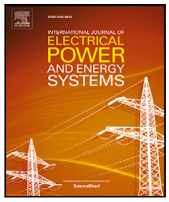
Electrical Power and Energy Systems 147 (2023) 108780

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes



Enhanced deep neural networks with transfer learning for distribution LMP considering load and PV uncertainties ☆

Boming Liu [a], Jin Dong [a,*], Jianming Lian [a], Teja Kuruganti [b], Xiaofei Wang [c], Fangxing Li [c]

[a] Electrification and Energy Infrastructure Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[b] Computational Sciences & Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[c] Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996, USA

ARTICLE INFO

Keywords:
Distribution locational marginal price
Leaning-assisted power system operation
Long short-term memory
Demand response

ABSTRACT

As the flexibility of generation and demand increases in distribution systems, the residential loads are emerging as a promising means to participate in demand response and the transactive energy market. Market pricing is an instrumental mechanism for the distribution system operator to exploit the full potential of the flexible resources. The distribution locational marginal price (DLMP) can be used to guide the residential load consumption. This type of market signal helps the distribution system operator to optimize the scheduling of all resources while satisfying related network constraints through a day-ahead market. However, solving the optimization problem for large-scale systems can be computationally expensive. To address the scalability and practicability limitations of the DLMP framework, a learning-based approach is proposed in this paper to complement the day-ahead distribution market framework. The proposed approach combines long short-term memory and transfer learning to develop deep neural network that can capture the spatial–temporal correlation of the input data. The model can determine the optimal DLMP for each node in a distribution system without the system parameters required to formulate the optimization problem. Testing results on IEEE 33-bus and 123-bus systems show that the proposed approach can generate a comparable DLMP against the optimization solutions.

1. Introduction

As distribution networks transition from passive components of the power system to active players in the electricity market, new opportunities are opening for consumers to participate in and take advantage of electricity cost savings while providing various grid services. The electricity market is designed for economic dispatch in power systems to achieve the supply–demand balance by solving the optimal power flow (OPF) problem [1]. The locational marginal price (LMP) has been well implemented by independent system operators/regional transmission organizations at whole-sale transmission-level power markets in the past decade [2]. With the increasing use of distributed generation and flexible load in the distribution networks, researchers have extended LMP to distribution LMP (DLMP) considering the different characteristics of the modern distribution systems [3–10]. DLMP is a pricing signal for consumers to make economic decisions by adjusting their electricity load. Because of the uncertainty introduced from increasing penetration of flexible loads and renewables such as PVs, the OPF problem needs to be solved more frequently to achieve optimal dispatching at the distribution level. However, the high computation complexity of the OPF problem can be a major obstacle to practical applications for large-scale systems [11,12].

1.1. Related work

Various learning-based approaches has been proposed to tackle these subjects. Deep neural networks (DNNs) are used to predict the result of OPF problems without solving the optimization problem. Liu et al. [13] proposed a learning-based aggregated response load model

☆ This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (hxxp://energy.gov/downloads/doe-public-access-plan). This material is based upon work supported by the US Department of Energy (DOE), Grid Modernization Laboratory Consortium (GMLC), DOE Energy Efficiency and Renewable Energy, Building Technology Office, and DOE Office of Electricity.

* Corresponding author.
E-mail address: dongj@ornl.org (J. Dong).

https://doi.org/10.1016/j.ijepes.2022.108780
Received 9 August 2022; Received in revised form 6 October 2022; Accepted 5 November 2022
Available online 1 December 2022

that can participate in a transactive energy market. García et al. [14] presented a learning-based methodology to forecast different components of the LMP. Pan et al. [15] determined the OPF solution via a DNN, which learns the mapping between the load inputs and decision variables. Huang et al. [16] demonstrated that the DNN-based approach can achieve a speed increase up to three orders of magnitude compared with conventional OPF solvers. Chatzos et al. [17] provided a two-stage distributed learning approach to predict AC–OPF solutions that exploit a spatial network decomposition. Owerko et al. [18] and Liu et al. [19] leveraged graph neural networks to predict OPF solutions with grid topology-based features, the approach directly predicted the LMPs, which are the actual OPF outputs for the electricity market. These works show that learning-based approaches can predict reasonable OPF solutions while reducing the computational burden. Most of the mentioned learning-based approaches focus on predicting OPF solutions at the transmission level, and limited research has been done on the distribution level considering the bidding of flexible load.

