Physics-informed Multi-agent Learning for Corrective Optimal Power Flow

Yinyin Ge, Hongxing Ye, *Senior Member, IEEE,* Xingtong Peng, Yadong Zhai, Wangqing Mao

***Abstract*—Maintaining security is the priority for power system operation. The time window for the operator to take corrective action is quite limited post contingency. This work presents a novel physics-informed approach to the corrective optimal power flow (COPF) problem post contingency. COPF aims to determine a new optimal state post-contingency while meeting various physical laws and engineering limits. It is fundamentally important to find the corrective action when a contingency is detected. The traditional deep neural networks (DNNs)-based approach confronts challenges of explainability and solution accuracy. We introduce a physics-informed multiple DNNs to solve the problem. An affine matrix, containing physics information, is extracted from robust optimization for uncertainty management. Then it is fed to the multi-DNN model for network partitioning. We present a two-stage learning process with Lagrangian relaxation. Comprehensive case studies are conducted with IEEE 118-bus systems, IEEE 300-bus systems, and European 1354-bus systems. The proposed approach achieves an optimality gap of less than 0.1% while maintaining 99.9% constraint feasibility. Compared to the traditional optimization approach, the proposed method reduces the solution time by up to a factor of 414. The simulation results indicate that the proposed approach has improved accuracy and speed.**

***Index Terms*—Multi-agent Deep Learning, physics-informed Learning, Partition, Corrective Optimal Power Flow**

NOMENCLATURE

$\mathcal{N}, \mathcal{N}^U$	Set of all buses, set of buses with uncertainty
$\mathcal{L}$	Set of lines
$\mathcal{G}, \mathcal{G}_i^b$	Set of generators, set of generators at bus i
P_g^G	Active output of generator g
Q_g^G	Reactive output of generator g
P_i^d	Active load of bus i
Q_i^d	Reactive load of bus i
V_i	Voltage magnitude of bus i
θ_i	Voltage phase angle of bus i
ϵ	The net power injection uncertainty
$\boldsymbol{G}$	Affine Policy
K	Total number of partitions
$\mathcal{N}_k, \mathcal{G}_k, \mathcal{L}_k$	Bus, generator, and line sets in partition D_k
$\boldsymbol{W}_l^{(1)}$	Weight matrix for DNN at layer l in stage 1
$\boldsymbol{W}_l^{(2)}$	Weight matrix for DNN at layer l in stage 2
$\boldsymbol{y}_l^{(1)}, \boldsymbol{y}_l^{(2)}$	Output of DNN at layer l in stage 1 and 2
$\boldsymbol{W}_k$	Weight parameter for DNN in partition D_k
ξ	Binary line contingency vector
λ_c	Lagrange multipliers
α	Learning rate
ρ	Lagrange multiplier update factor
v_c	Violation degree of constraint
$\mathcal{S}$	Training data set
L_k^{total}	Training objective function for partition D_k

H. Ye is with College of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, China.

Y. Ge and X. Peng are with College of Cyber Science and Engineering at Xi'an Jiaotong University, Xi'an, China.

Y. Zhai is now with Microsoft Corporation, Suzhou, China. (Work was done when he studied at Xi'an Jiaotong University.)

W. Mao is with State Grid Jiangsu Electric Power Co., LTD, Suqian, China.

Corresponding author: Hongxing Ye (yehxing@xjtu.edu.cn). This work is supported by Science and Technology Program of State Grid No. 5700-202358590A-3-2-ZN.

I. INTRODUCTION

ACCORDING to the U.S. National Academy of Engineering, the electric power infrastructure is one of the largest and most critical infrastructures in the world. Maintaining security is a top priority in power system operations. Failing to address contingencies in a timely manner may trigger cascading outages, leading to unpredictable losses. Many countries are experiencing increasing pressure from power outage due to extreme weather [1]. For example, Hurricane Ida caused at least 1.2 million people to lose power in 2021 [2]. There were large-scale power outages in southern China in December 2020 and January 2021 due to the extreme cold weather [1]. In April 2025, Spain and Portugal experienced a historical blackout, starting with 2200 MW of generation trips [3].

With the increasing severity, it is fundamental to take appropriate and timely actions to address power system contingencies. Power system operators often consider potential contingencies or N-1 security constraints [4], [5]. N stands for the number of critical equipment, such as transmission lines. In operation, preventive and corrective measures are used. In a preventive model, contingencies are considered in advance, and no re-dispatch actions are taken post-contingency. It inevitably leads to conservativeness, and thus only considers limited credible contingencies in the power system [5]. In a corrective model, available resources are preserved, and they can be dispatched on demand post-contingency. In order to take corrective actions, the system operator first must guarantee there are enough flexible resources, and then the timeliness requirement must be met post contingency.

The COPF model schedules the dispatch for both the normal operating condition and corrective actions post contingency. Therefore, flexible resources are optimally preserved in the COPF. Due to its nonconvex nature, the COPF is computationally intractable. It is challenging to quickly find the COPF solution that satisfies the timeliness requirement.

This article has been accepted for publication in *IEEE Transactions on Power Systems*. Citation information: DOI 10.1109/TPWRS.2026.3672692

Although existing model-based approaches can deliver high accuracy, they struggle to satisfy real-time requirements, as the problem is NP-hard and computations are slow for large systems. While some warm-start methods expedite optimal power flow computation, many nonlinear constraints still require iterative calculation. Additionally, to meet real-time requirements, researchers have proposed DC optimal power flow and model linearization methods, which ensure some level of real-time performance but remain insufficient to ensure solution feasibility and constraint satisfaction rates [6] [7]. Therefore, there is a persistent need to continuously improve these methods and explore new algorithms for more accurate and stable solutions.

Emerging deep learning techniques provide another tool in solving problems with complex nonlinear models. These models can express intricate variable relationships through the training of neural networks [8] and achieve acceleration by several orders of magnitude while maintaining an acceptable level of accuracy [9], [10]. Employing DNN to solve optimal power flow not only reduces solution time and avoids iterative processes but also delivers acceptable accuracy [9]. However, relying solely on the deep learning, without considering fundamental physics, can result in poor interpretability. The performance of DNN-based approaches is challenged by larger system sizes. This can adversely affect solution quality, increase training time, and compromise solution feasibility.

Therefore, we propose a domain-tailored deep learning approach for solving optimal power flow (OPF) problems in power systems affected by contingency events. The paper discusses using physical model information to guide the DNN in more effectively approximating the OPF model. This proposal aims to improve the accuracy and efficiency of power grid management post contingencies.

II. Related Work

Researchers employ network partitioning to address the power system black-start restoration problem. For example, authors propose an efficient network partitioning approach based on mixed-integer programming for power grid restoration [11]. The original power network is partitioned into multiple subnetworks based on the recursive spectral dichotomy in graph theory [12]. The idea is to divide the power network into clusters of buses and transmission elements to enhance the computational efficiency of steady-state analysis [13]–[15]. As the problem size grows nonlinearly with network scale, network partitioning can alleviate both computational and memory requirements. For example, the authors in [14] provide a network partitioning method based on slow coherence, which means that some generators exhibit similar rotor angle swings in inter-area mode oscillation.

Machine learning has rapidly developed in recent years and has performed well in learning tasks in many fields, including the power system [16]. [17] applies convolutional neural networks to learn the mapping between load/network topology and voltage, thereby partially addressing the OPF problem. In [18], the deep neural network is used to solve distributed DC optimal power problems for islanded microgrid systems. In [19], the authors review various machine learning methods, including the Support Vector Machine, Neural Network, and Decision Tree, to solve preventive control problems. In our earlier study [20], multiple DNNs are employed. However, it may encounter islanding issues during partitioning and cause infeasibility due to power mismatches among adjacent buses. In [21], graph neural network is also employed to predict heavily loaded lines, accelerating OPF. The graph neural network is studied for the topology-changed OPF problem in [22]. Another method, introduced in [23], utilizes a penalty function and a zero-order optimization technique, which is to estimate the gradient based on the two-point difference rather than the derivatives. Another approach, described in [24], utilizes a neural network embedded with KKT conditions to estimate optimal power flow while ensuring strict performance benchmarks. [25] proposes a voltage constraint method based on neural networks, which can efficiently predict the solution to AC-OPF problems. [26] presents a deep learning method for solving multiple AC-OPF problems with flexible topology and line admittance, enhancing the DNN's prediction capability across various grid topologies. Although deep learning models are known for their accuracy, their training can be problematic due to inherent nonconvexity and the multitude of local optima they exhibit [27].

The paper presents a new method that uses deep learning to address corrective AC-OPF problems. The proposed approach utilizes appropriate model information to assist the deep neural network (DNN) in approximating the solution. The paper suggests deploying multiple neural networks, each trained for a grid subnetwork, to tackle this task. By leveraging domain-tailored clustering, the approach can achieve high precision and faster speed [11], [14]. The main innovations and contributions of this work are as follows.

