

Topology-Dependent Neural Surrogate for Fast Alternating-Current Power Flow in Reconfigurable Active Distribution Networks¹

Mohammad Yasin Rahmaty Darvazi*, Hossein Afrakhte

Department of Electrical Engineering, Faculty of Engineering, University of Guilan, Rasht, Iran

* Corresponding author; Email: yasin_rahmaty@webmail.guilan.ac.ir

Abstract:

Active distribution networks with distributed energy resources, energy storage systems, and remotely operated switches exhibit coupled continuous–discrete behavior: nodal injections vary continuously, whereas feeder reconfiguration changes the network topology. This paper develops a topology-dependent neural surrogate for fast alternating-current power flow evaluation in reconfigurable distribution systems. The proposed model incorporates switch statuses together with load-scaling factors and active and reactive power setpoints of distributed generation and energy storage systems, while graph-theoretic connectivity and radiality tests are used during scenario generation to retain only admissible feeder configurations. A fixed-dimensional target representation is adopted to predict voltage magnitudes at all buses, active power flows on fixed physical branches, and total feeder losses under changing radial topologies. The surrogate is trained using Newton–Raphson solutions generated in Pandapower for 100,000 feasible operating scenarios of a modified 33-bus benchmark active distribution feeder. Test results show close agreement with the Newton–Raphson reference solutions, with voltage, flow, and loss coefficients of determination of 0.99971, 0.99946, and 0.99912, respectively. The mean absolute error of total feeder loss is 0.0079 MW. Batched central processing unit inference requires 0.54 ms per scenario, corresponding to an approximately 44-fold throughput speedup relative to the Newton–Raphson solver. The results indicate that the proposed surrogate is suitable for rapid screening and many-query reconfiguration or optimization studies, provided that final operational decisions are validated using a full alternating-current power flow solution.

Keywords:

active distribution networks, alternating-current power flow, neural surrogates, feeder reconfiguration, distributed energy resources

1- Introduction

Climate-change mitigation and the transition toward low-carbon energy systems require rapid decarbonization of electricity supply, reduced dependence on fossil-fuel-based generation, and accelerated deployment of renewable energy resources [1]. At the distribution level, the growing penetration of distributed energy resources (DERs), including solar photovoltaic (PV) generation, energy storage systems (ESSs), and dispatchable distributed generation (DG) units, is transforming conventional passive feeders into active distribution networks (ADNs), which provide key enabling infrastructure for flexible and sustainable electricity systems [2, 3]. Compared with traditional unidirectional operation, DER-rich ADNs exhibit bidirectional power flows and greater operating variability, increasing the need for coordinated voltage and reactive power control, loss-aware and economically

¹ **Declaration of Artificial Intelligence Use:** Artificial intelligence methods employed in this study include a multilayer perceptron neural network for approximating alternating-current power-flow solutions, as well as XGBoost [30] and random forest [31] for comparative loss prediction. Additionally, ChatGPT (OpenAI; <https://chatgpt.com/>) was used to assist with English-language editing, including improvements to grammar, phrasing, and readability. The authors take full responsibility for the scientific content and final manuscript.

efficient dispatch, and feeder reconfiguration through remotely controllable sectionalizing and tie switches to support loss reduction, load balancing, and service restoration [4, 5].

Monitoring and control of reconfigurable ADNs require repeated evaluation of bus voltages, branch flows, and losses under varying operating points and switch configurations. These quantities follow from the nonlinear alternating-current power flow (ACPF) equations, which are commonly solved using iterative numerical methods such as Newton–Raphson (NR). Although accurate, iterative solvers can become a practical bottleneck when embedded in higher-level routines that require thousands of evaluations, and their convergence behavior can be sensitive to initialization and operating conditions. Distribution-oriented load-flow studies have therefore sought faster and more robust computational methods for feeders with extensive DG penetration [6], while reconfiguration studies have further shown that binary switch statuses and nonlinear power-flow constraints make the operating problem intrinsically mixed-integer and topology dependent [7].

To reduce the cost of repeated evaluations, prior work spans both algorithmic and learning-assisted strategies. Distribution-oriented load-flow formulations and modeling approaches streamline computations for radial feeders with distributed generation [6], whereas reviews of surrogate modeling and deep neural network (DNN) methods in power systems emphasize the promise of fast data-driven approximations for repeated operating studies [8, 9]. Recent artificial intelligence (AI)-driven estimation work further demonstrates the potential of deep-learning models for efficient power-network assessment and voltage prediction [10]. In parallel, data-driven surrogates learn nonlinear relationships between injections and steady-state variables and can deliver near-instantaneous inference after offline training. Feasibility-aware deep proxies for optimal power flow (OPF), including its alternating-current formulation (AC-OPF), and stability-aware variants illustrate the potential for substantial speedups while explicitly managing infeasibility risks and disturbance-related constraints [11, 12].

Because network topology strongly shapes power-flow behavior, graph-based learning has received increasing attention in surrogate modeling. Graph neural networks (GNNs) can exploit the grid graph to capture spatial coupling and have been used, for example, to generate high-quality initial estimates for AC-OPF that reduce solver effort [13]. Physics-informed graph learning extends this idea by incorporating physical laws and system structure into the learning process, making such approaches attractive for reconfiguration studies [14]. In distribution networks, learning-based approaches have also been explored for fast reconfiguration under variable loading [15] and for hybrid model-based and data-driven optimization tasks such as reactive power management [16]. More broadly, physics-informed learning aims to improve trustworthiness and extrapolation by penalizing power-balance violations or embedding constraint-aware objectives, often improving loss-related fidelity and reducing

implausible predictions [17, 18]. Physics-informed and graph-based surrogates can improve physical consistency by incorporating power-balance residuals, network connectivity, or constraint-related terms into the learning process. However, these approaches require additional implementation choices, including the construction of topology-dependent graph operators, repeated evaluation of physical residuals during training, and selection of appropriate weights for data-fitting and physics-based loss components. Moreover, the use of physical or graph structure does not by itself guarantee accurate predictions for operating conditions or topologies that are insufficiently represented in the training dataset.

A complementary research direction considers graph-native surrogate solvers, in which GNN architectures approximate power-flow behavior by directly minimizing Kirchhoff-law residuals rather than solely imitating NR solutions [19]. At the same time, distribution-specific reviews emphasize adoption barriers, notably data requirements, robustness across operating regimes, and sensitivity to topology or model mismatch [20]. For reproducible large-scale labeling and benchmarking, the open-source Pandapower toolchain provides convenient network modeling and NR-based ACPF computation in Python [21]. The IEEE 33-bus feeder and its reconfiguration variants remain widely used test systems [22].

Topology-aware learning has emerged as an important direction in power-system surrogate modeling. Reference [23] proposed a single DNN for multiple AC-OPF problems with flexible network topology and parameters, while [24] developed a topology-aware GNN with feasibility-related regularization for adaptive AC-OPF. The comparative study in [25] showed that graph-based architectures provide limited additional benefit under fixed topology but become more advantageous when topology varies. More recently, [26] proposed a topology-aware gated GNN surrogate for AC power-flow estimation under load variation and topological uncertainty. These works highlight the value of explicitly representing network topology in learning-based models; however, they mainly address AC-OPF or rely on graph-specific architectures. In contrast, the present study adopts a feeder-specific MLP that directly combines binary switch statuses with continuous load, DG, and ESS variables to approximate the topology-dependent ACPF mapping, emphasizing implementation simplicity and efficient batched inference. Recent studies have explicitly addressed topology variability through topology-aware gated GNNs for power-flow estimation under varying or uncertain switch statuses [26], systematic benchmarking of GNN families using unified datasets and metrics [27], and graph-attention-based models that account for grid uncertainty [28]. Graph-based reinforcement learning for outage management and topology control further illustrates the operational value of topology-sensitive predictors [29]. Despite this progress, accurate surrogate modeling under discrete topology changes remains challenging: many approaches assume a fixed topology or require specialized graph

architectures, whereas practical reconfiguration applications benefit from surrogates that can handle both continuous operating variability and discrete switching actions in a deployable form.

Accordingly, this study develops a fast, topology-dependent surrogate for ACPF in reconfigurable ADN. Using a modified IEEE 33-bus feeder with DERs, ESS units, and seven controllable switches, we learn a mapping from operating conditions and switch statuses to bus voltage magnitudes, branch active power flows, and total feeder losses. Training data are generated by sampling feasible operating points and admissible switching states and labeling each scenario with NR solutions in Pandapower [21].

Key Contributions

This paper makes the following contributions:

1. **Joint learning of operating-point variability and reconfiguration actions:** We formulate a topology-dependent ACPF surrogate that simultaneously captures continuous operating variability (varying load levels, PV outputs, and dispatchable DER and ESS setpoints) and discrete switch decisions (binary statuses), enabling fast evaluation of reconfigurable ADN states rather than restricting learning to a single fixed feeder configuration.
2. **Topology-filtered scenario generation with explicit connectivity and radiality checks:** We design a data-generation pipeline that samples a wide range of ADN scenarios and filters topologically inadmissible switching states using graph-theoretic connectivity and radiality checks. This ensures that the training set contains only connected, radial configurations and prevents invalid topologies from entering the learning process.
3. **Compact DNN surrogate with high-fidelity ACPF prediction and sub-millisecond batched inference:** Using NR solutions generated in Pandapower as references, the proposed surrogate predicts bus-voltage magnitudes, fixed-branch active-power flows, and total feeder losses with low prediction error while providing low-latency batched inference for rapid screening and many-query optimization studies.
4. **Transparent benchmarking against learning-based and physics-based baselines:** In addition to comparison with a conventional AC solver, we benchmark the surrogate against (i) XGBoost [30], (ii) random forest (RF) [31], and (iii) LinDistFlow, a linearized distribution-

flow approximation derived from the DistFlow family [32], thereby clarifying accuracy–runtime tradeoffs under high-DER and topology-changing scenarios.

The remainder of the paper is organized as follows. Section 2 presents the system model and topology-dependent ACPF formulation. Section 3 describes scenario generation, topology validation, and DNN training. Section 4 evaluates prediction accuracy, distribution-shift sensitivity, prediction confidence, computational efficiency, baseline performance, practical integration, and scalability. Section 5 discusses future research directions, and Section 6 concludes the paper.