Moreover, most of the existing learning-based methods can only be applied on the source system and not on a different system with different topology and parameters. Transfer learning is an emerging training method that retains and reuses previously learned information to solve new problems. Transfer learning can adapt an old model to a new situation with fewer training data and less training time, and it enables the model to learn from the information of the source domain to solve a problem in the target domain [20]. Transfer learning has been widely applied on image recognition and text classification. It has also been used on different power system-related learning tasks [21–27]. A gradient-enhanced physics-informed neural networks is proposed to solve OPF problem in [28], transfer learning is applied for different initial conditions. Zhang et al. [29] demonstrated that the convergence rate can be dramatically accelerated by extreme transfer learning for the proposed multi-agent optimizer for decentralized optimal carbon-energy combined flow. An optimal scheduling strategy for microgrids based on the deep deterministic policy gradient and transfer learning was proposed by Fan et al. [30], which effectively used the existing scheduling knowledge. Ren and Xu [31] applied a transfer learning-based data-driven method for pre-fault dynamic security assessment of power systems, which enabled the dynamic security assessment to be used for unknown different but related faults by reducing the distribution difference between the trained data and unknown data. These studies demonstrate the significance of transfer learning in boosting the performance of the learning model through effectively using the knowledge from the source domain. However, in general, the application of transfer learning in the field of power systems is still in the exploratory stage.

1.2. Challenges and opportunities

In the previous tri-level bidding and dispatching framework [32], the DLMP was used as a price signal to guide the electricity consumption of the customers. The optimal DLMP was obtained by solving a bi-level optimization problem. A key feature of the DLMP algorithm is to design a customized utility tariff based on the temporal features of the whole network and the spatial patterns of the loads/generation across the whole distribution network. Such a comprehensive DLMP solution has great potential to provide benefits to both the utility company and end users. Three main barriers exist for DLMP to be widely adopted:

- For a large-scale system, solving the optimization problem can be time-consuming, or a feasible solution may not be obtained with limited computation power.
- For an arbitrary distribution network, the system parameters that are required to formulate the optimization problem may not all be available/observable.
- Utility companies require a plug-and-play solution that can efficiently and robustly generate DLMP curves for real-world operation scenarios involving uncertainties or even faults.

To address these challenges, a data-driven approach was proposed to determine the optimal DLMP for each node in a distribution network. The proposed approach combines long short-term memory (LSTM) units and transfer learning to develop a DNN that can capture the temporal dependencies of the time series inputs, as well as the spatial patterns of the distribution network. Thus, the DLMP of each node can be approximated based on certain input features without solving the OPF problem. To increase robustness of the DNN method against uncertainty, a sliding window method is demonstrated to mitigate the effect of uncertainty/noise in the prediction profiles of PVs and flexible loads. In addition, to address the insufficient training data in practice, the proposed DNN model can be applied on different distribution systems through transfer learning, which enables a pretrained DNN to learn the features from the target system more efficiently. The proposed approach provides a possible pathway to transfer knowledge between systems with different topologies by isolating the topology related inputs with an additional layer. By doing this, it allows one to reuse the common relationship between certain inputs and the optimal results from one system to another.

1.3. Contribution

Opposing to directly employing classical OPF and bi-level optimization frameworks to calculate the DLMP, we derive a data-driven solution to reduce the modeling and computational complexities of state-of-the-art DLMP solution techniques, thereby improving scalability and practicability. The proposed approach combines LSTM and transfer learning to develop DNNs that can capture the spatial–temporal correlations of the input data. The principle contributions of this work are summarized as follows:

1. **Data-driven with machine learning:** A DNN model was developed that considers both temporal dependencies of the time series inputs and spatial characteristics of the distribution network. The proposed DNN model can approximate the optimal DLMP across different nodes in a distribution system with acceptable errors and significantly reduce the computation time compared with the optimization-based approach.
2. **Robust with rolling horizon:** An ancillary rolling horizon approach was applied to increase the robustness of the proposed data-driven model. Through multiple simulations, the effect of rolling horizon window length selection against the uncertainty of the inputs such as PVs and flexible loads is demonstrated.
3. **Adaptive with transfer learning:** A transfer learning method was applied on the proposed data-driven model by leveraging pretrained model parameters, which address the issue of insufficient training data and improve the adaptiveness and practicality of the proposed approach on different distribution systems. The performance and effectiveness of the transfer learning were evaluated through different simulation scenarios with PV generation and flexible loads. Simulation results on various cases show that the proposed approach can generate robust pricing decisions with less than 5% error compared with the benchmark data provided by the optimization-based approach. In addition, the proposed learning-based DLMP framework can be quickly and conveniently deployed on a new target distribution network with very limited data by leveraging the transfer learning techniques.

The rest of this paper is organized as follows. Section 2 describes the tri-level bidding and dispatching framework and the formulation of the problem. Section 3 proposes the data-driven approach to obtain an optimal DLMP. Case studies on IEEE 33-bus and 123-bus systems in Section 4 demonstrate the effectiveness of the proposed approach. Section 5 provides conclusions.

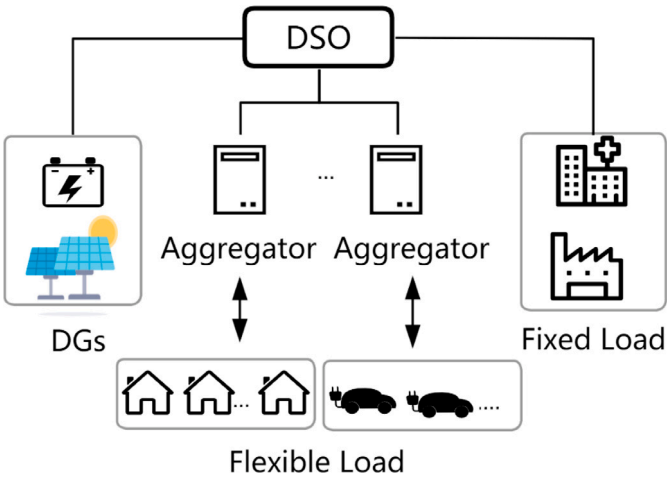


Fig. 1. Tri-layer day-ahead distribution-level electricity market.