- A multi-agent-based training framework is proposed for the corrective OPF problem. Each agent with DNN represents one subnetwork, and these DNNs can be trained in parallel. The proposed framework reduces the complexity of the training model and improves accuracy. It provides a closed-form expression for re-dispatch post-contingency in large-scale power systems. The framework decouples voltage and power training into two stages. In the first stage, voltage magnitude and phase angle are trained. The voltage information is then fed to the training model in the second stage.
- A novel partitioning method is proposed for DNN to solve the corrective OPF problem. The affine matrix in adaptive robust optimization is employed for K-means in power network partitioning. The affine matrix, originally used for uncertainty management, respects various physical and engineering constraints. Based on the partitioning result, each DNN is trained per grid subnetwork. This approach enables modular training and reduces the computational burden compared to monolithic DNNs.
- This work introduces a strategy to address the feasibility challenge within the multi-DNN framework. As adjacent buses in different partitions are trained by separate agents, their independently learned outputs may violate

KCL/KVL laws. Therefore, virtual buses for adjacent buses are introduced into their neighboring partitions. Incorporating them into the loss function ensures voltage and flow consistency across subnetwork boundaries. Additionally, Lagrangian-like relaxation is employed to address the lower and upper limit constraints for the voltage magnitude, real and reactive power.

The rest of this paper is organized as follows. Section III presents the proposed method. Section IV describes the extensive experiments and analyzes. Finally, conclusions are drawn in Section V.

III. Affine Matrix Based Partitioning

The OPF model is presented to ensure a self-contained presentation. The OPF typically computes the least-cost generation dispatch that satisfies both physical and operational constraints. In generalized form, an OPF model can be expressed as

$$\min \quad \sum_{g\in\mathcal{G}} C_g(P_g^G), \tag{1}$$

$$\text{s.t.} \quad P_{ij}^L = Y_{ij}^R V_i^2 - V_i V_j (Y_{ij}^I \sin(\theta_i-\theta_j) + Y_{ij}^R \cos(\theta_i-\theta_j)), \quad \forall (i,j)\in\mathcal{L}, \tag{2}$$

$$Q_{ij}^L = -Y_{ij}^I V_i^2 - V_i V_j (Y_{ij}^R \sin(\theta_i-\theta_j) - Y_{ij}^I \cos(\theta_i-\theta_j)), \quad \forall (i,j)\in\mathcal{L}, \tag{3}$$

$$\sum_{g\in\mathcal{G}_i^b} P_g^G - P_i^d - \sum_{(i,j)\in\mathcal{L}} P_{ij}^L = 0\ , \quad \forall i\in\mathcal{N}, \tag{4}$$

$$\sum_{g\in\mathcal{G}_i} Q_g^G - Q_i^d - \sum_{(i,j)\in\mathcal{L}} Q_{ij}^L = 0, \quad \forall i\in\mathcal{N}, \tag{5}$$

$$V_i^{\min} \le V_i \le V_i^{\max}, \quad \forall i\in\mathcal{N}, \tag{6}$$

$$P_g^{G\,\min} \le P_g^G \le P_g^{G\,\max}, \quad \forall g\in\mathcal{G}, \tag{7}$$

$$Q_g^{G\,\min} \le Q_g^G \le Q_g^{G\,\max}, \quad \forall g\in\mathcal{G}, \tag{8}$$

$$(P_{ij}^L)^2 + (Q_{ij}^L)^2 \le S_{ij}^{\max}, \quad \forall (i,j)\in\mathcal{L}, \tag{9}$$

$$-\theta_{ij}^{\Delta} \le \theta_i - \theta_j \le \theta_{ij}^{\Delta}, \quad \forall (i,j)\in\mathcal{L}, \tag{10}$$

where $\mathcal{N}, \mathcal{L}$, and $\mathcal{G}$ denote the sets of bus, branch, and generator, respectively. $\mathcal{G}_i^b$ denotes the generator set at bus i. $C_g(\cdot)$ denotes the cost function for generator g. Y_{ij}^R and Y_{ij}^I represent the real and imaginary parts of the line admittance, respectively. V_i and θ_i are the magnitude and phase angle of bus voltage, respectively. P_i^d, Q_i^d, P_g^G, and Q_g^G represent aggregated active and reactive loads, active and reactive outputs of the generators, respectively. The objective function (1) aims to minimize generation cost, typically formulated as a linear or quadratic function. Note that the objective function can be changed to other functions, such as minimizing energy loss. Constraints (2) and (3) denote the power flow equations following Kirchhoff's Voltage Law (KVL) law, and Equations (4) and (5) represent the nodal balance constraints. Constraint (6) represents the voltage magnitude limit. Equations (7) and (8) are the upper and lower limits for real and reactive power, respectively. Equation (9) denotes the transmission thermal constraint, ensuring that the power flow remains within the capacity limits of the transmission lines. Equation (10) denotes the limit of the phase angle difference between connected buses.

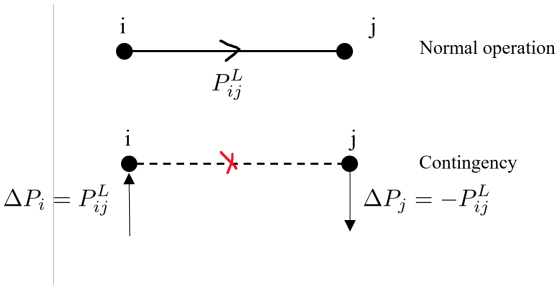


Fig. 1. Illustrative example of line contingency with flow P_{ij}^L. Line outage is viewed as admittance change and net-power injection uncertainty, which is used to generate redispatch matrix $\boldsymbol{G}$ for partitioning purpose.

To address the Corrective OPF problem, we first partition the power system network and then represent grid subnetworks with multiple deep neural networks. With proper partitioning, the benefits are twofold. First, the training process can be executed in parallel. Second, each DNN models a significantly smaller subnetwork. To achieve this goal, we classify buses based on power flow redistribution post contingency. In the proposed partitioning method, line outage is viewed as uncertainty with admittance change according to Line Outage Distribution Factor (LODF), as shown in Fig. 1. A surrogate affine approximation (SAA) in [28] is employed to capture the relationship between net power injections and voltage data. The surrogate affine policy is originally designed to address the renewable integration problem, and is essentially an optimized solution for adjusting controllable resources against uncertainty. It is noted that the uncertainty-based outage model is solely used for partitioning purpose.

An abstract formulation of the affinely adjustable robust model is provided below.

$$\min_{\boldsymbol{z},\boldsymbol{G}} \quad f(\boldsymbol{z},\boldsymbol{G}), \tag{11}$$

$$\text{s.t.} \quad g(\boldsymbol{z}) \le 0, \tag{12}$$

$$h(\boldsymbol{z^c}) \le 0, \tag{13}$$

$$\boldsymbol{z^c} = \boldsymbol{z} + \boldsymbol{G\epsilon}, \forall \boldsymbol{\epsilon}\in\mathcal{U}, \tag{14}$$

where $\boldsymbol{z}$ presents decision variables in the base-scenario case. $\boldsymbol{G}$ represents the affine policy, i.e., redispatch policy, the recourse action when the uncertainty is materialized. Variable $\boldsymbol{z}$ and $\boldsymbol{G}$ are decision variables. $\boldsymbol{\epsilon}$ represents the uncertainty in net power injection, and $\mathcal{U}$ is the uncertainty set. The function f can be the operation cost or the expected cost considering uncertainty. Equation (12) represents constraints in normal condition, including (2-10). The constraints for re-dispatch are represented by Equation (13), with $\boldsymbol{z}^c$ representing the voltage and power variables after the uncertainty is materialized. The system operator aims to achieve a new stable state that meets physical and operational constraints by re-dispatching controllable resources. This is achieved by the affine function as expressed in Equation (14). Matrix $\boldsymbol{G}$ maps the net power injection uncertainty to other variables. By solving the problem described in (11)-(14), we can obtain the matrix $\boldsymbol{G}$. Readers

are referred to [28] for more details on the affinely adjustable model.