2- System Model and Problem Formulation

This section formalizes the ADN model used (i) to generate the training set and (ii) to define the nonlinear mapping learned by the proposed surrogate. As illustrated in Fig. 1, the overall workflow comprises two stages: (1) sampling operating conditions and switch configurations, verifying topological admissibility using graph-theoretic criteria, and solving the steady-state ACPF using Pandapower; and (2) training a DNN to emulate the topology-dependent ACPF mapping for fast prediction of bus voltages, branch flows, and total losses. Throughout this study, the modified IEEE 33-bus feeder is represented by a balanced single-phase equivalent model. Accordingly, each bus is described by a single voltage phasor, and the formulation does not explicitly represent phase-specific load allocation, single-phase PV or other DER connections, missing-phase laterals, neutral-conductor effects, interphase mutual coupling, or voltage-unbalance indices. The errors reported in this study therefore quantify the approximation error of the neural surrogate relative to the balanced Pandapower ACPF reference model. They do not quantify the model-form error that may arise when a balanced equivalent is used to represent a practical unbalanced distribution feeder.

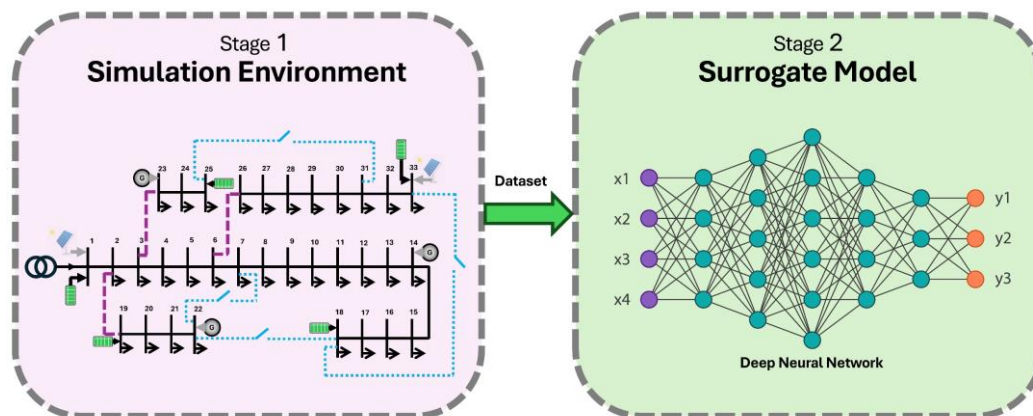


Fig. 1. Two-stage workflow for developing and applying the topology-dependent DNN surrogate

2-1- Network Model and Switching-Dependent Topology

The ADN is represented by a switching-dependent graph $\mathcal{G}(z) = (\mathcal{N}, \mathcal{E}(z))$, where $\mathcal{N} = \{1, \dots, n\}$ is the set of buses and bus 1 is the slack (substation) bus. The set of candidate branches is partitioned into fixed (always energized) branches $\mathcal{E}_{\text{fix}}$ and switchable branches $\mathcal{E}_{\text{sw}}$. Let $z \in \{0, 1\}^{|\mathcal{E}_{\text{sw}}|}$ denote the switch-status vector, where $z_e = 1$ indicates a closed (energized) switch and $z_e = 0$ indicates an open switch. The energized edge set is

$$\mathcal{E}(z) = \mathcal{E}_{\text{fix}} \cup \{e \in \mathcal{E}_{\text{sw}} : z_e = 1\}. \quad (1)$$

For a topology specified by the switch-status vector z , the corresponding bus-admittance matrix is denoted by $Y(z)$, whose elements are $Y_{ij}(z) = G_{Y,ij}(z) + j B_{Y,ij}(z)$. The matrix $Y(z)$ is assembled from the energized branch set $\mathcal{E}(z)$; therefore, switch actions modify its structure and entries. For compact power-balance notation, define $\mathcal{N}_i(z) = \{j \in \mathcal{N} | Y_{ij}(z) \neq 0\}$, including bus i through the self-admittance term. Thus, both $Y(z)$ and $\mathcal{N}_i(z)$ depend on the switch configuration, making the ACPF equations explicitly topology dependent.

2-2- ACPF Model

Let $V_i = |V_i| e^{j\theta_i}$ denote the complex voltage at bus i . Loads are modeled using scaling factors α_i applied to the nominal demands $(\bar{P}_i^L, \bar{Q}_i^L)$. Under the sign convention that positive injections correspond to net power injected into the network, the net nodal injections are

$$P_i^{\text{inj}} = P_i^{\text{DG}} + P_i^{\text{ESS}} - \alpha_i \bar{P}_i^L, \quad Q_i^{\text{inj}} = Q_i^{\text{DG}} + Q_i^{\text{ESS}} - \alpha_i \bar{Q}_i^L. \quad (2)$$

The slack bus provides the voltage reference:

$$|V_1| = \bar{V}_1, \quad \theta_1 = 0. \quad (3)$$

For a fixed topology z , expanding $S_i^{\text{inj}} = V_i I_i^*$ with $I_i = \sum_{j \in \mathcal{N}} Y_{ij}(z) V_j$ yields the standard AC nodal power-balance equations for all non-slack buses $i \in \mathcal{N} \setminus \{1\}$:

$$\begin{aligned} P_i^{\text{inj}} &= \sum_{j \in \mathcal{N}} |V_i| |V_j| \left(G_{Y,ij}(z) \cos(\theta_i - \theta_j) + B_{Y,ij}(z) \sin(\theta_i - \theta_j) \right), \forall i \in \mathcal{N} \setminus \{1\}, \\ Q_i^{\text{inj}} &= \sum_{j \in \mathcal{N}} |V_i| |V_j| \left(G_{Y,ij}(z) \sin(\theta_i - \theta_j) - B_{Y,ij}(z) \cos(\theta_i - \theta_j) \right), \forall i \in \mathcal{N} \setminus \{1\}. \end{aligned} \quad (4)$$

To define branch-level targets used by the surrogate, consider an energized branch $e = (i, j) \in \mathcal{E}(z)$ with series admittance y_e and resistance r_e . The branch current and complex power flow from i to j are

$$I_{ij} = y_e (V_i - V_j), \quad S_{ij} = P_{ij} + jQ_{ij} = V_i I_{ij}^*, \quad \forall (i, j) \in \mathcal{E}(z). \quad (6)$$

The corresponding active power loss on branch e and the total feeder loss are computed as

$$P_e^{\text{loss}} = r_e |I_{ij}|^2, \quad P_{\text{loss}}^{\text{tot}} = \sum_{e \in \mathcal{E}(z)} P_e^{\text{loss}}. \quad (7)$$

Eqs. (1)–(7) jointly define a topology-dependent ACPF mapping from operating conditions and switching decisions to steady-state voltages, branch flows, and losses.

2-3- Operational Limits and Admissible Operating Region

Training scenarios are generated within a feasible region defined by standard ADN limits. Load uncertainty is bounded by

$$\underline{\alpha}_i \leq \alpha_i \leq \bar{\alpha}_i, \forall i \in \mathcal{N}. \quad (8)$$

Voltage magnitudes are constrained by

$$\underline{V}_i \leq V_i \leq \bar{V}_i, \forall i \in \mathcal{N}. \quad (9)$$

Let $\mathcal{D} \subseteq \mathcal{N}$ and $\mathcal{S} \subseteq \mathcal{N}$ denote the sets of buses equipped with DG and ESS units, respectively. Their setpoints are restricted to the following capability ranges:

$$\begin{aligned} \underline{P}_i^{DG} \leq P_i^{DG} \leq \bar{P}_i^{DG}, \quad \underline{Q}_i^{DG} \leq Q_i^{DG} \leq \bar{Q}_i^{DG}, \forall i \in \mathcal{D}, \\ \underline{P}_i^{ESS} \leq P_i^{ESS} \leq \bar{P}_i^{ESS}, \quad \underline{Q}_i^{ESS} \leq Q_i^{ESS} \leq \bar{Q}_i^{ESS}, \forall i \in \mathcal{S}. \end{aligned} \quad (10)$$

For operational-feasibility assessment, each energized branch $e = (i, j) \in \mathcal{E}(z)$ is subject to the apparent-power limit

$$|S_{ij}| \leq \bar{S}_e, \quad |S_{ji}| \leq \bar{S}_e, \forall (i, j) \in \mathcal{E}(z). \quad (12)$$

Finally, each switch configuration is required to yield a connected radial feeder. Beyond reachability and cycle checks in the data pipeline, radiality can be expressed by the edge-count condition

$$|\mathcal{E}(z)| = |\mathcal{N}| - 1, \quad (13)$$

together with the requirement that all buses in $\mathcal{N}$ be reachable from the slack bus [33].

2-4- Surrogate Learning Problem

The surrogate is trained to approximate the implicit ACPF solution operator induced by Eqs. (1)–(7) over the admissible region Eqs. (8)–(13). The input vector aggregates operating conditions and topology decisions:

$$x = [\alpha, P^{DG}, Q^{DG}, P^{ESS}, Q^{ESS}, z]. \quad (14)$$

To ensure a fixed-dimensional learning target across different switching states, we include active power flows only for the fixed physical branches, arranged in a predefined order. The switchable connections E_{sw} are modeled as ideal switches and are therefore excluded from the learning target. The output vector is thus

$$\mathbf{y} = \left[|V_1|, \dots, |V_n|, P_{e_1}, \dots, P_{e_{|\mathcal{E}_{\text{fix}}|}}, P_{\text{loss}}^{\text{tot}} \right]. \quad (15)$$

Let n_α denote the number of load-scaling variables. The input vector contains n_α load factors, active and reactive power setpoints for each DG and ESS unit, and one binary status for each switchable branch. Its dimension is therefore

$$d_x = n_\alpha + 2|\mathcal{D}| + 2|\mathcal{S}| + |\mathcal{E}_{\text{sw}}|. \quad (15a)$$

Accordingly, $\mathbf{x} \in \mathbb{R}^{d_x - |\mathcal{E}_{\text{sw}}|} \times \{0,1\}^{|\mathcal{E}_{\text{sw}}|}$. The output vector contains one voltage magnitude for each bus, one active power flow for each fixed physical branch, and one scalar value representing the total feeder loss. Hence,

$$d_y = n + |\mathcal{E}_{\text{fix}}| + 1, \quad \mathbf{y} \in \mathbb{R}^{d_y}. \quad (15b)$$

When one load factor is assigned to each non-slack bus, $n_\alpha = n - 1$. The feeder-specific values of these dimensions are given in Section 3.2.

