, which empirically improves stability with heterogeneous data. The AdamW optimizer has a client learning rate $10^{-4}$ and a weight decay $10^{-4}$ applied to non-head parameters, consistent across all strategies.

During each communication round, each DSO trains for $E = 2$ local epochs with a batch size of 128. Performance is quantified using the root-mean-squared error (RMSE) and mean absolute error (MAE),

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}, \quad (11)$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|, \quad (12)$$

computed over the 270 predicted voltage components. Global metrics average these errors across all clients.

## V. Results and Discussion

### A. Centralized Benchmark

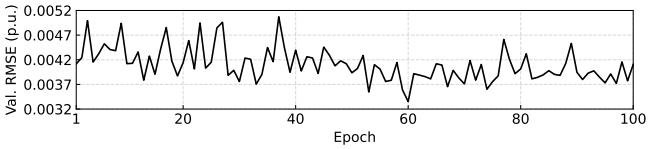
A centralized implementation of the neural estimator was trained to provide a reference for the federated strategies. It used the same warm-start initialization and the same set of 20 000 scenarios that participated in the federated rounds. This model provides an upper bound on the achievable accuracy, as it has direct access to the full dataset. In practice, however, such centralized data aggregation is often incompatible with multi-operator data ownership constraints, which motivates the federated approach.

### B. Convergence Behavior

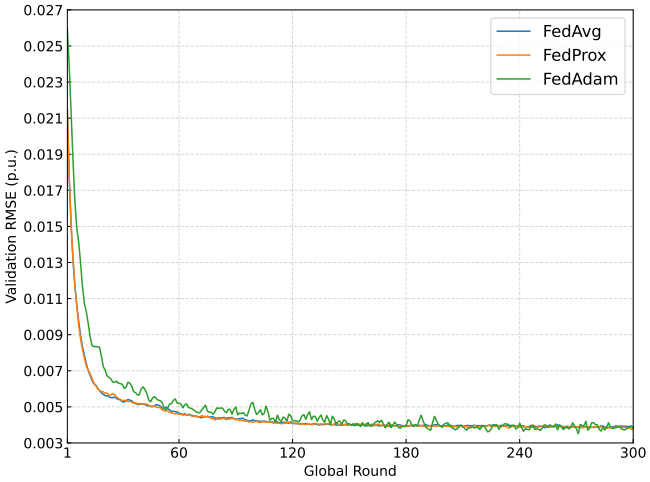
Fig. 4 shows the validation RMSE for both the centralized baseline and the federated strategies. The centralized model remains stable across training, with RMSE values between $3.35 \times 10^{-3}$ p.u. and $5.07 \times 10^{-3}$ p.u.

All three federated methods converge rapidly. FedAvg and FedProx reduce their errors smoothly during the first 30 global rounds, with FedProx slightly more stable due to the proximal term that limits client drift. FedAdam initially oscillates but quickly benefits from adaptive server-side moment updates and

reaches the lowest steady-state validation error. Convergence stabilizes after roughly 120 rounds.




(a) *Centralized model.*




(b) *Federated learning models.*

Fig. 4: Validation RMSE of (a) centralized and (b) federated estimators, in per-unit (p.u.).

### C. Accuracy Comparison

Fig. 5 summarizes the test errors. The centralized model achieves RMSE $=$ 0.0033 p.u. and MAE $=$ 0.0023 p.u. FedAdam reaches the best federated performance (RMSE $=$ 0.0158 p.u., MAE $=$ 0.0104 p.u.), followed by FedProx (RMSE $=$ 0.0167 p.u., MAE $=$ 0.0110 p.u.) and FedAvg (RMSE $=$ 0.0170 p.u., MAE $=$ 0.0112 p.u.). An RMSE on the order of $10^{-2}$ p.u. corresponds to voltage magnitude errors of around 1–2%, which are comparable to typical measurement uncertainty levels in operational power systems. FedProx improves robustness under imbalanced client datasets, while FedAdam benefits from adaptive server-side updates.