The linearization of the power flow equation employs a method delineated in [29], followed by the application of the SAA approach from [28] to determine the optimal solution. The matrix $\boldsymbol{G}$ is written accordingly as

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{g}_1^\top \\ \boldsymbol{g}_2^\top \\ \cdots \\ \boldsymbol{g}_{|\mathcal{N}|}^\top \end{bmatrix} \in \mathcal{R}^{2|\mathcal{N}|\times 2|\mathcal{N}^U|}, \quad \boldsymbol{g}_i \in \mathcal{R}^{2|\mathcal{N}^U|\times 2}, \tag{15}$$

where $|\mathcal{N}|$ represents the bus number, and $|\mathcal{N}^U|$ denotes the number of buses with uncertain power injection, and $\boldsymbol{g}_i$ represents the coefficients for uncertainty at bus i, with a dimension of $2|\mathcal{N}^U| \times 2$. Given matrix $\boldsymbol{G}$, K-means algorithm is employed to partition the network into K partitions. The partitioning process is as follows:

- Step 1, Optimization. Formulate problem (11)–(14), and solve it to attain $\boldsymbol{G}$.
- Step 2, Clustering. Treat $\boldsymbol{g}_i$ in $\boldsymbol{G}$ as a data point, and implement K-means clustering classification with $\arg\min_k d(\boldsymbol{g}_i, u_k)$, where $\boldsymbol{u}_k$ is the cluster center.
- Step 3, Fine tuning. Employ Depth-First Search algorithm to fine-tune bus classification.

Fundamentally, line contingency can be viewed as admittance change and power injection uncertainty, which motivates clustering $\boldsymbol{G}$ for network partitioning. A contingency often impacts only nearby buses. Therefore, clustering the affine matrix $\boldsymbol{G}$, i.e., the coefficients for re-dispatch after uncertainty, essentially captures the intrinsic relationships within the system. The resulting partitions ensure strong correlations within each partition and weak interconnections between different partitions. The partition number K can be selected based on the elbow method. In addition, the fine-tuning step is used to eliminate the islanding buses as a subroutine. It is noted that $\boldsymbol{G}$ is a solution derived from robust optimization. Thus, the corresponding re-dispatch respects all constraints (14).

IV. Multi-DNN Model

This work presents a learning framework grounded in the physical model. It can reduce training time and enhance deep learning accuracy—an essential benefit when scaling to large power systems. The general framework is illustrated in Fig. 2. Two phases are involved. In the first phase, we formulate a robust optimization model considering uncertainty, and then get the affine matrix G. By clustering vectors in G, we partition the original grid network into subnetworks, $D \to D_1, D_2, \cdots, D_k, \cdots, D_K$. The original bus set can be written as

$$\mathcal{N} = \mathcal{N}_1 \cup \mathcal{N}_2 \cdots \cup \mathcal{N}_k \cdots \cup \mathcal{N}_K, \tag{16}$$

where $\mathcal{N}_k$ denotes the bus set in partition D_k. In the second phase, multiple DNNs are trained to predict the solution to the COPF problem.

Deep neural networks refer to networks that have many layers. Deep neural networks can approximate complex variable mappings by training models from a labeled dataset.

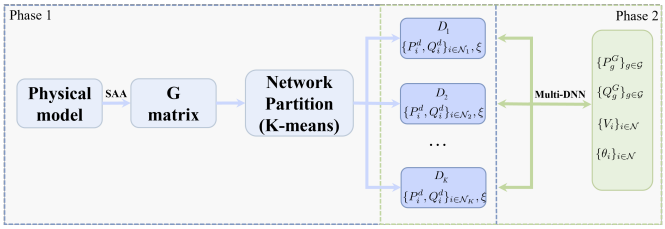


Fig. 2. The overall framework of the proposed method. In Phase 1, the power grid is partitioned by clustering redisppatch policy matrix from robust model. In Phase 2, the COPF solution is the learning output of multi-DNN with physical information embedded.

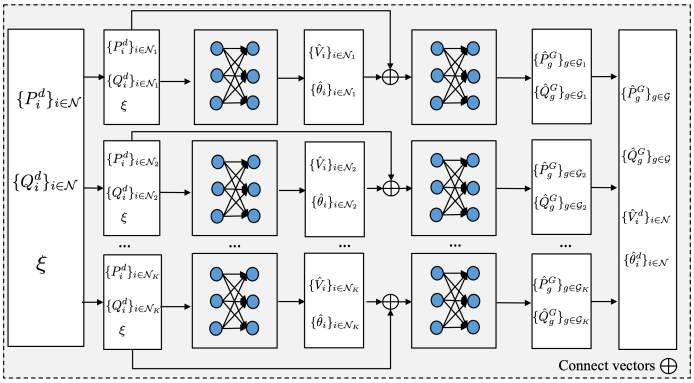


Fig. 3. Training flow chart of multiple DNNs. Each DNN is trained per partition. In the first stage, $\hat{V}_i$ and $\hat{\theta}_i$ are the output. $\hat{P}_g^G$ and $\hat{Q}_g^G$ are the outputs in the second stage.

For brevity, we temporarily omit the partition index k of the model. For each partition D_k, the model in the first stage is represented as

$$\boldsymbol{y}_0^{(1)} = \{P_i^d, Q_i^d, \xi\}_{i\in\mathcal{N}_k} \in \mathcal{R}^{2|\mathcal{N}_k|+1} \tag{17a}$$

$$\boldsymbol{y}_l^{(1)} = \phi_1\left(\boldsymbol{W}_l^{(1)}\boldsymbol{y}_{l-1}^{(1)} + \boldsymbol{b}_l^{(1)}\right),\ l = 1, ..., L_1. \tag{17b}$$

ξ stands for line contingency vector with 0-1 encoding. In the second stage, the model is denoted as

$$\boldsymbol{y}_0^{(2)} = \{\hat{V}_i, \hat{\theta}_i, P_i^d, Q_i^d, \xi\}_{i\in\mathcal{N}_k} \in \mathcal{R}^{4|\mathcal{N}_k|+1} \tag{18a}$$

$$\boldsymbol{y}_l^{(2)} = \phi_2\left(\boldsymbol{W}_l^{(2)}\boldsymbol{y}_{l-1}^{(2)} + \boldsymbol{b}_l^{(2)}\right),\ l = 1, ..., L_2. \tag{18b}$$

Superscripts $^{(1)}$ and $^{(2)}$ denote the network parameters in the first and second stages, respectively. Let $\hat{}$ denote the DNN learning output. $\boldsymbol{y}_{L_1}^{(1)} = \{\hat{V}_i, \hat{\theta}_i\}_{i\in\mathcal{N}_k} \in \mathcal{R}^{2|\mathcal{N}_k|}$ and $\boldsymbol{y}_{L_2}^{(2)} = \{\hat{P}_g^G, \hat{Q}_g^G\}_{g\in\mathcal{G}_k} \in \mathcal{R}^{2|\mathcal{G}_k|}$ are outputs of networks at stage one and two, respectively. $\mathcal{G}_k$ stands for the generator set in partition D_k. We have

$$|\mathcal{G}| = \sum_k |\mathcal{G}_k|.$$

$\boldsymbol{W}_l^{(1)}$ and $\boldsymbol{W}_l^{(2)}$ are weight matrices of the networks. $\phi_1(\cdot)$ and $\phi_2(\cdot)$ are nonlinear functions, such as sigmoid and relu activation functions.

The proposed multi-DNN framework is presented in Fig. 3. Each part of the network consists of a multi-layer fully connected neural network with nonlinear functions, such as sigmoid and relu activation functions [30]. There are two

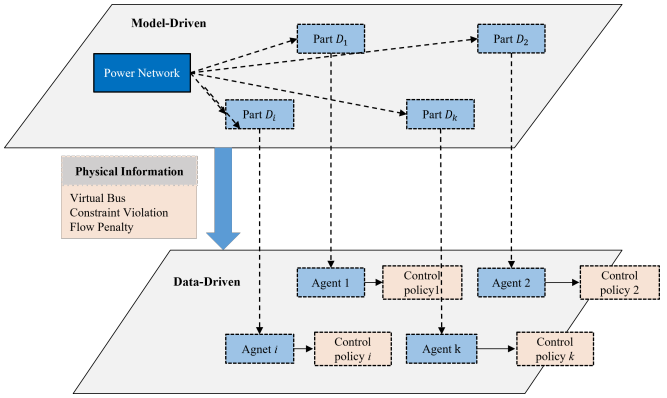


Fig. 4. Model- and data-driven framework. The grid is partitioned based on model information, and fed to DNN, which is trained with physical information, such as virtual bus, constraint violation, and flow penalty.

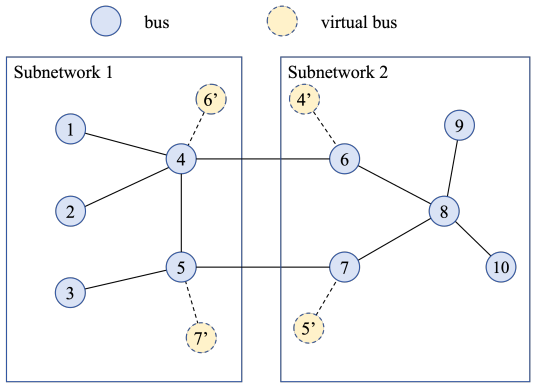


Fig. 5. An illustrative example of virtual buses, which are introduced to eliminate voltage/power mismatch of adjacent buses in different partitions.

stages in the training process. In the first stage, the input comprises the load vector and contingency information $\{P_i^d, Q_i^d, \xi\}_{i\in\mathcal{N}_k}$. In the second stage, the input includes the load vector, contingency information, voltage magnitude, and phase angle. The learning output of the second stage includes the generators' active and reactive outputs. The method for combining model- and data-driven strategies is shown in Fig.4. The control policy refers to $(P_g^G, Q_g^G, V_i, \theta_i)$.