Let $\Phi(\cdot)$ denote the ACPF mapping defined by solving Eqs. (4)–(5) and post-processing with Eqs. (6)–(7), i.e., $\mathbf{y} = \Phi(\mathbf{x})$. A DNN $f_\theta(\cdot)$ is trained to approximate this mapping, $\hat{\mathbf{y}} = f_\theta(\mathbf{x})$, by minimizing a weighted regression loss over N_s training samples:

$$\min_{\theta} \frac{1}{N_s} \sum_{k=1}^{N_s} \left(w_V \|\mathbf{y}_V^{(k)} - \hat{\mathbf{y}}_V^{(k)}\|_2^2 + w_P \|\mathbf{y}_P^{(k)} - \hat{\mathbf{y}}_P^{(k)}\|_2^2 + w_L |y_L^{(k)} - \hat{y}_L^{(k)}|^2 \right) + \lambda \|\theta\|_2^2, \quad (16)$$

Here, $\mathbf{y}_V$ denotes the subvector of bus voltage magnitudes, $\mathbf{y}_P$ the subvector of branch active power flows, and $y_L = P_{\text{loss}}^{\text{tot}}$. The weights w_V, w_P, w_L balance the heterogeneous output scales, and λ is an ℓ_2 regularization coefficient [34].

3- Data Generation and DNN Methodology

This section details the dataset construction and the DNN training procedure used to learn the topology-dependent ACPF mapping defined in Section 2. For each scenario, the inputs comprise load-scaling factors applied to the nominal active and reactive demands, DG setpoints, ESS setpoints, and the switch-

status vector. The corresponding outputs are obtained from the ACPF solutions and comprise bus voltage magnitudes $|V_i|$, branch active power flows P_{ij} , and total feeder active power loss $P_{\text{tot}}^{\text{loss}}$. The overall workflow follows the two-stage pipeline illustrated in Fig. 1.

3-1- Data Generation Pipeline

The base network is a modified IEEE 33-bus radial distribution feeder [22]. To emulate an ADN with high DER penetration and reconfiguration capability, the benchmark is augmented with five DG units (three diesel units with a maximum active-power output of 0.3 MW each and two PV units with a maximum active-power output of 0.5 MW each), five ESS units, each rated at 0.2 MW with a maximum energy capacity of 1 MWh, and seven controllable switches (four normally open tie switches and three normally closed sectionalizing switches). The resulting ADN layout, including DER placements and controllable switches, is shown in Fig. 2.

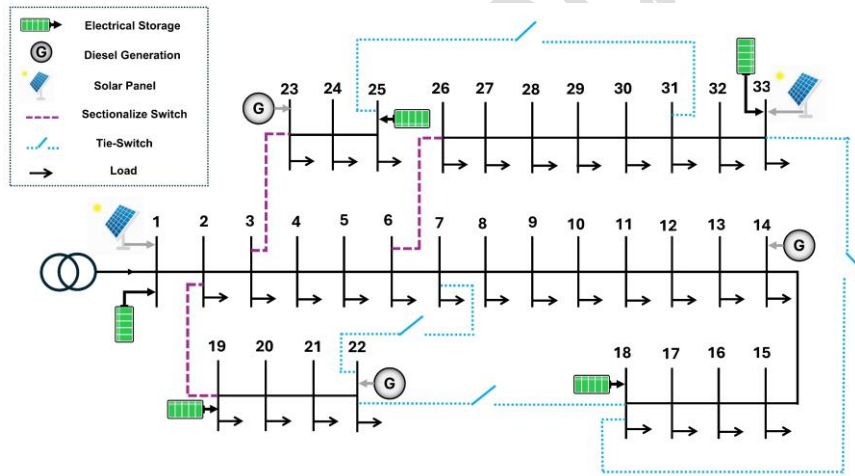


Fig. 2. Modified IEEE 33-bus feeder used in this study

Probabilistic sampling generated 864,900 candidate scenarios, from which 100,000 feasible operating scenarios were retained, covering 15 admissible radial topologies. Of the sampled cases, 101,280 satisfied the connectivity and radiality requirements, whereas 1,280 of these cases subsequently failed to converge with the NR solver. The complete sequential data-generation process required 1.10 h, corresponding to 39.6 ms per retained scenario; the mean NR solution time was 24.0 ms per converged case. This represents a one-time offline computational cost that is amortized over subsequent many-query evaluations. Consistent with the variable definitions in Section 2, load uncertainty is represented

by α_i , DG dispatch by P_i^{DG} and Q_i^{DG} , ESS dispatch by P_i^{ESS} and Q_i^{ESS} , and reconfiguration by the switch statuses z_e . Table 1 summarizes the adopted sampling distributions and ranges for all input components.

Table 1. Sampling distributions and ranges for input variables

Variable	Distribution	Range
Load scaling factors α_i	Uniform	[0.8, 1.2] p.u.
DG active outputs P_i^{DG}	Uniform	[0, 0.3] MW (diesel), [0, 0.5] MW (PV)
DG reactive outputs Q_i^{DG}	Uniform	$Q_i^{DG} \sim \mathcal{U}\left[-Q_{i,\max}^{DG}\left(P_i^{DG}\right), Q_{i,\max}^{DG}\left(P_i^{DG}\right)\right]$ with capability circle ($pf_{\min} = 0.9$)
ESS active setpoints P_i^{ESS}	Uniform	[-0.2, 0.2] MW
ESS reactive setpoints Q_i^{ESS}	Uniform	$Q_i^{ESS} \sim \mathcal{U}\left[-Q_{i,\max}^{ESS}\left(P_i^{ESS}\right), Q_{i,\max}^{ESS}\left(P_i^{ESS}\right)\right]$ with capability circle ($pf_{\min} = 0.9$)
Switch states z_e	Random binary choice for tie switches and sectionalizing switches	[0, 1]

To ensure that sampled reactive setpoints remain physically plausible, DGs and ESS units are assigned an apparent-power capability envelope derived from their active-power rating and a minimum power factor $pf_{\min}$. For each DER with rated active power $P_{\max}$, the corresponding apparent-power rating is defined as $S_{\text{rated}} = P_{\max} / pf_{\min}$, and the reactive power is bounded by the standard circular capability constraint

$$|Q| \leq \sqrt{S_{\text{rated}}^2 - P^2}. \quad (17)$$

Accordingly, after sampling P , the corresponding Q is sampled uniformly from $[-Q_{\max}(P), Q_{\max}(P)]$. For ESS units, $|P^{ESS}|$ is used in the capability relation to accommodate both charging and discharging setpoints. Loads are modeled at a constant power factor at each bus by scaling both the nominal active and reactive demands with the same α_i ; thus, the load at bus i is $(\alpha_i P_i^L, \alpha_i Q_i^L)$, where P_i^L and Q_i^L denote the nominal demands.

Candidate switch configurations are filtered before ACPF simulation to retain only connected, radial topologies. Following the graph model of Section 2, the energized topology is constructed by treating fixed branches and closed switches as edges. Connectivity is verified using a depth-first search (DFS) from the slack bus to confirm that all buses are reachable [33]. Radiality is enforced using cycle detection together with the tree edge-count condition. Configurations that violate connectivity or radiality are discarded, ensuring that the dataset contains only connected, radial feeder configurations. For each retained scenario, the network model is instantiated using the sampled load factors, DG setpoints, ESS setpoints, and switch-status vector. The ACPF equations are then solved in Pandapower using the NR algorithm. The resulting targets comprise voltage magnitudes at all buses, signed active power flows on the 29 fixed physical branches recorded in a predefined branch order and reference direction, and the total active power loss over all energized branches. Flows on the seven switchable connections are not included in the target vector. The complete dataset is divided into 80% training, 10% validation, and 10% test subsets. The input features and output targets are standardized using scaling parameters derived from the training set.

3-2- MLP Architecture and Training

The surrogate is implemented as a fully connected feedforward multilayer perceptron (MLP). The MLP was selected because the present study considers a feeder-specific setting in which the buses, candidate branches, controllable devices, and input and output orderings remain fixed. Topological variation is restricted to seven predefined controllable switches and can therefore be represented directly by the binary switch-status vector appended to the continuous operating variables. Under these conditions, a fixed-dimensional MLP provides a simple implementation, efficient batched CPU inference, and straightforward integration into repeated screening and optimization routines. The objective is not to establish the superiority of MLPs over graph neural networks (GNNs), but to evaluate whether explicit switch-status encoding can provide an accurate and computationally efficient surrogate for the considered reconfigurable feeder. Active and reactive power setpoints of the DG and ESS units are included directly in the input vector, enabling the model to capture the effects of varying operating conditions and admissible radial topologies on voltages, active power flows, and losses.

For the modified IEEE 33-bus feeder, $n_\alpha = 32$, $|\mathcal{D}| = 5$, $|\mathcal{S}| = 5$, $|\mathcal{E}_{\text{sw}}| = 7$, and $|\mathcal{E}_{\text{fix}}| = 29$. The implemented input vector is

$$\mathbf{x} = \left[\alpha_2, \dots, \alpha_{33}, (\mathbf{P}^{\text{DG}})^\top, (\mathbf{Q}^{\text{DG}})^\top, (\mathbf{P}^{\text{ESS}})^\top, (\mathbf{Q}^{\text{ESS}})^\top, \mathbf{z}^\top \right]^\top \in \mathbb{R}^{52} \times \{0,1\}^7,$$

whereas the output vector is

$$\mathbf{y} = \left[\mathbf{y}_V^\top, \mathbf{y}_P^\top, y_L \right]^\top \in \mathbb{R}^{63}.$$

Thus, the surrogate contains 59 input features and 63 output variables. The dimensional composition, physical units, and fixed ordering of the input and output components are summarized in Table 2.

Table 2. Dimensional composition of the surrogate input and output vectors.

