Spatio-Temporal Graph Neural Networks for Multi-Period Optimal Power Flow

Ali Rajaei
Delft University of Technology
Delft, The Netherlands
a.rajaei@tudelft.nl

Olayiwola Arowolo
Delft University of Technology
Delft, The Netherlands
o.a.arowolo@tudelft.nl

Jochen L. Cremer
Delft University of Technology
Delft, The Netherlands
j.l.cremer@tudelft.nl

2025 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe) | 979-8-3315-2503-3/25/$31.00 ©2025 IEEE | DOI: 10.1109/ISGTEurope64741.2025.11305408

***Abstract*—The increasing integration of renewable energy sources (RES) and the inter-temporal constraints of generation units necessitate real-time solutions to the AC multi-period optimal power flow (MP-OPF) problem. RES exhibit spatio-temporal correlations due to their geographically distributed nature and time-varying generation patterns. This paper proposes a novel graph recurrent neural networks (GRNN)-based approach to learn an optimization proxy for the AC MP-OPF problem. The proposed approach: (i) uses a graph attention mechanism to extract grid topology features, enhancing scalability for larger networks and improving topology adaptiveness; (ii) uses a recurrent structure to capture temporal correlations, and enable scalability for longer prediction horizons; and (iii) jointly consider spatial and temporal dependencies in end-to-end learning to improve prediction accuracy. Additionally, a feasibility restoration layer minimizes constraint violations during training and ensures feasibility during testing. Numerical results on the IEEE 118-bus and PEGASE 1354-bus systems demonstrate the superior performance of the proposed GRNN over the baseline neural architectures, achieving up to 50% lower prediction error, minor optimality gap of 0.5%, and 2-4 orders of magnitude speed-ups. Under N-1 line outages, the GRNN approach reduces the optimality gap by 4.5%, showcasing its robustness to topology changes. These results highlight the GRNN-FR as a promising approach for real-time applications in large-scale power networks, whether for fast warm-start initialization or rapid solution of numerous MP-OPF instances.**

***Index Terms*—Multi-period optimal power flow, Graph recurrent neural networks, Optimization proxy.**

I. Introduction

The ongoing energy transition increases the share of renewable energy sources (RES). However, the inherent intermittency and variability of RES pose significant operational challenges. In particular, in systems with a high share of RES, rapid fluctuations in the net load (demand minus renewable generation) force system operators to apply more re-dispatching and emergency costly measures. In addition, the growing presence of distributed energy resources (DER) with inter-temporal constraints, such as energy storage systems and flexible loads, increases the operational complexity of the system. Transmission system operators use the optimal power flow (OPF) optimization to schedule the committed resources in the system. As the share of RES and DER increases, the OPF problem extends to a multi-period OPF (MP-OPF) problem, considering inter-temporal constraints such as generator ramping limits. However, this extension would drastically increase the computational time of finding suitable schedules for large-scale systems [1], [2].

Recent research has used supervised machine learning (ML) to learn the mapping between net load input and generation output of the single-period OPF (SP-OPF) problem. Prior works train fully connected neural networks (FNN) to directly predict OPF solutions [3], [4], classify active sets or detect infeasibility [5], [6], and apply feasibility correction through post-processing, differentiable layers, or dual-based penalties [7], [8] to improve constraint feasibility. However, methods in [3]–[8] struggle to scale to large-scale networks and assume fixed topologies, despite practical transmission systems frequently undergoing changes [9].

Graph neural network (GNN) architectures are specifically suited for learning graph-structured problems. Works such as [10]–[13] use GNNs to exploit grid topology and solve OPF and unit commitment (UC) problems. However, these approaches do not consider the inter-temporal constraints of practical power systems.

With the growing integration of RES and DER, inter-temporal constraints are increasingly important in practical system operation, requiring MP-OPF solutions for day-ahead planning. A FNN in [14] predicts day-ahead generation, but trains a separate model per hour, limiting its ability to capture inter-temporal relations. [15] uses an RNN to learn temporal dependencies, though scalability and adaptability to topological changes remain challenges. To address this, spatio-temporal GNNs (ST-GNNs) have been proposed to capture the spatial structure and temporal dependencies of time-varying graph data [16], e.g., [17] presents an ST-GNN classifier to predict the on/off status of units in UC problems.