A. Modeling the constraints

To capture the OPF constraints, we employ a Lagrangian-like relaxation approach. Following [31], [32], the Lagrangian relaxation function of an optimization problem

$$\begin{aligned} \min \quad & f_0(\boldsymbol{z}) \\ \text{s.t.} \quad & \boldsymbol{f}_1(\boldsymbol{z}) = 0 \\ & \boldsymbol{f}_2(\boldsymbol{z}) \le 0 \end{aligned}$$

is $f_0(\boldsymbol{z}) + \boldsymbol{\lambda}_1^\top \boldsymbol{f}_1(\boldsymbol{z}) + \boldsymbol{\lambda}_2^\top \boldsymbol{f}_2(\boldsymbol{z})$, where $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2 \ge 0$ are Lagrangian multiplier vectors. Let c denote the constraint index. We define the violation degree as follows.

$$v_c = \begin{cases} f_c(\boldsymbol{z}) & \text{if } f_c \text{ is equality constraint} \\ \max(0, f_c(\boldsymbol{z})) & \text{if } f_c \text{ is inequality constraint} \end{cases} \tag{19}$$

The penalty function is then expressed as $\lambda_c v_c$, where λ_c is the Lagrangian multiplier. In this work, λ_c is updated together with the deep network parameters, which distinguishes it from the conventional Lagrangian relaxation approach. Further information will be provided in a subsequent section.

The relaxation technique can handle the majority constraints. However, residual errors may occur in very few cases. Feasibility tolerance can be imposed to mitigate constraint violations.

B. Introducing virtual buses

As an independent DNN is trained per partition, serious voltage and power mismatch issues may arise for adjacent buses in different partitions, also referred to as subnetworks. To address this, we propose fine-tuning the loss function.

Let $D_{k'}$ be an adjacent partition of partition D_k. If bus i in partition $D_{k'}$ is connected to any bus in partition D_k, then we introduce a virtual bus i' in partition D_k. The idea is illustrated in Fig. 5. An additional term is then introduced regarding the difference between a virtual bus i' and a real bus i. The loss term for the virtual bus is expressed as follows:

$$L_{i'} = C^{V'} \left(V_{i'} - V_i\right)^2 + C^{\theta'} \left(\theta_{i'} - \theta_i\right)^2,$$

where $V_{i'}$ and $\theta_{i'}$ represent the predicted voltage and phase angle values of the virtual bus, V_i and θ_i represent the corresponding values of the real bus in the training set. $C^{V'}$ and $C^{\theta'}$ are coefficients.

Let $\mathcal{V}_k$ denote the set of all virtual buses in partition D_k. The virtual bus loss for each partition D_k is given by:

$$L_k^{virtual} = \sum_{i'\in\mathcal{V}_k} L_{i'}, \tag{20}$$

We integrate virtual bus losses into the loss function.

C. Multi-DNNs training steps

The load and system contingencies are utilized to predicate state and control variables $(P_g^G, Q_g^G, V_i, \theta_i)$ via a multi-DNN framework. This approach is conceptualized as a mapping function $F : \mathcal{R}^{2|\mathcal{N}|+1} \to \mathcal{R}^{2|\mathcal{N}|+2|\mathcal{G}|}$. Let the subscript k denote the parameters and variables in partition D_k. We add subscript s to represent data for the s^{th} training instance. For partition k, the dataset is written as $\boldsymbol{\mathcal{D}}_k \triangleq \{\boldsymbol{x}_k, \boldsymbol{y}_k\}$, with

$$\begin{aligned} \boldsymbol{x}_k &\triangleq \{P_{i,s}^d, Q_{i,s}^d, \xi_s\}_{i\in\mathcal{N}_k, s\in\mathcal{S}} \\ \boldsymbol{y}_k &\triangleq \{P_{g,s}^G, Q_{g,s}^G, V_{i,s}, \theta_{i,s}\}_{g\in\mathcal{G}_k, i\in\mathcal{N}_k, s\in\mathcal{S}}, \end{aligned}$$

where $\mathcal{S}$ denote the set of training dataset. The DNN is trained per partition. In partition D_k, the objective function of the training process is outlined below.

$$L_k^{total} \triangleq L_k^y + L_k^{vt} + L_k^{flow} + \sum_{c\in\mathcal{C}_k} \sum_{s\in\mathcal{S}} \lambda_{k,c} v_{k,c}(\boldsymbol{x}_k, \hat{\boldsymbol{y}}_k), \tag{21}$$

where

$$L_k^y = C^y \|\boldsymbol{y}_k - \hat{\boldsymbol{y}}_k\|^2 \tag{22}$$

$$L_k^{vt} = C^v \sum_{s\in\mathcal{S}} L_{k,s}^{virtual} \tag{23}$$

$$L_k^{flow} = \sum_{s\in\mathcal{S}} \sum_{(ij)\in\mathcal{L}_k} \left(C^p (P_{ij,s}^L - \hat{P}_{ij,s}^L)^2 + C^q (Q_{ij,s}^L - \hat{Q}_{ij,s}^L)^2 \right) \tag{24}$$

The last term in (21) is the constraint violation penalty following (19). $\mathcal{C}_k$ is the set of constraint indices. $\lambda_{k,c}$ is updated using the subgradient method. C^y, C^v, C^p, and C^q are loss coefficients. Equation (24) represents the penalty for power flow error. As a side note, power flow can be directly calculated with learned voltage magnitude and angle according to the KVL/KCL equation. They are introduced to mitigate the overfitting issues by embedding physical model in the DNN. They can be calculated once the voltages are known and act as side information to enhance learning efficiency [33].

Let $(\boldsymbol{W}_k, \boldsymbol{b}_k)$ denote the two-stage DNN parameters for all layers in partition D_k. Given $\lambda_{k,c}$, the training aims to minimize the loss by optimizing parameters $\boldsymbol{W}_k$ and $\boldsymbol{b}_k$

$$\min_{\boldsymbol{W}_k,\, \boldsymbol{b}_k} \quad \frac{1}{|\mathcal{S}|} L_k^{total}. \tag{25}$$

The overall training scheme is presented in Algorithm 1.

Algorithm 1 Training of Multi-DNN.

Input: $\boldsymbol{\mathcal{D}}_1, \cdots, \boldsymbol{\mathcal{D}}_K$: Training data.

1: α, ρ, $batchsize$: Parameters.
2: $\lambda_c = 0$: Initialization.
3: **for** epoch $<$ epochMax **do**
4: **for** $k = 1, 2, \cdots K$ **do**
5: Compute loss w.r.t Multi-DNN output using (21)
6: Update $\boldsymbol{W}_k$ and $\boldsymbol{b}_k$ by backpropagation:
7: $\boldsymbol{W}_k \leftarrow \boldsymbol{W}_k - \alpha \cdot \nabla_W L_k^{total}$
8: $\boldsymbol{b}_k \leftarrow \boldsymbol{b}_k - \alpha \cdot \nabla_b L_k^{total}$
9: **for** $c \in \mathcal{C}_k$ **do**
10: Update $\lambda_{k,c}$ by subgradient method:
11: $\lambda_{k,c} \leftarrow \lambda_{k,c} + \rho \cdot v_{k,c}(\boldsymbol{x_k}, \hat{\boldsymbol{y}}_{\boldsymbol{k}})$
12: **end for**
13: **end for**
14: epoch $\leftarrow$ epoch + 1
15: **end for**

The training data are partitioned according to subnetworks. Partitions share the same contingency input for each training instance. $\boldsymbol{x}_k$ and $\hat{\boldsymbol{y}}_k$ denote the input and learned output for partition D_k. The training consists of the following steps.

- Step 1: Formulate an affinely adjustable robust OPF according to (11)–(14). Solve the optimization problem to determine the optimal matrix $\boldsymbol{G}$.
- Step 2: Apply clustering algorithm on $\boldsymbol{G}$ for network partitioning.
- Step 3: Generate the dataset for each partition.
- Step 4: Train the DNN for each partition D_k, starting with epoch = 1, and calculate the DNN output $\hat{\boldsymbol{y}}_k$ using equations (17b) and (18b).
- Step 5: Calculate the loss function using equation (21).
- Step 6: Update the two-stage DNN parameters using backpropagation in Line 7-9, and adjust the Lagrangian multiplier with the subgradient method in Line 11.
- Step 7: If the maximum number (epochMax) of training epochs is not reached, then go to step 4 to continue training epoch = epoch + 1. If so, save the training model.