Vector	Feature block and fixed ordering	Dimension	Type/unit
Input $\mathbf{x}$	Load factors: $\alpha_2, \dots, \alpha_{33}$	32	Continuous, p.u.
	DG active power setpoints: $\mathbf{P}^{\text{DG}}$	5	Continuous, MW
	DG reactive power setpoints: $\mathbf{Q}^{\text{DG}}$	5	Continuous, MVar
	ESS active power setpoints: $\mathbf{P}^{\text{ESS}}$	5	Continuous, MW
	ESS reactive power setpoints: $\mathbf{Q}^{\text{ESS}}$	5	Continuous, MVar
	Switch statuses: $\mathbf{z}$	7	Binary $\{0,1\}$
	Total input dimension	59	52 continuous + 7 binary
Output $\mathbf{y}$	Bus-voltage magnitudes: $ V_1 , \dots, V_{33} $	33	p.u.
	Fixed-branch active power flows: P_e , $e \in \mathcal{E}_{\text{fix}}$	29	MW
	Total feeder active power loss: $P_{\text{loss}}^{\text{tot}}$	1	MW
	Total output dimension	63	-

As shown in Table 2, the same bus, device, switch, and branch ordering is retained throughout data generation, training, and inference. A single load factor α_i scales both the nominal active and reactive demands at bus i . The 33 voltage outputs include the slack-bus voltage magnitude, whereas the 29 branch-flow outputs correspond only to the fixed physical branches. Flows on the four tie switches and three sectionalizing switches are excluded, thereby preserving a fixed output dimension under topology changes.

The fixed-dimensional representation also limits the transferability of the adopted MLP. Unlike GNNs, the MLP does not explicitly exploit network adjacency and does not naturally accommodate changes in the number or identity of buses, branches, switches, or controllable devices. Consequently, the trained model is not expected to transfer directly to a structurally different feeder. Such structural modifications would generally require input and output remapping, representative new training data, and model retraining.

The 59 input features represent physically defined operating variables rather than 59 independent historical datasets. During inference, the feature vector is constructed by ordering and standardizing load estimates or forecasts, DG and ESS measurements or setpoints, and switch-status data. Therefore, the principal offline challenge is obtaining representative coverage of the joint operating-point and topology space, rather than constructing the feature vector itself.

The hidden stack consists of three layers with 512, 256, and 128 neurons, respectively, using ReLU activation functions. A dropout rate of 0.3 is applied to the first two hidden layers to reduce overfitting [35], and the output layer uses a linear activation function for multivariate regression. The full layer-by-layer specification is provided in Table 3.

Table 3. DNN surrogate model architecture

Layer Type	Neurons	Activation	Dropout
Input	59	-	-
Hidden 1	512	ReLU	0.3
Hidden 2	256	ReLU	0.3
Hidden 3	128	ReLU	-
Output	63	Linear	-

Training follows the supervised regression objective described in Section 2. The model is optimized using Adam, with mean squared error (MSE) as the training loss and mean absolute error (MAE) reported as an additional metric. Training is performed with a batch size of 32 for up to 100 epochs, with the validation set used for hyperparameter tuning and monitoring of generalization performance [36].

4- Results and Analysis

This section evaluates the performance of the proposed DNN surrogate for topology-dependent ACPF approximation in the modified IEEE 33-bus ADN. The assessment uses the dataset and trained model described in Section 3. All experiments were conducted on a computing platform equipped with a 64-bit Intel Core i7-12650H processor operating at 2.30 GHz and 16 GB of RAM. The evaluation examines dataset characteristics, training dynamics, prediction accuracy, distribution-shift sensitivity, prediction confidence, computational

efficiency, baseline comparisons, and practical scalability. The results assess the extent to which the surrogate can approximate the balanced Pandapower ACPF reference solutions with low inference latency under the operating conditions and topology set considered in this study.

4-1- Dataset Statistics

The data-generation process yielded 100,000 feasible operating scenarios, partitioned into 80,000 training samples, 10,000 validation samples, and 10,000 test samples. Each sample contains 59 input features and 63 output variables, representing the operating conditions, topology status, and corresponding ACPF quantities defined in Section 3.

The retained dataset spans the specified operating ranges and the admissible radial topologies represented in the study. Load-scaling factors are sampled uniformly over the specified range and have a mean close to 1.0 p.u. in the retained dataset. DG outputs are distributed across their rated intervals, namely [0, 0.3] MW for diesel units and [0, 0.5] MW for PV units. ESS active-power setpoints are sampled over both positive and negative values, representing discharging and charging operation, respectively. Randomly sampled switch statuses are subjected to the connectivity and radiality checks described in Section 3.1, yielding 15 admissible radial topologies. For the outputs, bus-voltage magnitudes typically lie within 0.95–1.05 p.u., while total feeder losses range from approximately 0.05 to 1.76 MW across the retained scenarios. The branch active power flows also span a broad range of positive and negative values under the sampled operating conditions and topologies. Lower-loss scenarios generally occur under moderate loading and favorable local balancing of demand by DER and ESS injections, whereas higher-loss scenarios are associated with heavier loading, greater reactive-power demand, or radial configurations that increase feeder currents.

Table 4. Training-size sensitivity of the proposed DNN surrogate

Training samples	Voltage MAE (p.u.)	Flow MAE (MW)	Loss MAE (MW)
5,000	1.42×10^{-3}	0.0346	0.0184
20,000	8.41×10^{-4}	0.0217	0.0115
80,000	6.20×10^{-4}	0.0154	0.0079

A training-size sensitivity analysis was conducted using nested subsets of 5,000, 20,000, and 80,000 training samples while retaining the same validation and test sets. As shown in Table 4, all three error measures decrease as the number of training samples increases, with progressively smaller improvements at larger dataset sizes. No

exact input-vector duplicates were found between the training and test sets. Moreover, although the 15 admissible topology classes are represented in the dataset, each input vector also contains 52 continuously valued operating variables; therefore, the 100,000 scenarios do not enumerate the continuous operating space. The observed test-set accuracy and training-size trend consequently support the interpretation that the DNN approximates the continuous ACPF relationship within the represented topology classes rather than performing exact input–output lookup.

4-2- Training Dynamics and Evaluation Metrics

The proposed surrogate learns a nonlinear mapping from the 59-dimensional input vector, comprising load-scaling factors, DG and ESS active and reactive power setpoints, and switch statuses, to 63 ACPF output quantities. The outputs comprise 33 bus-voltage magnitudes, 29 fixed-branch active power flows, and one scalar representing total feeder active power loss. The output targets were standardized during training to improve numerical conditioning, and all reported errors were calculated after inverse transformation to physical units.

We report four complementary metrics to characterize predictive accuracy and empirical error behavior. Mean absolute error (MAE) summarizes the typical absolute deviation in physical units. Mean squared error (MSE) assigns greater weight to larger deviations and therefore reflects sensitivity to relatively high-error predictions. For each output group, $\text{Max}|\text{error}|$ denotes the largest observed absolute error over the evaluated test set and is defined as

$$\text{max}|\text{error}| = \max_{n,j} |y_{n,j} - \hat{y}_{n,j}|,$$

where n indexes the test scenarios, j indexes the outputs in the relevant output group, $y_{n,j}$ is the NR reference value, and $\hat{y}_{n,j}$ is the corresponding surrogate prediction. This finite test-set maximum is an empirical statistic and should not be interpreted as a mathematical or operational guarantee for unseen or out-of-distribution scenarios. The coefficient of determination, R^2 , summarizes goodness of fit by quantifying how well the surrogate reproduces the variability of the reference ACPF solutions.

Fig. 3 shows the training and validation MSE in standardized output space together with the corresponding MAE values across training epochs. The training and validation curves decrease smoothly and remain closely aligned, indicating stable optimization without an evident widening

generalization gap over the selected training horizon. Both training and validation metrics decrease rapidly during the initial epochs and then improve more gradually. The absence of a widening gap between the training and validation curves is consistent with stable generalization on the validation set. The slightly lower validation values observed during some epochs may partly reflect the fact that dropout is active during training but disabled during standard validation inference.

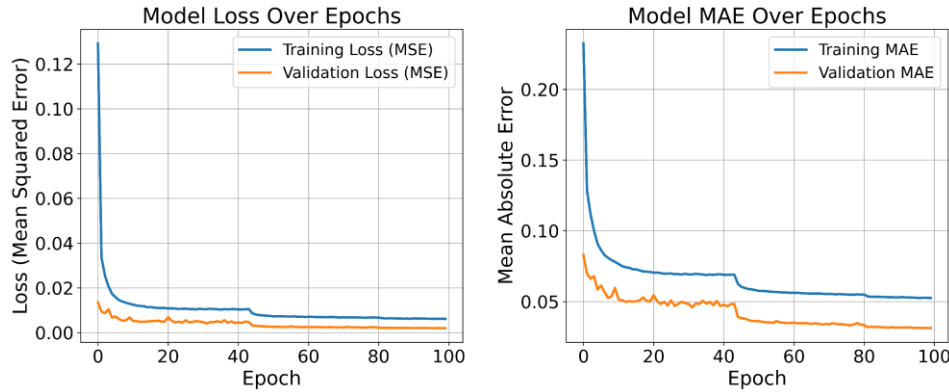


Fig. 3. Training and validation MSE and MAE across training epochs

4-3- Prediction Accuracy and Empirical Error Analysis

Table 5 summarizes the surrogate accuracy in physical units on the 10,000-scenario test set. Across all output groups, the surrogate predictions show close agreement with the NR reference, with low average errors and the empirical maximum deviations reported in Table 5. In addition to MAE and MSE, the reported $\text{Max}|\text{error}|$ values characterize the largest absolute prediction errors observed in the test set. These values can inform the selection of preliminary screening margins, but they do not guarantee the same error limits for unseen operating conditions or topologies.

Table 5. Surrogate prediction accuracy in physical units on the test set

Output	MAE	MSE	$\text{Max} \text{error} $	R^2
Voltage magnitudes (p.u.)	6.20×10^{-4}	1.10×10^{-6}	2.10×10^{-3}	0.99971
Active power flows (MW)	0.0154	6.15×10^{-4}	0.252	0.99946
Total feeder loss (MW)	0.0079	1.72×10^{-4}	0.154	0.99912

Fig. 4 compares the NR reference total feeder losses with the DNN predictions for 100 randomly selected test scenarios. The predicted trace nearly overlaps the NR reference over the full loss range (approximately 0.05–1.76 MW), including stressed high-loss cases. Together with the low MAE and MSE reported in Table 5, this agreement indicates that the surrogate accurately reproduces the nonlinear

variation of feeder losses across the evaluated operating conditions, including higher-loss cases associated with increased feeder currents.