This paper develops a topology adaptive graph recurrent neural network (GRNN) model to approach the MP-OPF problem. A graph attention mechanism extracts the grid topology information to improve topology adaptability, while the recurrent structure captures temporal dependencies across time steps to predict the generation output for the next 24 hours. Despite the previous works with FNN [3]–[8], [14], GNN [10]–[13], and RNN [15] approaches for OPF problem; to the best of authors' knowledge, joint consideration of spatial and temporal information for the AC MP-OPF problem has

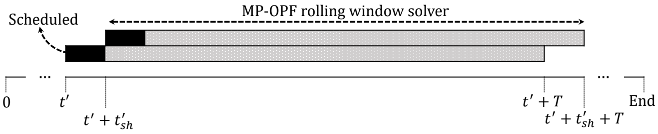

Fig. 1: The RHC-based solution algorithm for MP-OPF problem.

not been investigated before. By incorporating both temporal and spatial features in an end-to-end learning approach, our approach achieves better accuracy and scalability to larger networks and longer prediction horizons. To ensure feasibility, a restoration layer is considered to guide the training by minimizing the power balance and ramping violations, and ensuring feasibility during testing. Furthermore, receding horizon control (RHC) is employed during the MP-OPF optimization data generation to imitate a more realistic operation of the system [18]. Instead of focusing on day-ahead prediction, we train our model to predict for the next 24 hours at any given time of the day, ensuring adaptability to varying load patterns.

II. Multi-Period AC Optimal Power Flow

The MP-OPF determines the optimal dispatch of generators for a given time horizon to supply the demand while satisfying the operational limits of the generators and the network. Consider a transmission network with a set of buses (nodes) denoted by $i \in \mathcal{V}$, and a set of lines (edges) denoted by $l \in \mathcal{E}$. The set of nodes with a generator are called *generator nodes* and are represented by $i \in \mathcal{V}^g$. The time horizon of the problem is represented by $t \in \mathcal{T} = \{t', ..., t' + T\}$, while t' and T represent the current time step and number of time steps in the future, respectively. The net load demand at each node is $x_{i,t}$, and the generation output is $y_{i,t}$. The AC MP-OPF optimization with SOCP relaxation [19] is written as:

$$\min \sum_{i \in \mathcal{V}^g, t \in \mathcal{T}} c_i(y_{i,t}) \tag{1a}$$

$$\text{s.t.} \quad g(y_{i,t}) \leq 0 \qquad i \in \mathcal{V}, t \in \mathcal{T} \tag{1b}$$

$$h(x_{i,t}, y_{i,t}) \leq 0 \qquad i \in \mathcal{V}, t \in \mathcal{T} \tag{1c}$$

where the objective function (1a) minimizes the operational cost of the generators for the time horizon $\mathcal{T}$. (1b) denotes operational limits of the generators including minimum/maximum generation and ramping limits. (1c) presents the convex SOCP PF equations [19].

This paper uses RHC to solve the MP-OPF problem over a period of 1 year of historical demand data to generate a training data set for supervised ML models. Fig. 1 depicts a schematic of the RHC-based solution algorithm. Assume the system is operating at time t'. The RHC solver uses a rolling time window horizon $\mathcal{T} = \{t', t'+T\}$ to solve the problem for the next T steps. The dispatch solution for for the next t'_{sh} time steps (i.e., $\{t', t'+t'_{sh}-1\}$) is applied, where t'_{sh} is the number of scheduled time steps. Then, the current time step and the rolling window are shifted by $t' = t' + t'_{sh}$. Now, the MP-OPF is solved for the new time horizon $\mathcal{T} = \{t'+t'_{sh}, t'+t'_{sh}+T\}$, and the algorithm repeats until the final time step.

III. Proposed MP-OPF Optimization Proxy

This section aims to learn an optimization proxy for the MP-OPF problem. Here, we aim to learn a mapping $f : x \rightarrow \hat{y}$, where $x \in \mathbb{R}^{N \times T}$ represent a sequence of T input net loads on a N-node network and $y \in \mathbb{R}^{N^g \times T}$ represents the corresponding sequence of N^g generator outputs.

A. Graph Convolutional Neural Network

The power grid can be modelled as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where nodes $i \in \mathcal{V}$ represent substations and edges $l \in \mathcal{E}$ represent transmission lines. A layer of a GCN could be written as follows:

$$\hat{y} = \sigma(H(S)x) = \sigma(\sum_{k=0}^{K} H_k S^k x) \tag{2}$$

where S is a graph shift operator (GSO) such as the adjacency matrix, $H_k \in \mathbb{R}^{F_o \times F_i}$ is the learnable filter parameter (convolutional filter), K is the number of shift operations (or filter taps), and σ is a point-wise non-linear function. The GCN uses the GSO to propagate information between adjacent graph nodes.