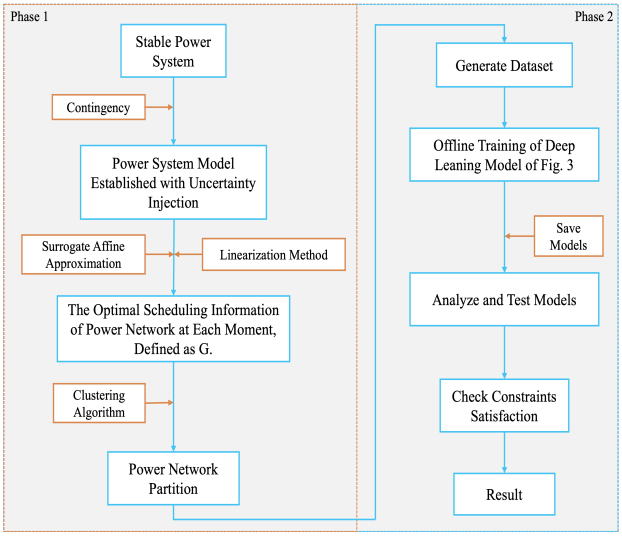


Fig. 6. The flow chart of the proposed method.

- Step 8: Analyze the model and record the results using the test dataset.

The overall flowchart of the previous analysis of the method proposed in this paper is depicted in Fig. 6. Check Constraints Satisfaction is to verify any violations using the learning output.

D. Datasets description and preprocessing

The datasets for training and testing this method are generated as follows.

Initially, the upper and lower bounds for each load bus are set. Denote the nominal load as $\boldsymbol{P_0^d}$ and $\boldsymbol{Q_0^d}$. The lower bound of the load is set to $(0.7\boldsymbol{P_0^d}, 0.7\boldsymbol{Q_0^d})$, and the upper bound is set to $(1.5\boldsymbol{P_0^d}, 1.5\boldsymbol{Q_0^d})$. We consider the N-1 security rule. The load vector is generated randomly within the lower and upper bounds under line contingency ξ, in which one line is out of service. The corrective optimal power flow is then solved using the MATPOWER Interior Point Solver (MIPS), which is capable of producing close-to-optimal solutions for AC-OPF. If the obtained OPF solution is feasible, it is added to the dataset; otherwise, the current solution is ignored.

We use 80% of the generated datasets for training and 20% for testing to evaluate model performance. To enhance training efficiency, the output data should be preprocessed before the training process to mitigate adverse effects from numerical issues and outliers. The zero-mean normalization method is employed for preprocessing.

V. Case Study

The simulations are conducted with the IEEE 118-bus, IEEE 300-bus, and European 1354-bus systems. Experiments are executed on a platform equipped with an Intel i5-12400 CPU, 64GB RAM, Matlab 2021a, MATPOWER 7.1, Python 3.8, Torch 1.12.1, and a RTX2080Ti GPU. Details of the cases are presented in Table I. To demonstrate the effectiveness of the proposed approach, we conducted a series of comparative experiments. We compared the following methods:

Table I. The introduction of case.

Case	Bus	Branch	Gen	P_d(MW)	Q_d(MVA)
IEEE 118-bus sys.	118	186	54	4242	1438
IEEE 300-bus sys.	300	411	69	23526	7788
European 1354-bus sys.	1354	1991	260	73060	13401

M0: Single DNN.

M1: Multi-DNN with sensitivity matrix partition method [34].

M2: Multi-DNN with the partition method proposed in [13].

M3: Multi-DNN with the partition method proposed in [11].

M4: Multi-DNN with our proposed partition method.

M5: Multi-DNN with our proposed partition method and multi-stage learning framework.

M6: Multi-DNN employing our proposed partition method and multi-stage learning framework with a modified loss function.

M7: Multi-DNN using our proposed partition method and multi-stage learning framework with a modified loss function and partition processing method utilizing virtual buses.

M8: DeepOPF-FT: An embedded training method to solve OPF with flexible topology and line admittances by designing a DNN [26].

M9: DeepOPF-V: A DNN-based voltage-constrained method to solve AC-OPF problems with high computational efficiency [25].

Table II lists these comparison methods along with their corresponding intentions. M8 and M9 are trained on the same dataset as the other methods. The DNN network layers and training dataset are shown in Table III. In multi-DNN cases, they are structures for each DNN.

Active Power Mismatch (APM), ΔP, ΔQ, ΔV, and $\Delta\theta$ are introduced to assess the quality of the learning output as follows.

$$APM = \sqrt{\sum_{i\in\mathcal{N}} \Delta\tilde{P}_i^2}$$

$$\Delta\tilde{P}_i^2 = \sum_{g\in\mathcal{G}_i} \hat{P}_g^G - P_i^d - \hat{V}_i \sum_{(i,j)\in\mathcal{L}} \hat{V}_j (Y_{ij}^R \cos\hat{\theta}_{ij} + Y_{ij}^I \sin\hat{\theta}_{ij})$$

$$\Delta P = \frac{1}{N}\sqrt{\sum_{g\in\mathcal{G}} (P_g^G - \hat{P}_g^G)^2}, \quad \Delta Q = \frac{1}{N}\sqrt{\sum_{g\in\mathcal{G}} (Q_g^G - \hat{Q}_g^G)^2}$$

$$\Delta V = \frac{1}{N}\sqrt{\sum_{i\in\mathcal{N}} (V_i - \hat{V}_i)^2}, \qquad \Delta\theta = \frac{1}{N}\sqrt{\sum_{i\in\mathcal{N}} (\theta_i - \hat{\theta}_i)^2}$$

The constraint satisfaction ratios of voltage magnitude, active generation, and reactive generation are denoted by η_V, η_{P_g} and η_{Q_g}. Let η_{opt} denote the average relative difference, i.e., optimality gap, between the optimal target values achieved by MIPS and the multiple DNNs.

A. Validations of the proposed partition method and Multi-DNN learning

1) Validations of the Multi-DNN: The experimental results suggest that deep learning parameters play a crucial role in determining the effectiveness of training outcomes. Based on our experimental findings, we recommend the selection of appropriate values for the deep learning parameters during the training process, as presented in Table IV. It is critical to carefully consider the impact of these parameters, such as batch size and learning rate, on the performance of the model. With same DNN structure, the training time for them are close. For example, the training time for the European 1354-bus systems is about 4 hours and 11 minutes.

Table II. Comparison method and the intentions.

Methods	Corresponding intentions
M0 and M1	Verify the effectiveness of the partition about the power system.
M1, M2, M3 and M4	Verify the effectiveness of the proposed partition method.
M4 and M5	Verify the effectiveness of the proposed multi-stage learning structure.
M4, M5 and M6	Verify the effectiveness of the proposed multi-stage learning structure and modified loss function.
M6, M7	Verify the effectiveness of partition processing method.
M7, M8 and M9	Verify the superiority of the proposed method.

Table III. Network Parameters for different cases.

Case	Structure of DNN	Training dataset
IEEE 118-bus sys.	[237 500 500 236 200 344]	20000
IEEE 300-bus sys.	[601 500 500 600 400 738]	80000
European 1354-bus sys.	[2709 1000 1354 1000 3120]	200000

Table V presents a performance comparison between M0 and M1. The table shows the deviation of predicted values from the actual data. The results indicate that our Multi-DNN learning framework significantly enhances performance compared to a single deep neural network. In particular, our framework demonstrates superior performance in large power systems.

For the IEEE 118-bus system, the M1 method resulted in a significantly smaller power mismatch than the M0 method, and also achieved better performance in calculating active power output, reactive power output, bus voltage magnitude, and phase angle. Similarly, for the 300 and 1354 systems, M1 exhibited a smaller power mismatch compared to M0, highlighting the advantages of partitioning large-scale power systems and using Multi-DNN for learning and calculation. Furthermore, the active output deviations of generators for M1 were less than 0.2 MW, 1.1 MW, and 1.4 MW for the case118, case300, and case1354 systems, respectively. Additionally, while the M0 method resulted in a voltage amplitude deviation of less than 0.4 KV, the M1 method achieved a deviation of less than 0.06 KV for the IEEE 118-bus system.

Table IV. Hyper Parameters for different cases.