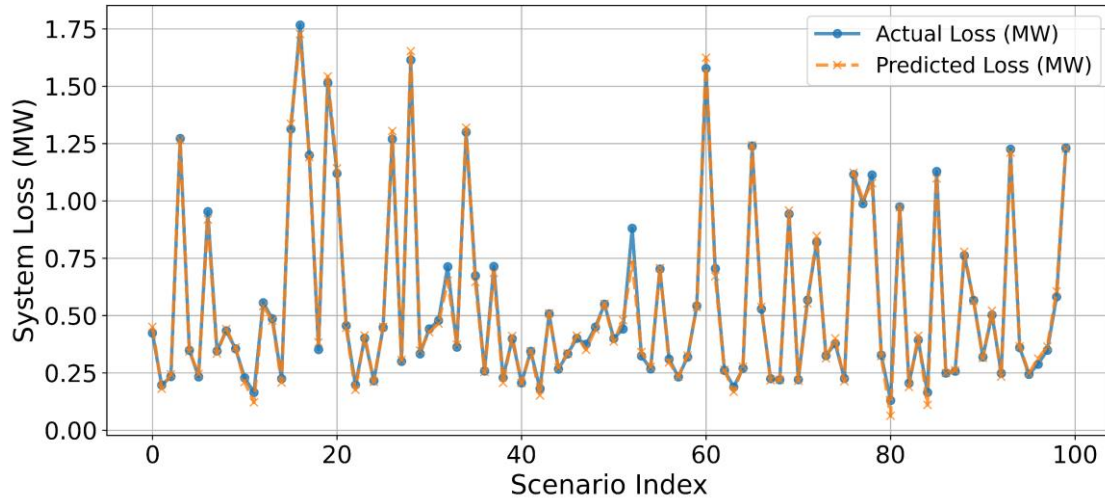


Fig. 4. NR reference and DNN-predicted total feeder losses for 100 randomly selected test scenarios.

As shown in Fig. 4, the DNN predictions closely track the NR reference across the selected test scenarios, including cases with relatively high feeder losses. These higher-loss cases correspond to operating points with increased feeder currents due to heavier loading, less favorable radial configurations, or greater reactive-power circulation. Accurate tracking across both lower- and higher-loss operating conditions is particularly important for applications such as feeder reconfiguration and DER dispatch, where feeder loss is often included in the objective function.

The voltage-magnitude predictions closely follow the NR reference values and preserve the spatial voltage-drop pattern along the feeder. As illustrated in Fig. 5, the DNN-predicted voltage profiles for five randomly selected test scenarios closely follow the NR reference profiles across all 33 buses, including the minimum-voltage locations toward the end of the feeder. The largest observed voltage-magnitude error on the test set is 2.10×10^{-3} p.u. (Table 5); this finite empirical maximum can inform the selection of a preliminary voltage-screening margin but should not be interpreted as a guaranteed bound for unseen cases.

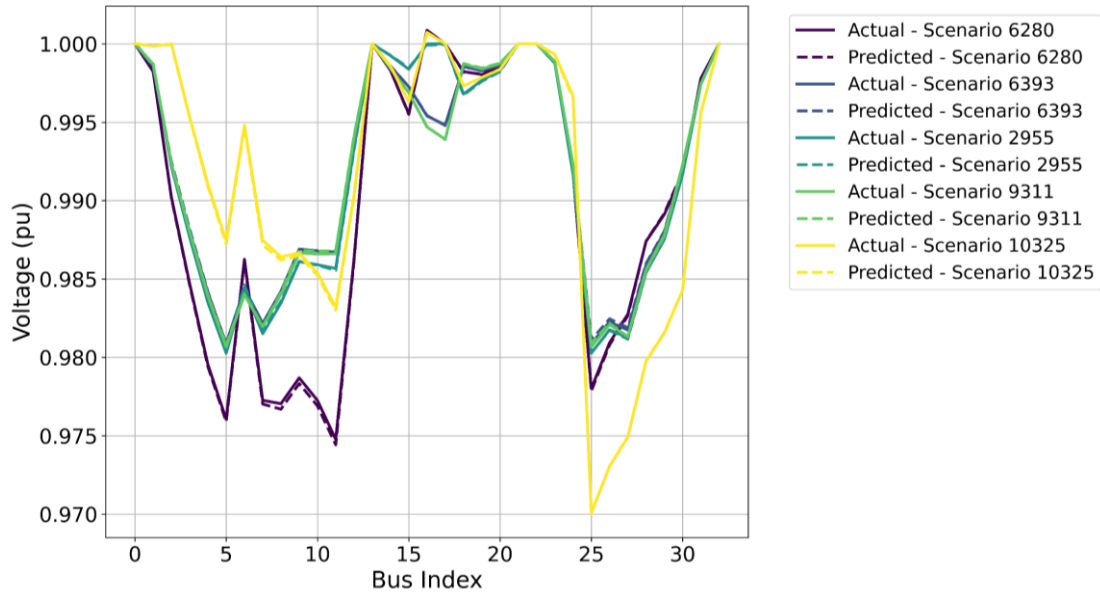


Fig. 5. NR reference and DNN-predicted bus-voltage magnitudes for five randomly selected test scenarios.

Fig. 5 further shows that the surrogate preserves the feeder voltage-profile shape across the illustrated operating points. The closest agreement is observed across the mid-feeder buses, while the larger deviations occur near the minimum-voltage locations toward the end of the feeder, where voltages are more sensitive to loading and reactive-power variations. At these locations, the predicted profiles remain closely aligned with the NR reference, supporting the use of the surrogate for preliminary voltage screening when an empirical margin informed by the observed $\text{Max}|\text{error}|$ is applied.

Branch active power predictions show close agreement with the NR reference across the evaluated operating conditions. Fig. 6 demonstrates close tracking of both flow magnitude and direction for five randomly selected test scenarios, including sign reversals associated with reverse power flow under high DER output. The largest observed absolute flow error is $\text{Max}|\text{error}| = 0.252 \text{ MW}$ (Table 5). This value represents the empirical maximum over the evaluated test set and should not be interpreted as a guaranteed error bound for unseen scenarios.

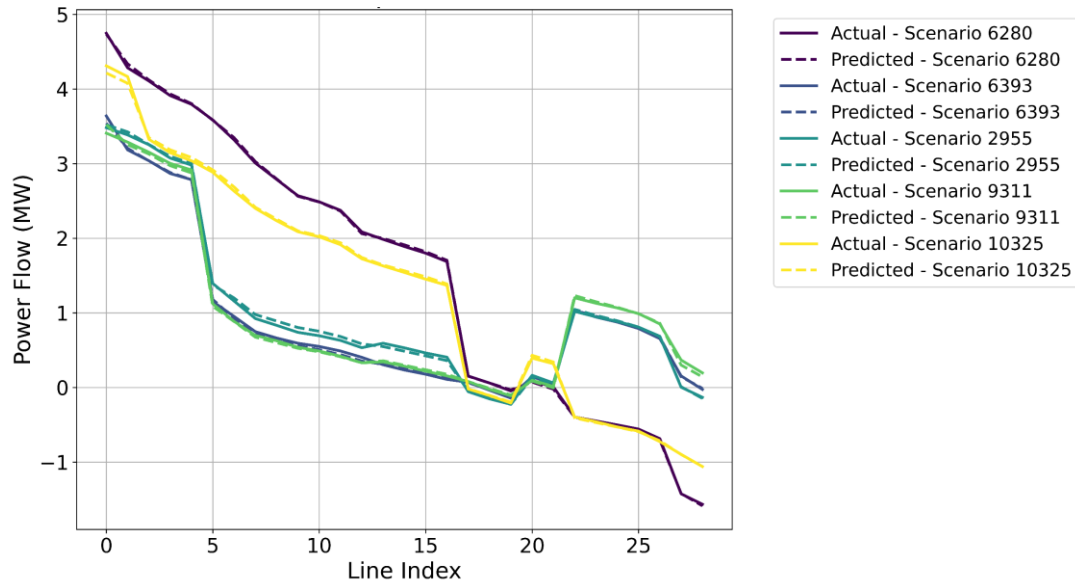


Fig. 6. NR reference and DNN-predicted active power flows for five randomly selected test scenarios.

In Fig. 6, the DNN reproduces the branch-flow patterns of the NR reference, including abrupt changes at branching points and variations in downstream flow magnitude. Importantly, the surrogate captures sign reversals (negative flows) in scenarios with substantial DER injection, which is important for applications that depend on correct flow direction, such as congestion management and protection-aware dispatch. Residual mismatches are most noticeable at the larger flow magnitudes in the illustrated scenarios; the overall test-set Max|error| is reported in Table 5. The accuracy results above should be interpreted within the balanced-model scope defined in Section 2. In a strongly unbalanced feeder, phase-specific voltage sensitivities, current paths, and branch loadings may differ considerably from their balanced-equivalent values. Therefore, the present surrogate should not be applied directly to an unbalanced field feeder without generating three-phase ACPF training data and redesigning the input and output representations to preserve phase identity.

4-4- Distribution-Shift Sensitivity

Distribution-shift sensitivity was assessed using the original 10,000-scenario test set, generated from the same sampling distributions as the training data, as the reference. Without retraining, the surrogate was then evaluated on three additional 10,000-scenario test sets independently labeled using NR ACPF solutions. The correlated in-envelope set introduced common feeder-level load variation, correlated PV production, and coordinated ESS operation within the original training bounds. The stressed in-envelope set concentrated loads in the 1.10–1.20 p.u. range with low PV output and predominantly

charging ESS operation, whereas the mild out-of-envelope set extended load factors mainly to 1.20–1.30 p.u. as an extrapolation diagnostic. The original training-set scalers were used, and only feasible operating scenarios were retained.