B. Recurrent Neural Network

Given a sequence of F_i-dimensional data points $\{x_t\}_{t \in \mathcal{T}}$, a RNN learns the mapping to a *hidden state* $z_t \in \mathbb{R}^{F_h}$ and output $y_t \in \mathbb{R}^{F_o}$, using the current input x_t, and previous hidden state z_{t-1}, as shown:

$$z_t = \sigma(Ax_t + Bz_{t-1}) \tag{3a}$$

$$\hat{y}_t = \sigma(Cz_t) \tag{3b}$$

where $A \in \mathbb{R}^{F_h \times F_i}$, $B \in \mathbb{R}^{F_h \times F_h}$, and $C \in \mathbb{R}^{F_o \times F_h}$ are linear operators, and σ is a pointwise non-linear function.

C. The Proposed Graph Recurrent Neural Network

We propose a graph recurrent neural network (GRNN) to learn the MP-OPF optimization proxy. GRNN can adapt the recurrent operations of the RNN to consider the power system graph structure [16]. Considering the inherent graph structure and the recurrent structure of the MP-OPF data by GRNN enables us to 1) efficiently parameterize our model for MP-OPF problems with larger grid size and longer time horizon, and 2) achieve higher accuracy due to better inductive biases provided by the spatial and temporal information. Therefore, GRNN is a suitable model for the MP-OPF problem that is solved on large-scale power systems with spatial and temporal dependencies.

Assume that for each time step, the net load input data $x_t \in \mathbb{R}^{F_i}$, generator output data $y_t \in \mathbb{R}^{F_o}$, and hidden states $z_t \in \mathbb{R}^{F_h}$ could be seen as graph signals. Therefore, the hidden state can be calculated by parameterizing the linear maps A, B in (3a) to give the generic GRNN cell architecture written as:

$$z_t = \sigma(A(\Phi)x_t + B(\Phi)z_{t-1}) \tag{4}$$

where $A(\Phi)$ and $B(\Phi)$ are parameterized graph attention filters with K filter taps similar to (2). Here, Φ is the attention-learned graph, i.e., $S = \Phi$. The graph attention mechanism

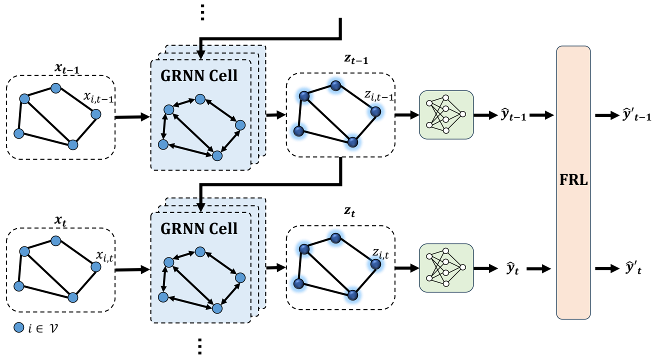

Fig. 2: The proposed GRNN-based approach for learning the MP-OPF optimization proxy. The GRNN layers take the current step's net load x_t and the previous step's hidden state z_{t-1} as inputs. The output from the final GRNN layer is processed by the FNN decoder to predict generator outputs $\hat{y}_t$, which are then refined by the FRL to ensure feasibility.

provides parametric edge weights from the node features by computing attention coefficients α_{ij} between nodes i and j:

$$\alpha_{ij} = \text{softmax}(e_{ij}) \tag{5a}$$

$$e_{ij} = \sigma(a^T (W x_i || W x_j)) \tag{5b}$$

where $||$ denotes concatenation, $W \in \mathbb{R}^{F_o \times F_i}$ is a learnable weight matrix, and $a \in \mathbb{R}^{2F_o}$ is a learnable attention vector [20].

Notably, the number of learnable parameters in the filters $A(\Phi)$ and $B(\Phi)$ do not depend on the MP-OPF time horizon T, nor the power system grid size N. This parameter sharing across the time and spatial dimension enables scalability to larger power systems and longer horizons for the task of MP-OPF prediction.