Case	learning rate	batch size	ρ	λ_0
IEEE 118-bus sys.	5e-4	200	1e-3	0
IEEE 300-bus sys.	1e-4	200	1e-3	0
European 1354-bus sys.	1e-4	400	1e-3	0

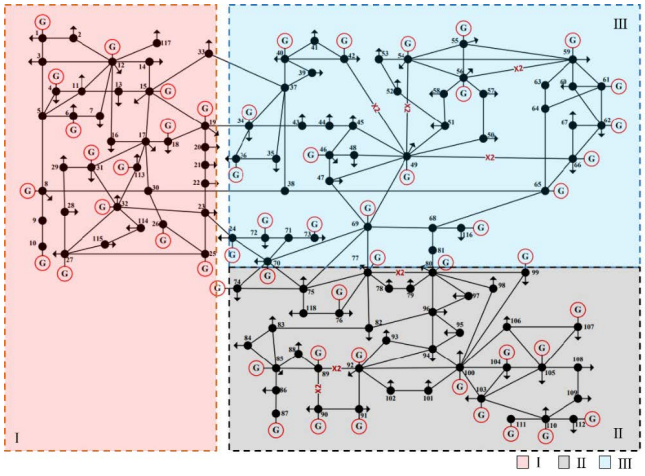


Fig. 7. The partition results of IEEE 118-bus system.

Table V. Comparison of prediction errors for M0 and M1 (MW, MVA,KV, deg).

Case	Methods	APM	ΔP	ΔQ	ΔV	$\Delta\theta$
IEEE 118-bus sys.	M0	1.2232	1.7806	2.2009	0.3884	1.2508
	M1	**0.4139**	**0.1314**	**0.1951**	**0.0571**	**0.0944**
IEEE 300-bus sys.	M0	2.5255	3.4599	8.9291	3.182	3.0208
	M1	**0.3517**	**1.0461**	**1.3131**	**0.4678**	**0.7500**
European 1354-bus sys.	M0	2.1099	3.5219	3.7749	1.3833	6.5224
	M1	**0.1554**	**1.3546**	**1.4536**	**0.1124**	**3.3850**

2) Validations of our proposed partitioning method: Experiments comparing the performance of the M1, M2, M3, and M4 methods are presented in Table VI. The partition result of M4 is shown in Fig. 7. Bold font denotes the best outcome, while underlined text represents the second best.

These results clearly demonstrate the superiority of our proposed power system partitioning approach over other algorithms. For the IEEE 118-bus, 300-bus, and 1354-bus systems, our method outperforms all others. For the IEEE 118-bus system, our approach achieves the best results in power mismatch, generator active output, and bus voltage phase angle, and is ranked second in generator reactive output and bus voltage amplitude. Overall, our method's performance outshines that of other partitioning algorithms.

For example, in the IEEE 300-bus system, the predication errors for generator active power output and voltage magnitude are less than 0.8 MW and 0.4 kV, respectively. Although method M2 achieves the best performance for $\Delta\theta$, our proposed method (M4) outperforms all other algorithms in terms of overall metrics. In particular, the prediction errors for generator active output, reactive output, and nodal voltage magnitude are less than 0.8 MW, 0.9 MW, and 0.2 kV, respectively, with our partitioning method. In the European 1354-bus system, the proposed partitioning method performs the best in four out five metrics.

3) Validations of modified loss function and improved multi-stage learning framework: To test the performance of the multi-stage learning process, we compare the computational results of M4, M5, and M6. Table VII shows the performance comparison of M4, M5, and M6 methods. The results for

Table VI. Comparison of prediction errors attained from different partitioning methods (MW, MVA,KV, deg).

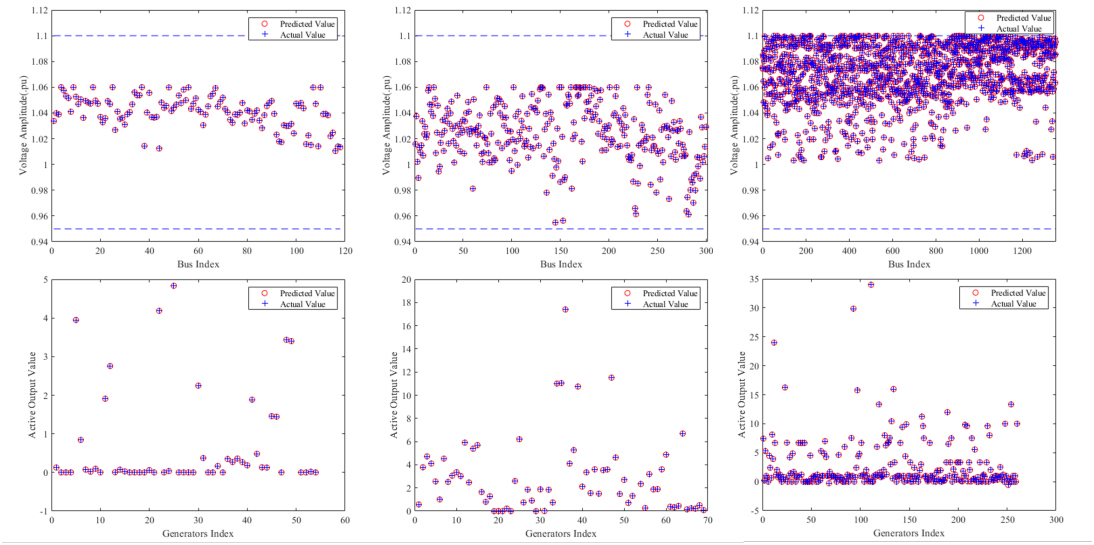
Case	Methods	APM	ΔP	ΔQ	ΔV	$\Delta\theta$
IEEE 118-bus sys.	M1	0.4139	0.1314	0.1951	0.0576	0.0944
	M2	0.7831	0.1804	0.2305	0.0808	0.1062
	M3	0.2975	0.0708	**0.1070**	**0.0548**	0.0472
	M4	**0.2503**	**0.0697**	0.1089	0.0571	**0.0472**
IEEE 300-bus sys.	M1	0.3517	1.0461	1.3131	0.7354	0.7500
	M2	0.4416	1.2404	4.8509	0.4678	**0.0095**
	M3	0.7165	3.1922	3.4901	0.9683	1.5300
	M4	**0.2113**	**0.7417**	**1.1617**	**0.3963**	0.5100
European 1354-bus sys.	M1	0.1554	1.3546	1.4536	0.1124	3.3850
	M2	0.2740	1.0667	1.0822	0.2003	1.6248
	M3	1.4249	1.5814	1.6006	**0.0996**	3.3850
	M4	**0.1474**	**0.7830**	**0.8559**	0.1414	**1.3450**

Table VII. Comparison of prediction errors with multistage learning process and modified loss function (MW, MVA,KV, deg).

Case	Methods	APM	ΔP	ΔQ	ΔV	$\Delta\theta$
IEEE 118-bus sys.	M4	0.2503	0.0697	0.1089	0.0576	0.0472
	M5	0.0302	0.0084	0.0117	0.0049	**0.0207**
	M6	**0.0280**	**0.0078**	**0.0102**	**0.0045**	0.0221
IEEE 300-bus sys.	M4	0.2113	0.7417	1.1617	0.3963	0.2006
	M5	0.0101	0.0343	0.0524	0.0022	0.0563
	M6	**0.0098**	**0.0270**	**0.0345**	**0.0022**	**0.0504**
European 1354-bus sys.	M4	0.1474	0.783	0.8559	0.1414	1.3450
	M5	0.0982	**0.0388**	0.0364	0.0031	0.0146
	M6	**0.0737**	0.0394	**0.0364**	**0.0009**	**0.0141**

M4 and M5 indicate that the proposed multi-stage prediction outperforms the single-stage approach in case 118, case 300, and case 1354 systems. Based on the comparison of the results from M5 and M6, it can be concluded that the modified loss function also contributes to performance improvement. Compared to the single-stage prediction of M4, the multi-stage prediction accuracy of M5 and M6 is improved many times. In all test comparisons, M6 has the lowest prediction error. In summary, the M6 method has a good prediction performance of less than 0.005 KV, 0.003 KV, and 0.001 KV for bus voltage magnitude in the three test systems. The overall prediction of active power is less than 0.04 MW. Figure 8 visualizes the predicted bus voltage, generator active output, and actual values under the M6 method, demonstrating excellent agreement between the predicted and actual values for these quantities.

4) Verification of using virtual buses to process partitions: To validate the performance of the proposed method, which employs virtual buses to process partitions for enhancing predictive capabilities, we compared the computational results of M6 and M7. Table VIII presents the comparative data of the M6 and M7 methods. Based on the comparison of the results of M6 and M7, it is evident that our proposed multistage prediction process exhibits certain performance improvements in the case 118, case 300, and case 1354 systems. In case 118, we can observe that the performance enhancement brought by M7 is almost negligible in terms of generator output predication. However, in the more intricate case 300 and case 1354 systems, there are significant performance boosts. It can be concluded that the utilization of virtual buses is more conducive to strengthening the model's ability to capture the correlation information between partitions, thereby boosting



(a) IEEE-118-bus sys. (b) IEEE-300-bus sys. (c) European-1354-bus sys.