The experiments use small distribution feeders, increasing heterogeneity between clients. In real systems, larger distribution grids provide richer measurement sets, reducing heterogeneity; the results here are therefore conservative relative to realistic TSO–DSO deployments.

## VI. Conclusions and Future Work

This work presents a federated learning framework for multi-operator state estimation in an integrated T–D system, enabling TSOs and DSOs to train a shared neural-network estimator while keeping all measurements local.

Using a synthetic model of the Dutch transmission system augmented with three reduced distribution feeders, we

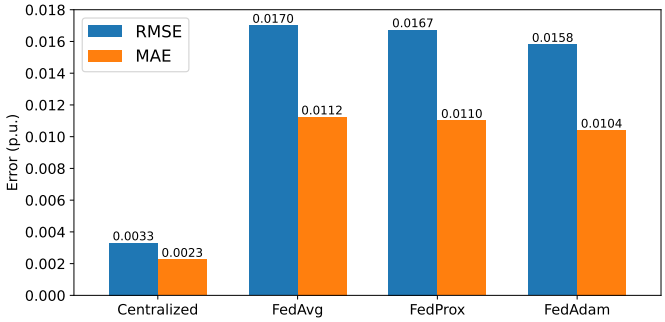



Fig. 5: Test performance of centralized and federated models (RMSE and MAE), in per unit (p.u.).

compared FedAvg, FedProx, and FedAdam under partitioned operator data. All methods showed stable convergence, with FedAdam achieving the lowest errors and FedProx improving robustness under client heterogeneity. Because the study employs small distribution feeders, the level of heterogeneity is likely higher than in realistic deployments, making the results conservative.

These findings indicate that FL is a viable approach for coordinated TSO–DSO state estimation under operator-level data separation. Future work will consider larger distribution models, communication constraints, and the use of the estimator for initializing digital twins for dynamic assessment and control.

## References


[1] J. Aviles-Cedeno, J. Rueda, and A. de Roos, "Toward a Digital Twin of the Dutch EHV Network: Analyzing Future Multi-Energy Scenarios," in *2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, Oct. 2024, pp. 1–5.

[2] T. H. B. Huy, D. N. Vo, H. D. Nguyen, H. P. Truong, K. T. Dang, and K. H. Truong, "Enhanced power system state estimation using machine learning algorithms," in *Proc. 2023 Int. Conf. on System Science and Engineering (ICSSE)*, 2023, pp. 401–405.

[3] L. Wang, Q. Zhou, and S. Jin, "Physics-guided deep learning for power system state estimation," *J. Mod. Power Syst. Clean Energy*, vol. 8, no. 4, pp. 607–615, 2020.

[4] C. H. Veena, P. Jithendar, A. R. J. Anandhi, A. Singla, M. Bansal, and R. R. Ali, "Federated learning for real-time state estimation and dynamic phasor analysis in wide-area power grids," in *Proc. 2023 Int. Conf. on Power Energy, Environment & Intelligent Control (PEEIC)*. IEEE, 2023, pp. 678–682.

[5] R. Raghuvamsi, Y. T. Teeparthi, B. Mao, and A. Ukil, "Incentive edge-based federated learning for false data injection attack detection on power grid state estimation," *Applied Energy*, vol. 314, p. 118828, 2022.

[6] H. Wu, Z. Xu, and J. Ruan, "PFL-DSSE: A personalized federated learning approach for distribution system state estimation," *CSEE Journal of Power and Energy Systems*, vol. 10, no. 5, pp. 2265–2270, 2024.

[7] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera-Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*. PMLR, Apr. 2017, pp. 1273–1282.

[8] S. J. Reddi, Z. Charles, M. Zaheer, Z. Garrett, K. Rush, J. Konečný, S. Kumar, and H. B. McMahan, "Adaptive federated optimization," in *Proceedings of the 9th International Conference on Learning Representations (ICLR)*, 2021.

[9] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proceedings of Machine Learning and Systems*, I. Dhillon, D. Papailiopoulos, and V. Sze, Eds., vol. 2, 2020, pp. 429–450.