Fig. 2 presents the schematic of the proposed GRNN-based architecture for learning MP-OPF optimization proxy. The net load $x_{i,t}$ for each node $i \in \mathcal{V}$ and each time step $t \in \mathcal{T}$ is the input feature to the first GRNN layer, i.e. $F_i = 1$. The GRNN takes current step features x_t and previous step hidden features z_{t-1} as input and outputs the hidden state features z_t. The hidden state features have extracted the spatial and temporal dependencies in MP-OPF data. Furthermore, a *global* FNN is employed as a decoder to map the GRNN hidden state z_t into the target value of generator outputs $y_{i,t}$ for $i \in \mathcal{V}^g$ at the current time step t. Therefore, the output dimension of the FNN decoder layer is $F_o = N^g$.

D. Feasibility Restoration Layer

Supply-demand imbalances within power systems can trigger various operational issues. Additionally, violations of generator ramping constraints may cause generators to trip, further compromising system security. Therefore, prioritizing the feasibility of MP-OPF predictions takes precedence over the pursuit of absolute accuracy and optimality to ensure the secure operation of power systems. We use the implicit differentiable layers in [21] to develop a feasibility restoration layer (FRL). The FRL repairs the GRNN dispatch predictions to satisfy power balance and ramping constraints.

Assume a total input demand of x_t for $t \in \mathcal{T}$. Given a GRNN prediction $\hat{y}_{i,t}$, the FRL finds a feasible prediction $\hat{y}'_{i,t}$ by solving:

$$\min_{\hat{y}'_{i,t}} \quad \mathcal{L}^{FRL} = ||\hat{y}'_{i,t} - \hat{y}_{i,t}||_2 \tag{6a}$$

$$\text{s.t.} \quad h_i^r(\hat{y}'_{i,t}) \leq 0 \qquad i \in \mathcal{V}^g, t \in \mathcal{T} \tag{6b}$$

$$h^b(x_t, \hat{y}'_t) = 0 \qquad t \in \mathcal{T} \tag{6c}$$

where h^b models the total power balance equality constraint, and h_i^r denotes the generator ramping limit inequality constraints. The objective (6a) finds the closest feasible point to the GRNN prediction. Additionally, to enforce the minimum and maximum generation limits, the last output layer employs a Sigmoid activation function, which maps outputs to the range [0,1] and subsequently scales them to the specified limits.

Consider an MP-OPF training data set $\{(x, y)\}_\Omega$, including input net loads x and the corresponding generator outputs y. The proposed GRNN predicts outputs $\hat{y}$, and the FRL refines outputs $\hat{y}'$. The GRNN-FR model is trained by minimizing a loss function $\mathcal{L}(x, y, \hat{y}, \hat{y}')$ over the training data set Ω:

$$\mathcal{L}(x, y, \hat{y}, \hat{y}') = \mathcal{L}^o(y, \hat{y}) + \lambda \mathcal{L}^{FRL}(x, \hat{y}, \hat{y}') \tag{7}$$

where $\mathcal{L}^o$ represents standard ML loss functions, such as mean squared error, while $\mathcal{L}^{FRL}$ is the loss function in (6). λ is a hyper-parameter to balance the two loss terms.

IV. Numerical Results

A. Settings and Test Networks

The proposed GRNN-based model is tested on the IEEE 118-bus and PEGASE 1354-bus networks [22]. To provide time-varying input demand for the MP-OPF problem, the hourly net load profile of the Netherlands in 2021 [23] is used. Respectively, the 8760 instances of the hourly load demand profile are scaled down to the nominal demand values of the IEEE networks in addition to a random noise drawn from a Kumaraswamy(1.6,2.8) distribution with a Pearson correlation coefficient of 0.75 to model the correlation pattern between loads. Moreover, ramping limits of 8 hours are assumed for generators. In order to gather instance labels, the MP-OPF problem is solved by the RHC-based algorithm with a rolling time horizon of $\mathcal{T} = 6$, and $t'_{sh} = 4$ hours. Then, random sequences of $\mathcal{T} = 24$ hours are used for training and testing. The MP-OPF instances are modeled by Pyomo and solved with Gurobi 10.0 in Python 3.10.