Fig. 8. Comparison of predicted bus voltage magnitude and active generator power with actual values.

Table VIII. Prediction error improvement by introducing virtual buses (MW, MVA,KV, deg).

Case	Methods	ΔP	ΔQ	ΔV	$\Delta\theta$
IEEE 118-bus sys.	M6	0.0078	0.0102	0.0045	0.0221
	M7	**0.0077**	**0.0102**	**0.0020**	**0.0101**
IEEE 300-bus sys.	M6	0.0270	0.0345	0.0022	0.0504
	M7	**0.0124**	**0.0150**	**0.0020**	**0.0165**
European 1354-bus sys.	M6	0.0394	0.0364	0.0031	0.0146
	M7	**0.0203**	**0.0271**	**0.0021**	**0.0120**

the model's performance.

B. Feasibility analysis

To ensure that the predicted value $\hat{y}$ satisfies the OPF constraints, we've validated its constraints (6), (7), and (8) for the M2, M3, M4, M6, and M7 methods on the IEEE 118, IEEE 300, and European 1354 systems. It is noted that feasibility analysis does not include voltage and flow consistency for boundary buses, which are demonstrated in the last subsection. The results in Table IX indicate high satisfaction of the constraints for all three test systems, with the M6 and M7 exhibiting the highest percentage of constraint satisfaction, including a 100% satisfaction rate for its voltage constraint and a 99.9% satisfaction rate for the other constraints. Figure 9 provides a comparative figure of the constraint violation degree for various methods to clearly demonstrate the advantages of our proposed method.

C. Comparison with state-of-the-art approaches

The proposed method M7 is compared with contemporary methods (M8 and M9). The parameters of these methods are set in accordance with the original papers. Notably, all comparative experiments are conducted under the N-1 line-outage contingency scenario.

Table IX. Comparison of constraint satisfaction rates of several methods.

case	Methods	$\eta_V(\%)$	$\eta_{P_g}(\%)$	$\eta_{Q_g}(\%)$
IEEE 118-bus sys.	M2	99.7	98.8	98.7
	M3	99.4	98.7	98.6
	M4	99.9	99.8	99.4
	M6	**100**	**99.9**	**99.9**
	M7	**100**	**99.9**	**99.9**
IEEE 300-bus sys.	M2	99.3	98.6	99.1
	M3	99.2	98.4	98.8
	M4	99.6	99.3	98.9
	M6	**100**	**99.9**	**99.9**
	M7	**100**	**99.9**	**99.9**
European 1354-bus sys.	M2	98.2	97.9	97.9
	M3	98.3	98.1	98.0
	M4	98.9	98.6	98.2
	M6	**100**	**99.9**	**99.9**
	M7	**100**	**99.9**	**99.9**

As presented in Table X, the proposed method M7 showcases excellent performance in terms of optimality loss and constraint satisfaction ratio. Regarding the optimality gap metric, the M7 exhibits the smallest value. While all methods achieve favorable results in voltage prediction, with the constraint satisfaction rate reaching 100%, the M7 surpasses the others in terms of active and reactive generation constraints. In the test results of three systems, the M7 attains a satisfaction rate of 99.9% or higher. This superiority over other methods indicates that the M7 has superior overall performance.

D. Redispatch solution Time comparison

Table XI shows the average inference time of each system. The table also shows the solution time using traditional algorithms. Compared to traditional methods, when the number of epochs (set to 1200) is reasonable and the solution times are up to dozens or even hundreds of times faster than those of traditional methods, it is of great significance for the redispatch post contingency.

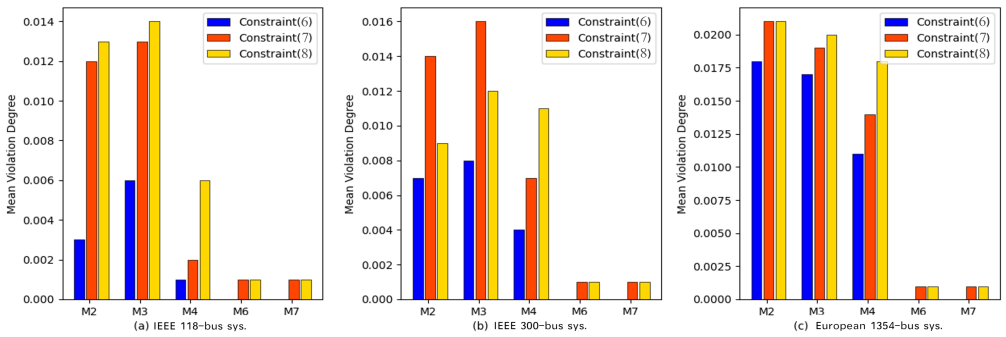


Fig. 9. Mean violation for violated AC-OPF constraints. The violation is expressed in kV, MW and MVA for constraints (6), (7), and (8), respectively.

Table X. Comparison between M7 and state-of-the-art methods.

case	Metric	M7	M8	M9
118-bus	$\eta_{opt}(\%)$	**<0.1**	-0.14	0.1
	$\eta_V(\%)/\nu_V(p.u.)$	**100/0**	100/0	100/0
	$\eta_{P_g}(\%)/\nu_{P_g}(p.u.)$	**99.9/0.0018**	99.14/0.0189	99.29/0.0156
	$\eta_{Q_g}(\%)/\nu_{Q_g}(p.u.)$	**99.9/0.0016**	99.12/0.0193	99.19/0.0178
300-bus	$\eta_{opt}(\%)$	**<0.1**	-0.12	0.11
	$\eta_V(\%)/\nu_V(p.u.)$	**100/0**	100/0	100/0
	$\eta_{P_g}(\%)/\nu_{P_g}(p.u.)$	**99.9/0.0012**	99.75/0.0050	99.9/0.0020
	$\eta_{Q_g}(\%)/\nu_{Q_g}(p.u.)$	**99.9/0.0014**	99.12/0.0176	99.3/0.0140
1354-bus	$\eta_{opt}(\%)$	**<0.1**	0.11	0.14
	$\eta_V(\%)/\nu_V(p.u.)$	**100/0**	100/0	100/0
	$\eta_{P_g}(\%)/\nu_{P_g}(p.u.)$	**99.9/0.0016**	99.28/0.0144	98.38/0.0324
	$\eta_{Q_g}(\%)/\nu_{Q_g}(p.u.)$	**99.9/0.0014**	99.34/0.0132	98.22/0.0356

Slash "/" is to split two metrics.

Table XI. Comparison between the average inference time of our proposed method and MIPS solution time.

Case	Proposed Method(sec)	AC-OPF(sec)	Speedup
IEEE 118-bus sys.	**1.6e-4**	0.0408	$\times 255$
IEEE 300-bus sys.	**6.2e-4**	0.2325	$\times 375$
EUR 1354-bus sys.	**3.3e-3**	1.3662	$\times 414$

VI. Conclusion

This paper presents a physics-informed deep learning approach to approximately solving the corrective AC power flow online. The paper outlines a multi-agent learning framework and introduces a two-stage learning process with a novel partition strategy. A customized loss function is proposed to enhance solution quality. Case studies in IEEE and European systems show the proposed approach outperforms other methods. Additionally, the proposed approach significantly improves computational speed, with online solution time tens to hundreds of times faster than traditional optimization algorithms.

Future studies can improve upon this method. At present, the method proposed in this paper mainly has two limitations:

- If the grid topology changes, the model needs to be retrained.
- Currently, we mainly consider the N-1 contingency in simulations.

We plan to investigate meta-learning or other emerging deep learning methods in future research to enhance applicability.

VII. Acknowledgment

We sincerely thank the reviewers' constructive and insightful comments, which help improve the quality of the manuscript.

References

[1] Y. L. Qiu, N. Deng, B. Wang, X. Shen, Z. Wang, N. Hultman, H. Shi, J. Liu, and Y. D. Wang, "Power supply disruptions deter electric vehicle adoption in cities in China," *Nature Communications*, vol. 15, no. 1, p. 6041, 2024.

[2] "Hurricane ida caused at least 1.2 million electricity customers to lose power - u.s. energy information administration (eia)." https://www.eia.gov/todayinenergy/detail.php?id=49556, Sept. 2021. (Accessed on 10/02/2024).

[3] ENTSOE, "Iberian Peninsula Blackout - 28 April 2025."

[4] M. Shahidehpour, H. Yamin, and Z. Li, *Market Operations in Electric Power Systems: Forecasting, Scheduling, and Risk Management*. Wiley-IEEE Press, 1st edition ed., 2002.

[5] S. Zhang, H. Ye, F. Wang, Y. Chen, S. Rose, and Y. Ma, "Data-aided offline and online screening for security constraint," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2614–2622, 2021.