Table 6. Distribution-shift sensitivity of the proposed DNN surrogate

Test regime	Voltage MAE (p.u.)	Flow MAE (MW)	Loss MAE (MW)	Voltage R^2	Flow R^2	Loss R^2
Original test set	6.20×10^{-4}	0.0154	0.0079	0.99971	0.99946	0.99912
Correlated in-envelope	7.35×10^{-4}	0.0183	0.0097	0.99955	0.99921	0.99868
Stressed in-envelope	1.18×10^{-3}	0.0289	0.0161	0.99891	0.99808	0.99641
Mild out-of-envelope	2.74×10^{-3}	0.0618	0.0365	0.99380	0.98720	0.96950

As shown in Table 6, the correlated in-envelope cases exhibit only modest accuracy degradation relative to the original test set, whereas prediction errors increase more noticeably under stressed but still in-envelope conditions. The largest degradation occurs for the mild out-of-envelope cases, with increased MAEs and reduced R^2 values. These results indicate that the surrogate retains high accuracy under moderate distribution shifts within the represented operating region, whereas out-of-envelope operating points should instead be evaluated using the complete ACPF solver.

4-5- Prediction-Confidence Analysis

To assess whether prediction uncertainty can identify difficult operating points, Monte Carlo dropout was applied to the final trained surrogate without retraining. The existing dropout layers remained active during inference, and 50 stochastic forward passes were performed for each scenario in the original test set and the three distribution-shift sets of Section 4.4. The predictive mean was used as the point estimate and the predictive standard deviation as an epistemic-uncertainty proxy. The model weights and training-set scalers were unchanged, and all statistics were calculated in physical units after inverse transformation. The error–uncertainty relationship was quantified using the Pearson correlation coefficient between the scenario-level predictive standard deviation and the corresponding absolute prediction error relative to the NR reference; larger positive values indicate that higher predictive uncertainty is associated with larger surrogate errors.

Table 7. Prediction-confidence analysis using Monte Carlo dropout

Test regime	Mean predictive standard deviation			Pearson correlation		
	Voltage (p.u.)	Flow (MW)	Loss (MW)	Voltage	Flow	Loss
Original test set	3.20×10^{-4}	0.0081	0.0045	0.61	0.67	0.71
Correlated in-envelope	4.00×10^{-4}	0.0100	0.0058	0.64	0.70	0.73
Stressed in-envelope	6.80×10^{-4}	0.0176	0.0107	0.70	0.75	0.79

Mild out-of-envelope	1.52×10^{-3}	0.0395	0.0248	0.76	0.81	0.84
----------------------	-----------------------	--------	--------	------	------	------

As shown in Table 7, predictive uncertainty increases under stressed and out-of-envelope operating conditions and is positively associated with the corresponding prediction error relative to the NR reference. The mild out-of-envelope set exhibits the highest uncertainty and strongest Pearson correlations. Thus, Monte Carlo dropout can support relative confidence ranking and help identify cases that should be evaluated using the complete ACPF solver. However, these uncertainty estimates are not calibrated probabilistic guarantees and do not replace complete ACPF validation.

4-6- Computational Efficiency

Inference runtime was measured on the same CPU platform using the trained model under two evaluation modes: (i) single-scenario inference, where each operating point is evaluated independently, and (ii) batched inference, where many scenarios are evaluated together. The single-scenario latency is representative of interactive or event-driven use (e.g., rapid checks during switching operations), whereas batched throughput is representative of Monte Carlo screening and repeated evaluations inside optimization loops.

Table 8 summarizes the measured CPU inference runtime under single-scenario and batched evaluation. For single-scenario inference (200 repeated calls), the average latency is 12.8 ms per scenario, corresponding to approximately 78.1 scenarios/s. For a batch of 1,000 scenarios, the total inference time is 540 ms, corresponding to 0.54 ms per scenario and approximately 1851.9 scenarios/s. These timings include practical Python-side overheads (data movement and call dispatch) and therefore provide a realistic estimate for integrating the surrogate into operational scripts. Compared with the Pandapower NR solver, which requires tens of milliseconds per scenario on the same platform, the surrogate provides a substantial speedup while maintaining high-fidelity ACPF approximations.

Table 8. CPU inference runtime for the trained surrogate

Mode	n	Time	Throughput (scenarios/s)
Single-sample	200	Mean 12.8 ms	78.1
Batch	1000	540 ms total; 0.54 ms per case	1851.9

The runtimes reported in Table 8 characterize only local surrogate evaluation on the stated computing platform, including Python-side data movement and model-call overhead. They exclude measurement acquisition and communication, state estimation or forecasting, data-quality checking, database access, topology processing, full-ACPF fallback calculations, and control-system or switching-actuation delays. Therefore, these values should not be interpreted as end-to-end operational decision times.

The observed gap between single-scenario and batched performance indicates that fixed overheads (such as input preparation and runtime dispatch) are more visible in the single-scenario setting, whereas batching amortizes these overheads and better reflects the underlying throughput of the neural-network forward pass. This behavior is advantageous in planning and operational studies that require thousands of power-flow evaluations, such as stochastic hosting-capacity assessment, topology search, and receding-horizon control.

4-7- Benchmarking Against Baseline Models

To complement the multi-output accuracy assessment in Subsection 4.3, this subsection benchmarks total-loss prediction against representative learning-based and physics-based baselines. The comparison focuses on total active loss because it is a scalar system-level metric frequently used in feeder reconfiguration, DER dispatch, and energy-storage scheduling. To ensure a consistent comparison, each baseline is trained or configured to predict total feeder loss using the same training set and is evaluated on the same 10,000-scenario test set. The proposed DNN is included in this benchmark as the full multi-output surrogate reported earlier.

The proposed DNN is compared with an XGBoost regressor, an RF regressor, and a LinDistFlow approximation. The two tree-ensemble baselines are trained on the same training set used for the DNN and evaluated on the same test set, ensuring that all reported errors correspond to identical operating scenarios. As strong tabular learners, XGBoost and RF are expected to perform well when the operating space is well covered by training data, although their piecewise structure can lead to larger tail errors under stressed conditions. In our implementation, XGBoost is configured as a boosted ensemble of 400 trees with a maximum depth of 6, a learning rate of 0.05, row and feature subsampling ratios of 0.9 and 0.9, respectively, and L_2 regularization; RF uses 400 trees with unconstrained depth. LinDistFlow is

implemented as a radial, linearized DistFlow surrogate following the simplified/linearized DistFlow formulation [32]; for each scenario, the energized feeder is formed by treating closed switches as zero-impedance connections and extracting a spanning tree from the slack bus, after which losses are recovered using the standard quadratic approximation. As a linearization that neglects higher-order loss terms, LinDistFlow is computationally efficient, but its accuracy typically degrades as loading increases and voltage drops or reactive circulation become significant. Reference loss values for all comparisons are obtained from the NR ACPF solution in Pandapower, and the loss-accuracy and runtime comparison is summarized in Table 9.

Table 9. Loss-prediction accuracy and runtime comparison with NR ACPF

Method	MAE (MW)	MSE (MW ²)	R^2	Time (ms)
DNN (proposed)	0.0079	1.72×10^{-4}	0.99912	0.54
XGBoost	0.0096	2.80×10^{-4}	0.99857	0.95
RF	0.0115	4.20×10^{-4}	0.99785	2.10
LinDistFlow	0.0185	9.20×10^{-4}	0.99529	0.38
NR ACPF	-	-	-	24.0

Table 9 shows that all data-driven surrogates achieve high goodness-of-fit, reflecting the well-covered training distribution. Among them, the proposed DNN provides the lowest MAE and MSE and the highest R^2 , indicating the most consistent tracking of AC losses across nominal and stressed operating points. XGBoost remains a strong baseline, as expected for structured regression, while RF exhibits slightly higher dispersion. LinDistFlow remains fast but less accurate because of its linearized formulation, with larger errors observed in high-loss regimes where I^2R losses become more pronounced.

In terms of runtime, the proposed DNN achieves 0.54 ms per scenario under batched evaluation, representing an approximately 44-fold speedup over the NR solver (24.0 ms/scenario). While LinDistFlow remains the fastest baseline due to its closed-form linear structure, its reduced fidelity under stressed conditions makes it less suitable for safety-critical screening without larger guard bands. Overall, among the evaluated methods, the proposed DNN provides a favorable accuracy–latency tradeoff for rapid screening and repeated objective evaluation prior to final ACPF validation.

Fig. 7 illustrates how loss estimates from the proposed DNN and the baseline models compare with the NR ACPF reference as operating conditions become more stressed. Under normal conditions, all

methods track the reference closely; however, differences become more pronounced at stressed operating points where approximation error is most likely to affect operational decisions. Across the full loss range, the proposed DNN remains tightly clustered around the reference line, indicating low bias and low dispersion even in high-loss cases. XGBoost also follows the reference well but shows a slight tail deviation at higher losses, whereas RF exhibits wider scatter consistent with its piecewise-constant partitioning of the input space. In contrast, LinDistFlow displays a systematic underestimation that increases with true loss, as expected from a linearized formulation that cannot fully capture quadratic current-driven loss behavior under heavy loading.