The data set is randomly divided into 70% training, 15% validation, and 15% testing sets. ML models are implemented using PyTorch 2.0.0 and PyG 2.5.3 in Python. The mean squared error (MSE) is used as the loss function $\mathcal{L}^o$ during training. $\lambda = 0.01$ is assumed after a hyper-parameter optimization on the validation data set. All models are trained with the Adam optimizer for a maximum of 20 epochs with a learning rate of 10^{-3} and a weight decay of 10^{-5}. Training and testing are done on the DelftBlue supercomputer with Intel(R) Xeon(R) Platinum 8358P CPU at 2.60GHz, 32GB RAM, and NVIDIA V100S GPUs [24]. For case studies, all models are run over 5 random seeds for data shuffling and weight initialization. The reported results are the average over these runs. Testing accuracy is measured as the mean absolute error (MAE) normalized by the total input load.

The proposed GRNN approach is compared with four baseline neural architectures. The models predict for a time horizon of $\mathcal{T} = 24$ hours ahead on a N-node power system (graph), with N^g generators. The tested approaches are:

- Gurobi: the MP-OPF optimization (1) solved by Gurobi for a 24-hour horizon ($\mathcal{T} = 24$), serving as the ground truth.
- Gurobi-TD: the MP-OPF optimization decomposed using a temporal decomposition and solved with a rolling time horizon of $\mathcal{T} = 1$, and $t'_{sh} = 1$.
- FNN: takes a $N\times$24-dimensional vector input demand, and predicts a $N^g\times$24-dimensional vector output generation. This model includes 2 FNN layers with 64-dimensional hidden features.
- GCN: takes a 24-dimensional vector input demand over a N-node graph. This model includes 2 GCN layers with 64-dimensional hidden features and 3 filter taps. The output of GCN on generator nodes $\mathcal{V}^g$ is concatenated with the respective generator features and passed through a *local* FNN that outputs a 24-dimensional vector. Note that the local FNN shares parameters across different generator nodes [13].
- RNN: takes a sequence of 24 inputs of N-dimensional load input, and predicts a sequence of N^g-dimensional output generation. This model includes 2 RNN blocks with 64-dimensional hidden features, followed by a FNN to map the output to a N^g-dimensional vector.
- GRNN: takes a sequence of 24 inputs of 1-dimensional load input over a N-node graph. We use two GRNN layers with 64-dimensional hidden features and $K = 3$ filter taps. A *global* FNN maps the flattened output of the GRNN cell (z_t) to a N^g-dimensional vector.
- GRNN-FR: extends the GRNN model by considering a feasibility restoration layer (FRL).

B. Prediction Accuracy and Computational Efficiency

Table I compares different models for the 118-bus and 1354-bus networks. In the case of the 118-bus network, the proposed GRNN effectively captures spatio-temporal correlations and outperforms the FNN and RNN models, achieving 21% lower MAE. In contrast, the GCN demonstrates poor performance with a 10% optimality gap, primarily due to parameter sharing among generator nodes, which can cause issues when generator production levels have different biases. Additionally, the RNN and GRNN exhibit less infeasibility for balance and ramp constraints, as the ignorance of inter-temporal dependencies by the FNN and GCN models results in more constraint violations. The proposed GRNN-FR incorporates the FRL into the training loop via (7) to guide the GRNN towards producing feasible predictions. As a result, the GRNN-FR not only eliminates constraint violations during testing but also reduces the optimality gap to 0.5%. For the 1354-bus network, a similar pattern is observed in Table I, with the GRNN effectively capturing spatio-temporal dependencies and outperforming other models by reducing the MAE by 70%, and the optimality gap to 0.2%. Moreover, the GRNN-FR eliminates the constraint violations while slightly increasing the MAE and achieving a minor gap of 0.3%.

TABLE I: Performance comparison of models on the 118-bus and 1354-bus networks over a 24-hour horizon.

Sys	Model	MAE (%)	Gap (%)	Violation (%) Balance	Violation (%) Ramp	Time (sec)
118-bus	Gurobi	-	-	-	-	4.0E+00
	Gurobi-TD	0.015	0.4	0	0	7.6E+00
	FNN	0.101	2.5	2.4	0.2	2.6E-4
	GCN	0.233	10.7	9.6	0.5	1.8E-3
	RNN	0.033	0.6	1.2	0.1	2.3E-5
	GRNN	0.026	0.6	1.1	0.1	9.4E-3
	GRNN-FR	0.027	0.5	0	0	3.4E-2
1354-bus	Gurobi	-	-	-	-	2.2E+2
	Gurobi-TD	0.001	0.0	0	0	1.5E+2
	FNN	0.037	0.8	8.7	0.1	5.0E-4
	GCN	0.109	1.1	11.8	1.5	4.8E-3
	RNN	0.058	1.3	13.6	0.3	2.6E-5
	GRNN	0.011	0.2	2.0	0.0	1.0E-2
	GRNN-FR	0.018	0.3	0	0	4.1E-2