[6] L. Roald and G. Andersson, "Chance-constrained ac optimal power flow: Reformulations and efficient algorithms," *IEEE Transactions on Power Systems*, pp. 1–1, 2017.

[7] Z. Yang, K. Xie, J. Yu, H. Zhong, N. Zhang, and Q. Xia, "A general formulation of linear power flow models: Basic theory and error analysis," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1315–1324, 2019.

[8] D. Elbrächter, D. Perekrestenko, P. Grohs, and H. Bölcskei, "Deep neural network approximation theory," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2581–2623, 2021.

[9] X. Pan, T. Zhao, and M. Chen, "Deepopf: Deep neural network for dc optimal power flow," in *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pp. 1–6, 2019.

[10] X. Pan, T. Zhao, M. Chen, and S. Zhang, "Deepopf: A deep neural network approach for security-constrained dc optimal power flow," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–1, 2020.

[11] A. Golshani, W. Sun, and K. Sun, "Advanced power system partitioning method for fast and reliable restoration: toward a self-healing power grid," *IET Generation Transmission & Distribution*, vol. 12, no. 1, pp. 42–52, 2018.

[12] B. Mahdad, T. Bouktir, and K. Srairi, "Optimal power flow using partitioning technique," in *First International Conference on Electrical Systems PCSE'05, Oum El Bouaghi University, Algeria, 2005*, 2005.

[13] N. Miiller and V. H. Quintana, "Network partition in power systems - sciencedirect," *Power Systems and Power Plant Control 1989*, pp. 87–92, 1990.

[14] S. B. Yusof and G. J. Rogers, "Slow coherency based network partitioning including load buses," *IEEE Trans Power Syst*, vol. 8, no. 3, pp. 1375–1382, 1993.

[15] Y. Yang, Y. Sun, Q. Wang, F. Liu, and L. Zhu, "Fast power grid partition for voltage control with balanced-depth-based community detection algorithm," *IEEE Transactions on Power Systems*, vol. 37, no. 2, pp. 1612–1622, 2022.

[16] G. Gurumoorthi, S. Senthilkumar, G. Karthikeyan, and F. Alsaif, "A hybrid deep learning approach to solve optimal power flow problem in hybrid renewable energy systems," *Scientific Reports*, vol. 14, p. 19377, Aug. 2024. Publisher: Nature Publishing Group.

[17] Y. Jia, X. Bai, L. Zheng, Z. Weng, and Y. Li, "ConvOPF-DOP: A data-driven method for solving AC-OPF based on CNN considering different operation patterns," *IEEE Transactions on Power Systems*, vol. 38, no. 1, pp. 853–860, 2022.

[18] F. Guo, B. Xu, W.-A. Zhang, C. Wen, D. Zhang, and L. Yu, "Training deep neural network for optimal power allocation in islanded microgrid systems: A distributed learning-based approach," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2021.

[19] L. Duchesne, E. Karangelos, and L. Wehenkel, "Recent developments in machine learning for energy systems reliability management," *Proceedings of the IEEE*, vol. PP, no. 99, pp. 1–21, 2020.

[20] Y. Zhai, H. Ye, L. Zhang, and Y. Ge, "Model information-aided deep learning for corrective ac power flow," in *2022 IEEE Power Energy Society General Meeting (PESGM)*, pp. 1–5, 2022.

[21] T. Pham and X. Li, "Reduced optimal power flow using graph neural network," in *2022 North American Power Symposium (NAPS)*, pp. 1–6, 2022.

[22] M. Gao, J. Yu, Z. Yang, and J. Zhao, "A physics-guided graph convolution neural network for optimal power flow," *IEEE Transactions on Power Systems*, vol. 39, no. 1, pp. 380–390, 2024.

[23] X. Pan, M. Chen, T. Zhao, and S. H. Low, "Deepopf: A feasibility-optimized deep neural network approach for ac optimal power flow problems," *IEEE Systems Journal*, vol. 17, no. 1, pp. 673–683, 2023.

[24] R. Nellikkath and S. Chatzivasileiadis, "Physics-informed neural networks for ac optimal power flow," *Electric Power Systems Research*, vol. 212, p. 108412, 2022.

[25] W. Huang, X. Pan, M. Chen, and S. H. Low, "Deepopf-v: Solving ac-opf problems efficiently," *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 800–803, 2022.

[26] M. Zhou, M. Chen, and S. H. Low, "Deepopf-ft: One deep neural network for multiple ac-opf problems with flexible topology," *IEEE Transactions on Power Systems*, vol. 38, no. 1, pp. 964–967, 2023.

[27] Y. Chen, Y. Shi, and B. Zhang, "Optimal control via neural networks: A convex approach," in *International Conference on Learning Representations*, 2019.

[28] H. Ye, "Surrogate affine approximation based co-optimization of transactive flexibility, uncertainty, and energy," *IEEE Transactions on Power Systems*, vol. 33, no. 99, pp. 4084–4096, 2018.

[29] Z. Yang, H. Zhong, A. Bose, T. Zheng, Q. Xia, and C. Kang, "A linearized opf model with reactive power and voltage magnitude: A pathway to improve the mw-only dc opf," *IEEE Transactions on Power Systems*, pp. 1–1, 2017.

[30] K. Y. Xiao, V. Tjeng, N. M. Shafiullah, and A. Madry, "Training for faster adversarial robustness verification via inducing relu stability," in *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May*, 2019.

[31] M. R. Hestenes, "Multiplier and gradient methods," *Journal of Optimization Theory and Applications*, vol. 4, no. 5, pp. 303–320, 1969.

[32] L. Yang and X. Zhou, "Constraint reformulation and a lagrangian relaxation-based solution algorithm for a least expected time path problem," *Transportation Research Part B: Methodological*, vol. 59, pp. 22–44, 2014.

[33] Y. Yang, Z. Yang, J. Yu, B. Zhang, Y. Zhang, and H. Yu, "Fast calculation of probabilistic power flow: A model-based deep learning approach," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2235–2244, 2020.

[34] T. Smed, "Feasible eigenvalue sensitivity for large power systems," *IEEE Transactions on Power Systems*, vol. 8, no. 2, pp. 555–563, 1993.

Yinyin Ge is currently an Associate Professor in the School of Cyber Science and Engineering, Xi'an Jiaotong University (XJTU). She received the B.S. degree in Automation and the M.S. degree in Systems Engineering from Jiaotong University, Xi'an, China, in 2008 and 2011, respectively, and the Ph.D. degree in Electrical Engineering from the Illinois Institute of Technology, Chicago, IL, USA, in 2016. She served as a Research Associate at Case Western Reserve University, Cleveland, OH, USA, before joining XJTU. Her research interests focus on cyber-physical energy systems (CPES), including optimization and privacy preservation of power and energy systems, cyber-physical security, and integration of smart grid and intelligent transportation systems.

Hongxing Ye (Senior Member, IEEE) is currently a Professor with Xi'an Jiaotong University (XJTU). He received the B.S. degree in Information Engineering and the M.S. degree in Systems Engineering from Xi'an Jiaotong University, Xi'an, China, in 2007 and 2011, respectively, and the Ph.D. degree in Electrical Engineering from the Illinois Institute of Technology, Chicago, IL, USA, in 2016. He was a visiting researcher at Argonne National Lab, Chicago. He was a Tenure-Track Assistant Professor with Cleveland State University, Cleveland, OH, USA, before joining XJTU. His research interests include power system planning and operation, and machine learning in cyber-physical energy system. Dr. Ye received the Sigma Xi Research Excellence Award. His research has been supported by NSFC, NSTMP, NSF, and Industry.

Xingtong Peng received the B.S. degree in Control from Chongqing University, Chongqing, and the M.S. degree in Cyber Science and Engineering from Xi'an Jiaotong University, Xi'an, China, in 2022 and 2025, respectively. He is currently an engineer at Huawei Technologies Co., Ltd., Shanghai, China. His research interests include machine learning and its applications in power systems.

Yadong Zhai received the B.S. degree in Control from Chongqing University, Chongqing, and the M.S. degree in Control Science and Engineering from Xi'an Jiaotong University, Xi'an, China, in 2021 and 2024, respectively. He is currently an engineer at Microsoft Corporation, Suzhou, Jiangsu, China. His research interests include machine learning and its applications in power systems.

Wangqing Mao received the B.S. and M.S. degrees in Electrical Engineering from North China Electric Power University, Baoding, China, in 2013 and 2016, respectively. She is currently working toward the Ph.D. degree in electrical engineering, Xi'an Jiaotong University, Xi'an, China. She is currently a science and technology planning and programming manager in Science and Technology Internet Department, State Grid Suqian Power Supply Company, Suqian, Jiangsu, China. Her research interests include renewable integration, and optimal control in power systems.