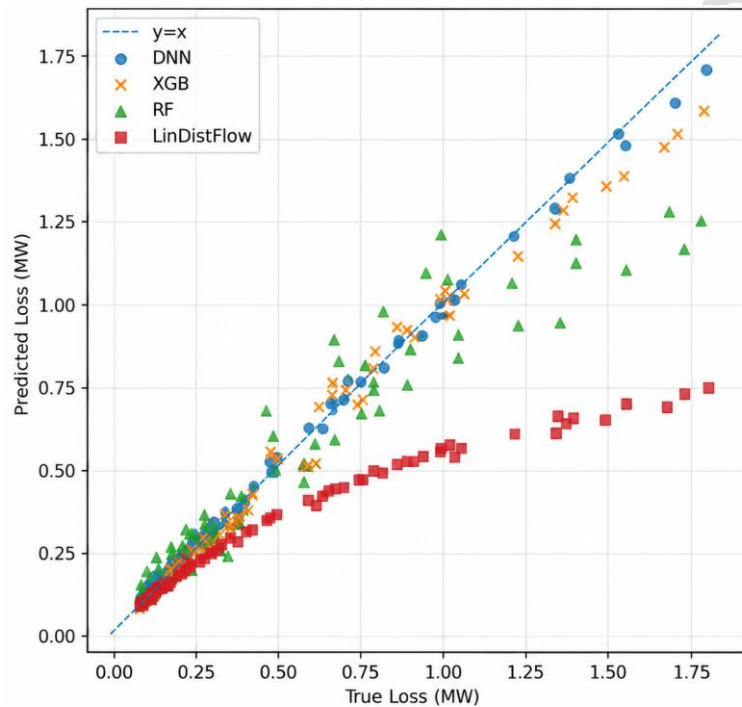


Fig. 7. Loss comparison over 100 test scenarios (DNN vs baselines)

4-8- Practical Integration and AC-Validation Workflow

The proposed surrogate is intended to operate as a rapid screening layer within a distribution-management, planning, or optimization workflow rather than as a replacement for the complete ACPF solver. For each candidate operating state, bus-load estimates obtained from distribution-system state estimation, smart-meter aggregation, forecasting, or pseudo-measurements are combined with DG and ESS active and reactive power setpoints obtained from the corresponding controllers or energy-management systems, together with switch statuses obtained from supervisory control and data

acquisition (SCADA) records. These quantities must be mapped to the same balanced single-phase feeder representation and fixed feature ordering used during training. Before inference, the continuous inputs should be checked against the training envelope, the candidate switch configuration should be verified as connected and radial, and the resulting 59-dimensional feature vector should be standardized using the training-set parameters. The surrogate predictions, 33 bus-voltage magnitudes, active power flows on 29 fixed physical branches, and total feeder active loss, can then be used to screen or rank candidate operating decisions.

Operating points that fall outside the training envelope, exhibit high predictive uncertainty, are insufficiently represented by the training data, or lie within a predefined safety margin of an operational limit should instead be evaluated using the complete ACPF solver. In feeder-reconfiguration applications, only a reduced set of promising candidates should be shortlisted using the surrogate; these candidates must subsequently be re-evaluated using full ACPF, and only an AC-validated solution should be considered for implementation. This final validation must verify voltage limits, device capability limits, and apparent-power or thermal limits on all energized branches. Such validation is essential because the surrogate predicts active-power flows only on the fixed physical branches; it does not predict reactive or apparent branch power and excludes flows on the seven switchable connections. Therefore, the surrogate outputs alone cannot provide a complete assessment of operational feasibility.

4-9- Scalability Considerations

Scalability should be considered separately in terms of surrogate dimensionality and coverage of the joint topology–operating-point space. From Eqs. (15a) and (15b), the input and output dimensions are

$$\dim(x) = n_{\alpha} + 2|\mathcal{D}| + 2|\mathcal{S}| + |\mathcal{E}_{\text{sw}}|,$$

$$\dim(y) = n + |\mathcal{E}_{\text{fix}}| + 1.$$

For sparse distribution feeders, if the numbers of DERs, ESS units, and controllable switches grow approximately proportionally with the number of buses, these dimensions increase approximately linearly with network size. With fixed hidden-layer widths, the parameter count and arithmetic cost of the input and output layers also grow approximately linearly. This dimensional observation does not

imply that accuracy or measured inference time will scale linearly, because larger feeders may require greater model capacity and the measured latency remains hardware- and implementation-dependent.

The more restrictive scalability issue is coverage of topology and operating-condition combinations.

With $|\mathcal{E}_{\text{sw}}|$ binary switches, the raw switching space contains $2^{|\mathcal{E}_{\text{sw}}|}$ status vectors, although only a subset produces connected radial feeders. For the seven-switch system studied here, this corresponds to 128 binary status vectors before topology filtering. As the number of controllable switches increases, exhaustive enumeration or uniform random sampling becomes increasingly inefficient. Representative training data must cover admissible topologies jointly with load variation, DG and ESS operation, reverse-power-flow conditions, and operating points close to voltage or loading limits. Otherwise, some admissible but infrequent topologies or stressed operating regions may remain poorly represented even when the total dataset is large.

Accordingly, the present results demonstrate computational viability for the modified IEEE 33-bus feeder but do not establish feeder-independent scalability. Studies on substantially larger feeders should consider topology-stratified and adaptive sampling, active learning, targeted generation of near-limit cases, modular or hierarchical feeder-zone surrogates, mixture-of-experts models, graph-based or physics-guided architectures, and transfer learning. These approaches should be combined with out-of-distribution detection and full ACPF evaluation for inputs that are insufficiently represented by the training data.

5- Future Work

Future studies should extend the present framework in four complementary directions. First, the balanced single-phase representation should be generalized to unbalanced three-phase distribution feeders by incorporating phase-indexed load, DG, and ESS injections, three-phase line-impedance matrices, missing-phase information, and phase-specific voltage and branch-power outputs. The output representation should also include active and reactive power flows on both fixed and switchable connections to support more complete operational-limit assessment. Practical distribution-system devices and controls, including inverter control modes, voltage regulators, switched capacitor banks,

neutral-conductor effects, and protection-related constraints, should also be incorporated. Second, scalability should be evaluated on larger feeders with more controllable switches and candidate branches. Topology-stratified or adaptive sampling could reduce redundant ACPF labeling, while modular or hierarchical surrogates, graph-based architectures, and transfer-learning methods may reduce the dependence on a fixed feeder structure and predefined feature ordering.

Third, the trustworthiness of the surrogate should be further strengthened through calibrated uncertainty quantification, formal out-of-distribution detection, broader distribution-shift testing, and physics- or constraint-aware training. These mechanisms should help identify unfamiliar or low-confidence operating points and direct near-limit or operationally critical cases to evaluation using the complete NR-based ACPF solver. Finally, the surrogate should be evaluated in applications such as feeder reconfiguration, coordinated DG and ESS dispatch, hosting-capacity assessment, stochastic scheduling, and real-time control. Such studies should assess not only prediction accuracy and inference time, but also end-to-end computational performance, robustness to measurement and parameter uncertainty, and the quality of the final AC-validated operating decisions. These extensions would support the development of a more transferable and reliable surrogate-based decision-support framework for practical active distribution networks.

6- Conclusion

This paper developed a DNN surrogate to accelerate steady-state ACPF analysis in active distribution networks subject to frequent operating changes and switching actions. A key contribution is the end-to-end data-generation framework that combines probabilistic sampling of loads and controllable resources with explicit topology validation to retain connected, radial network configurations. Using a modified IEEE 33-bus feeder with five DG units, five ESS units, and seven controllable switches, the approach generates 100,000 feasible operating scenarios. NR ACPF solutions generated in Pandapower provide the supervised-learning targets for DNN training.

The resulting MLP learns the nonlinear mapping from 59 input features, comprising load factors, DG and ESS active and reactive power setpoints, and switch statuses, to 63 outputs comprising bus-voltage

magnitudes, fixed-branch active power flows, and total feeder active power loss. Evaluation on held-out test cases shows that the surrogate preserves the overall voltage profile, captures reverse-flow behavior under high local generation, and tracks the nonlinear loss response with low average prediction errors. Batched inference provides a substantial computational speedup over repeated conventional ACPF solutions, making the surrogate suitable for rapid operating-point screening and repeated evaluations in reconfiguration or optimization routines. Nevertheless, the trained MLP is feeder-specific and should not be applied directly to a structurally different network without feature remapping, new representative data, and model retraining. Accordingly, the surrogate should be used for rapid candidate screening and ranking, while every final switching or dispatch decision must be confirmed using the complete ACPF model before implementation. The present computational and accuracy results are specific to the modified 33-bus feeder and should not be interpreted as evidence of feeder-independent scalability; substantially larger networks with many controllable switches would require more selective data generation and potentially more structured surrogate architectures.

Declarations

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics Declaration

This study is based solely on computational modeling and simulation. No human or animal subjects were involved, and no ethical approval was required.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Data Availability

The data-generation pipeline used to create the 100,000 scenarios and the modified IEEE 33-bus system parameters are available from the corresponding author upon reasonable request.

Author Contributions

Mohammad Yasin Rahmaty Darvazi: Conceptualization; data curation and pipeline development; methodology; software and validation; writing – original draft.

Hossein Afrakhte: Supervision; formal analysis; writing – review and editing.

Nomenclature

Bold lowercase symbols denote vectors and bold uppercase symbols denote matrices. An underbar and an overbar indicate lower and upper bounds, respectively; the superscript * denotes complex conjugation; and $j = \sqrt{-1}$ is the imaginary unit.

Symbols and notation

Symbol	Definition	Symbol	Definition
i, j, k	Bus indices and scenario index, respectively.	n, N_s	Number of buses and number of supervised-learning samples.
$e = (i, j)$	Candidate or energized branch connecting buses i and j .	$\mathcal{N}$	Set of buses; bus 1 is the slack (substation) bus.
$\mathcal{E}_{fix}$	Set of fixed, always-energized physical branches.	$\mathcal{E}_{sw}$	Set of switchable branches.

$\mathcal{E}(\mathbf{z})$	Set of energized branches under switch-status vector $\mathbf{z}$.	$\mathcal{G}(\mathbf{z})$	Switching-dependent network graph.
$\mathcal{N}_i(\mathbf{z})$	Buses contributing to the current injection at bus i , including i .	$\mathcal{D}, \mathcal{S}$	Sets of buses equipped with DG and ESS units, respectively.
$\mathbf{z}, \mathbf{z}_e$	Binary switch-status vector and status of switch e (1 closed; 0 open).	$\mathbf{Y}(\mathbf{z}), G_{Y,ij}, B_{Y,ij}$	Topology-dependent bus-admittance matrix and its conductance/susceptance entries.
V_i, V_i , θ_i	Complex voltage, voltage magnitude, and phase angle at bus i .	$\alpha_i; P_i^L, Q_i^L$	Load-scaling factor and nominal active/reactive demands at bus i .
P_i^{DG}, Q_i^{DG}	Active and reactive setpoints of the DG unit at bus i .	P_i^{ESS}, Q_i^{ESS}	Active and reactive setpoints of the ESS unit at bus i .
P_i^{inj}, Q_i^{inj}	Net nodal active and reactive power injections at bus i .	y_e, r_e	Series admittance and resistance of branch e .
I_{ij}	Complex current flowing from bus i to bus j .	$S_{ij} = P_{ij} + jQ_{ij}$	Complex, active, and reactive power flow from bus i to bus j .
$P_e^{loss}, P_{loss}^{tot}$	Active loss of branch e and total active feeder loss.	S_e^{max}	Apparent-power limit of energized branch e .
$P_{max}, S_{rated}, pf_{min}$	Rated active power, rated apparent power, and minimum power factor.	$Q_{max}(P)$	Maximum feasible reactive-power magnitude at active-power setpoint P .
Max error	Largest observed absolute prediction error in the reported test set.	R^2	Coefficient of determination.
$\mathbf{x}$	Input vector: load factors, DG/ESS P-Q setpoints, and switch statuses.	$\mathbf{y}, \hat{\mathbf{y}}$	Reference and surrogate-predicted output vectors.
$\mathbf{y}_V, \mathbf{y}_P, \mathbf{y}_L$	Voltage, fixed-branch active-flow, and total-loss output components.	$\Phi(\cdot)$	Topology-dependent AC power-flow solution operator.
$f_\theta(\cdot), \theta$	Neural surrogate and its trainable parameter vector.	w_V, w_P, w_L, λ	Output-group loss weights and L2-regularization coefficient.