Regarding the computational time, all ML models achieve significant speed-ups of 2 to 5 orders of magnitude compared to the Gurobi for the 118-bus network. Moreover, the Gurobi-TD, which considers a temporal decomposition, even increases computational time while yielding a 0.4% optimality gap. For the 1354-bus system, the advantage of ML models is even more pronounced, achieving speed-ups of 4 to 7 orders of magnitude. In comparison, the Gurobi-TD achieves only a marginal improvement in computational time for the larger system, remaining insufficient for real-time applications. This highlights the scalability and efficiency of the proposed GRNN-based approach for large-scale networks.

C. Robustness to Topology Changes

The previous section investigated the effectiveness of the ML models for the MP-OPF problem under normal operating conditions with a fixed network topology. However, practical power grids frequently undergo topology changes, due to events such as maintenance, or N-1 line outages. This case study investigates the robustness of the proposed GRNN approach compared to other baseline neural architectures to N-1 line outage conditions. To this end, a new training dataset is generated, where a random non-radial line is taken out of service in 50% of the samples. Moreover, for the FNN and RNN models, an additional feature vector of size $|\mathcal{E}|$, with one-hot-encoding of topology changes, is appended to the input features; whereas GCN and GRNN models use the adjacency matrix and the attention coefficients, respectively.

Table II compares different models under N-1 line outages. For the 118-bus network, while the FNN and RNN provided high-quality predictions for a fixed topology in Table I, their performance significantly drops under topology changes, with large optimality gaps about 5% to 6%. This drop in accuracy

TABLE II: Performance comparison of models on the 118-bus and 1354-bus networks under N-1 line outages over a 24-hour horizon.

Sys	Model	MAE (%)	Gap (%)	Violation (%) Balance	Violation (%) Ramp
118-bus	FNN	0.142	6.0	6.2	0.2
	GCN	0.255	6.3	5.9	0.4
	RNN	0.121	5.2	5.8	0.0
	GRNN	0.038	0.5	1.7	0.0
	GRNN-FR	0.043	0.6	0	0
1354-bus	FNN	0.012	1.6	1.7	0.0
	GCN	0.145	8.6	8.5	0.2
	RNN	0.013	2.0	14.4	0.0
	GRNN	0.004	0.3	1.1	0.0
	GRNN-FR	0.006	0.4	0	0

is expected, as these models do not effectively consider the topology features. The GCN faces a similar challenge as in Table I due to the parameter sharing across nodes. In contrast, the graph attention mechanism of the proposed GRNN effectively extracts graph structure information, achieving a minor optimality gap of 0.5% and outperforming the baseline models. Additionally, the GRNN-FR satisfies the ramping and total balance constraints, albeit with a slight increase in the optimality gap to 0.6%. For the 1354-bus system, the proposed GRNN similarly outperforms the baseline neural architectures. However, the difference in performance is less pronounced, as an N-1 line outage in a larger grid has a relatively smaller impact on generator dispatches compared to a similar outage in a smaller grid.

V. Conclusion

This paper proposes a novel GRNN-based approach for learning an optimization proxy for the AC MP-OPF problem with inter-temporal constraints. The proposed GRNN model considers the inherent graph structure and the recurrent structure of the MP-OPF data to effectively capture spatio-temporal dependencies. Additionally, a feasibility restoration layer (FRL) minimizes constraint violations during training and eliminates them during testing. Numerical case studies on the IEEE 118-bus and PEGASE 1354-bus system demonstrate that the proposed GRNN-FR outperforms the baseline learning models, achieving up to 50% lower MAE, a minor optimality gap of 0.2-0.5%, and 2-4 orders of magnitude speed-ups. Furthermore, the graph attention mechanism of the GRNN model improves robustness to N-1 line outages, reducing the optimality gap by up to 4.5%. These results highlight the GRNN-FR as a promising approach for real-time applications in large-scale networks, whether as a fast warm-start solution or for rapidly solving numerous MP-OPF instances, such as in stochastic optimization. However, two main limitations remain. First, the GRNN's transductive design limits its generalization to new power systems and cost functions. Second, incorporating the FRL increases training time due to the need to solve an LP in each forward pass. Future work should explore inductive GRNN architectures and develop more efficient training methods for better transferability and scalability.

References

[1] M. I. Alizadeh and F. Capitanescu, "A tractable linearization-based approximated solution methodology to stochastic multi-period ac security-constrained optimal power flow," *IEEE Transactions on Power Systems*, 2022.

[2] A. Rajaei, O. Arowolo, and J. L. Cremer, "Learning-accelerated admm for stochastic power system scheduling with numerous scenarios," *IEEE Transactions on Sustainable Energy*, 2025.

[3] A. S. Zamzam and K. Baker, "Learning optimal solutions for extremely fast ac optimal power flow," in *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 2020, pp. 1–6.

[4] 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. 36, no. 3, pp. 1725–1735, 2020.

[5] D. Deka and S. Misra, "Learning for dc-opf: Classifying active sets using neural nets," in *2019 IEEE Milan PowerTech*. IEEE, 2019, pp. 1–6.

[6] A. Lotfi and M. Pirnia, "Constraint-guided deep neural network for solving optimal power flow," *22nd Power Systems Computation Conference (PSCC 2022)*, vol. 211, p. 108353, 2022.

[7] P. L. Donti, D. Rolnick, and J. Z. Kolter, "Dc3: A learning method for optimization with hard constraints," in *International Conference on Learning Representations*, 2020.

[8] A. Velloso and P. Van Hentenryck, "Combining deep learning and optimization for preventive security-constrained dc optimal power flow," *IEEE Transactions on Power Systems*, vol. 36, no. 4, pp. 3618–3628, 2021.

[9] F. Bellizio, J. L. Cremer, and G. Strbac, "Machine-learned security assessment for changing system topologies," *International Journal of Electrical Power & Energy Systems*, vol. 134, p. 107380, 2022.

[10] F. Diehl, "Warm-starting ac optimal power flow with graph neural networks," in *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, 2019, pp. 1–6.

[11] D. Owerko, F. Gama, and A. Ribeiro, "Optimal power flow using graph neural networks," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5930–5934.

[12] S. Liu, C. Wu, and H. Zhu, "Topology-aware graph neural networks for learning feasible and adaptive ac-opf solutions," *IEEE Transactions on Power Systems*, 2022.

[13] S. Park, W. Chen, D. Han, M. Tanneau, and P. Van Hentenryck, "Confidence-aware graph neural networks for learning reliability assessment commitments," *IEEE Transactions on Power Systems*, 2023.

[14] W. Chen, S. Park, M. Tanneau, and P. Van Hentenryck, "Learning optimization proxies for large-scale security-constrained economic dispatch," *22nd Power Systems Computation Conference (PSCC 2022)*, vol. 213, p. 108566, 2022.

[15] M. Mohammadian, K. Baker, M. H. Dinh, and F. Fioretto, "Learning solutions for intertemporal power systems optimization with recurrent neural networks," in *2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. IEEE, 2022, pp. 1–6.

[16] M. Sabbaqi and E. Isufi, "Graph-time convolutional neural networks: Architecture and theoretical analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[17] A. V. Ramesh and X. Li, "Spatio-temporal deep learning-assisted reduced security-constrained unit commitment," *IEEE Transactions on Power Systems*, 2023.

[18] Y. Huang, D. Gordon, and P. Scott, "Receding horizon dispatch of multi-period look-ahead market for energy storage integration," *Applied Energy*, vol. 352, p. 121856, 2023.

[19] S. H. Low, "Convex relaxation of optimal power flow—part i: Formulations and equivalence," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 1, pp. 15–27, 2014.

[20] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[21] A. Agrawal, B. Amos, S. Barratt, S. Boyd, S. Diamond, and J. Z. Kolter, "Differentiable convex optimization layers," *Advances in neural information processing systems*, vol. 32, 2019.

[22] S. Babaeinejadsarookolaee *et al.*, "The power grid library for benchmarking ac optimal power flow algorithms," *arXiv preprint arXiv:1908.02788*, 2019.

[23] ENTSO-E, "Power statistics," *https://www.entsoe.eu/data/power-stats/*.

[24] Delft High Performance Computing Centre (DHPC), "DelftBlue Supercomputer (Phase 1)," https://www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase1, 2022.