Acronyms and abbreviations

Acronym	Definition	Acronym	Definition
ACPF	Alternating-current power flow	AC-OPF	Alternating-current optimal power flow
ADN	Active distribution network	AI	Artificial intelligence
CPU	Central processing unit	DER	Distributed energy resource
DG	Distributed generation	DNN	Deep neural network
DFS	Depth-first search	ESS	Energy storage system
GNN	Graph neural network	LinDistFlow	Linearized DistFlow approximation
MAE	Mean absolute error	MLP	Multilayer perceptron
MSE	Mean squared error	NR	Newton-Raphson
OPF	Optimal power flow	PV	Photovoltaic
ReLU	Rectified linear unit	RF	Random forest
SCADA	Supervisory control and data acquisition	p.u.	Per unit
XGBoost	Extreme gradient boosting		

References

- [1] International Energy Agency, *Renewables 2025: Analysis and Forecasts to 2030*, IEA, Paris, 2025.
- [2] K. Wang, Y. Huang, Y. Liu, T. Huang, S. Zang, A review of optimization scheduling for active distribution networks with high-penetration distributed generation access, *Energies*, 18 (2025) 4119.

- [3] S.P. Mirhoseini, S.M. Hashemi, B. Alizadeh, M. Tabarzadi, A comprehensive method for expansion planning of active distribution system considering reliability assessment, *AUT Journal of Electrical Engineering*, 57(1) (2025) 113-130. doi: 10.22060/eej.2024.23185.5593.
- [4] M.R. Behbahani, A. Jalilian, A. Bahmanyar, D. Ernst, Comprehensive review on static and dynamic distribution network reconfiguration methodologies, *IEEE Access*, 12 (2024) 9510-9525.
- [5] T. Thavaratnam, B. Venkatesh, ML assisted T-model feeder reduction and convex fast reconfiguration method for active distribution networks with DERs, *IET Generation, Transmission & Distribution*, 20(1) (2026) e70226. doi: 10.1049/gtd2.70226.
- [6] H. Ahmadi, B. Vahidi, S.S. Aghaee, S.H. Hosseini, A. Mosallanejad, A new load-flow method in distribution networks based on an approximation voltage-dependent load model in extensive presence of distributed generation sources, *AUT Journal of Electrical Engineering*, 52(2) (2020) 133-146. doi: 10.22060/eej.2020.18258.5348.
- [7] E. Azad-Farsanni, M. Zare, Stochastic multi-objective distribution network reconfiguration considering wind turbines, *AUT Journal of Electrical Engineering*, 53(1) (2021) 111-124. doi: 10.22060/eej.2021.19203.5384.
- [8] S. Mohammadi, V.H. Bui, W. Su, B. Wang, Surrogate modeling for solving OPF: A review, *Sustainability*, 16 (2024) 9851.
- [9] M. Khodayar, J. Regan, Deep neural networks in power systems: A review, *Energies*, 16 (2023) 4773.
- [10] M. Shahraeini, M. Besharatloo, From DC to holistic centrality: AI-driven estimation for enhanced power network analysis, *AUT Journal of Electrical Engineering*, 58(1) (2026) 121-148. doi: 10.22060/eej.2025.24278.5667.
- [11] X. Pan, M. Chen, T. Zhao, S.H. Low, DeepOPF: A feasibility-optimized deep neural network approach for AC optimal power flow problems, *IEEE Systems Journal*, 17(1) (2023) 673-683.
- [12] L. Piloto, S. Liguori, S. Madjiheurem, M. Zgubic, S. Lovett, H. Tomlinson, S. Elster, C. Apps, S. Witherspoon, CANOS: A fast and scalable neural AC-OPF solver robust to N-1 perturbations, *arXiv preprint arXiv:2403.17660*, 2024.
- [13] A. Deihim, D. Apostolopoulou, E. Alonso, Initial estimate of AC optimal power flow with graph neural networks, *Electric Power Systems Research*, 234 (2024) 110782.
- [14] J. Authier, R. Haider, A. Annaswamy, F. Dörfler, Physics-informed graph neural network for dynamic reconfiguration of power systems, *Electric Power Systems Research*, 235 (2024) 110817.
- [15] P. Yeptho, A.E. Saldaña-González, M. Aragüés-Peñalba, S. Barja-Martínez, Evaluation of XGBoost and ANN as surrogates for power flow predictions with dynamic energy storage scenarios, *Energies*, 18 (2025) 4416.
- [16] G. Abbas, W. Zhi, A. Ali, A two-stage reactive power optimization method for distribution networks based on a hybrid model and data-driven approach, *IET Renewable Power Generation*, 18 (2024) 3967-3979.
- [17] V. Eeckhout, H. Fani, M.U. Hashmi, G. Deconinck, Improved physics-informed neural network based AC power flow for distribution networks, *arXiv preprint arXiv:2409.09466*, 2024.
- [18] J. Jalving, M. Eydenberg, L. Blakely, A. Castillo, Z. Kilwein, J.K. Skolfield, F. Boukouvala, C. Laird, Physics-informed machine learning with optimization-based guarantees: Applications to AC power flow, *International Journal of Electrical Power and Energy Systems*, 157 (2024) 109741.
- [19] B. Donon, R. Clément, B. Donnot, A. Marot, I. Guyon, M. Schoenauer, Neural networks for power flow: Graph neural solver, *Electric Power Systems Research*, 189 (2020) 106547.
- [20] A. Yaniv, P. Kumar, Y. Beck, Towards adoption of GNNs for power flow applications in distribution systems, *Electric Power Systems Research*, 216 (2023) 109005.
- [21] L. Thurner, A. Scheidler, F. Schäfer, J.H. Menke, J. Dollichon, F. Meier, S. Meinecke, M. Braun, Pandapower—an open-source Python tool for convenient modeling, analysis, and optimization of electric power systems, *IEEE Transactions on Power Systems*, 33(6) (2018) 6510-6521.
- [22] M.E. Baran, F.F. Wu, Network reconfiguration in distribution systems for loss reduction and load balancing, *IEEE Transactions on Power Delivery*, 4(2) (1989) 1401-1407.
- [23] M. Zhou, M. Chen, S.H. Low, DeepOPF-FT: One deep neural network for multiple AC-OPF problems with flexible topology, *IEEE Transactions on Power Systems*, 38(1) (2023) 964-967. doi: 10.1109/TPWRS.2022.3217407.
- [24] S. Liu, C. Wu, H. Zhu, Topology-aware graph neural networks for learning feasible and adaptive AC-OPF solutions, *IEEE Transactions on Power Systems*, 38(6) (2023) 5660-5670. doi: 10.1109/TPWRS.2022.3230555.
- [25] T. Falconer, L. Mones, Leveraging power grid topology in machine learning assisted optimal power flow, *IEEE Transactions on Power Systems*, 38(3) (2023) 2234-2246. doi: 10.1109/TPWRS.2022.3187218.
- [26] S. Jadhav, B. Sevak, S. Das, W. Su, V.H. Bui, Enhancing power flow estimation with topology-aware gated graph neural networks, *arXiv preprint arXiv:2507.02078*, 2025.
- [27] A. Yaniv, C. Goebel, Benchmarking graph neural networks for power flow prediction in distribution systems, in: *Proceedings of IEEE Kiel PowerTech*, 2025. doi: 10.1109/PowerTech59965.2025.11180442.

- [28] H. Li, L. Liu, S. Yu, Q. He, Q. Wu, J. Zhang, Q. Lu, Graph attention convolution network for power flow calculation considering grid uncertainty, *International Journal of Electrical Power and Energy Systems*, 165 (2025) 110513. doi: 10.1016/j.ijepes.2025.110513.
- [29] R.A. Jacob, S. Paul, S. Chowdhury, Y.R. Gel, J. Zhang, Real-time outage management in active distribution networks using reinforcement learning over graphs, *Nature Communications*, 15 (2024) 4766. doi: 10.1038/s41467-024-49207-y.
- [30] T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785-794. doi: 10.1145/2939672.2939785.
- [31] L. Breiman, Random forests, *Machine Learning*, 45(1) (2001) 5-32.
- [32] M.E. Baran, F.F. Wu, Optimal capacitor placement on radial distribution systems, *IEEE Transactions on Power Delivery*, 4(1) (1989) 725-734.
- [33] T. Ishizaki, A. Chakraborty, J.I. Imura, Graph-theoretic analysis of power systems, *Proceedings of the IEEE*, 106 (2018) 931-952.
- [34] B. Pang, E. Nijkamp, Y.N. Wu, Deep learning with TensorFlow: A review, *Journal of Educational and Behavioral Statistics*, 45 (2020) 227-248.
- [35] G. Chen, P. Chen, Y. Shi, C.-Y. Hsieh, B. Liao, S. Zhang, Rethinking the usage of batch normalization and dropout in the training of deep neural networks, *arXiv preprint arXiv:1905.05928*, 2019.
- [36] A. Badola, V.P. Nair, R.P. Lal, An analysis of regularization methods in deep neural networks, in: *Proceedings of IEEE INDICON*, New Delhi, India, 2020, pp. 1-6.