2. Tri-level DLMP framework

A tri-level bidding and dispatching framework based on competitive distribution operation with a DLMP was established by Wang et al. [32]. Various market participants are shown in Fig. 1. The first two levels of the market form a bi-level model to optimize the aggregators' payment and to represent the interdependency between load aggregators and the distribution system operator (DSO) using the DLMP. The objective of the first level is to minimize the aggregators' power purchasing cost:

$$\min \sum_{t\in T}\left(\sum_{i\in H}\pi_{i,t}^{p}\cdot P_{i,t}^{Agg}\right) \tag{1}$$

where $P_{i,t}^{Agg}$ is the active power of the aggregator i at time t, $\pi_{i,t}^{p}$ is the active DLMP of node i at time t. The objective of the second level is to minimize system's generation cost while maintaining all operational constraints:

$$\begin{aligned}&\min \sum_{t\in T}\sum_{i\in G} c_{i,t}\cdot P_{i,t}^{G}+d_{i,t}\cdot Q_{i,t}^{G}\\ &where\ G=\{sub, PV, MT, SVC\}\end{aligned} \tag{2}$$

s.t.

$$\sum_{i\in G}P_{i,t}^{G}-\sum_{i\in F}P_{i,t}^{D}-\sum_{i\in H}P_{i,t}^{Agg}-P_{t}^{loss}=0:\ \lambda_{t}^{p},\ \forall t\in T \tag{3}$$

$$\sum_{i\in G}Q_{i,t}^{G}-\sum_{i\in F}Q_{i,t}^{D}-Q_{t}^{loss}=0:\ \lambda_{t}^{q},\ \forall t\in T \tag{4}$$

$$V_{j,t}=V_{1,t}+\sum_{i\in B}Z_{j,i}^{p}\left(P_{i,t}\right)+\sum_{i\in B}Z_{j,i}^{q}\left(Q_{i,t}\right) \tag{5}$$

$$V^{\min}\le V_{j,t}\le V^{\max},\forall j\in B,\ \forall t\in T \tag{6}$$

$$P_{i,t}^{G,\min}\le P_{i,t}^{G}\le P_{i,t}^{G,\max},\forall i\in sub,\ \forall t\in T \tag{7}$$

$$Q_{i,t}^{G,\min}\le Q_{i,t}^{G}\le Q_{i,t}^{G,\max},\forall i\in sub,\ \forall t\in T \tag{8}$$

where $P_{i,t}^{D}$ and $Q_{i,t}^{D}$, are fixed active and reactive load demand of node i at time t. Eqs. (3), (4) are the power balance constraints, Eq. (5) is the voltage expression derived from linearized power flow for distribution (LPF-D) [4]. Eqs. (6) and (7), (8) are the voltage constraints and power outputs limits respectively. Other constraints are omitted to conserve space. Details can be found in [32]. The third level dispatches the optimal load aggregation to all residents by the proposed priority list-based demand dispatching algorithm. The first level minimizes residents' electricity payment by optimizing aggregated residential flexible load schedules according to the DLMP in the second level. The second level clears the day-ahead distribution market and integrate power losses and voltage constraints in the DLMP. Because the first two levels are coupled, they are solved by reformulating it as a single-level mathematical programming with equilibrium constraints by Karush–Kuhn–Tucker optimality conditions and then mixed-integer linear programming by the big-M method. Once the optimal aggregated HVAC/electric vehicle schedule is obtained, the third level is solved to dispatch all participating residents based on the aggregated schedule.

In research from Wang et al. [32], the objective of the market clearing model is to minimize systems' generation cost while satisfying all the constraints. An appropriate pricing signal can efficiently guide consumers' load consumption and maintain economic operation of the distribution system. Therefore, the major goal of this paper is to learn the mapping between the given system profiles and optimal operating decision, namely DLMP in this study. Specifically, correlated variables in the optimization problem were taken as inputs and learn the dependency between these variables and the optimal result through a neural network. Neural networks have powerful learning abilities in approximating complex and nonlinear relationships in various domains and applications. Fig. 2 shows the schematic of the proposed data-driven approach. Information related to the constraints of the optimization problem, such as PV/DG capacity, PV/DG price, the amount of fixed load, and the flexible load ratio, are fed into the DNN model to approximate the appropriate DLMP. Based on the powerful learning ability of DNNs, such a data-driven method, which focuses on the prediction accuracy and the dependency between variables of the optimization problem and the constraints, can be applied.

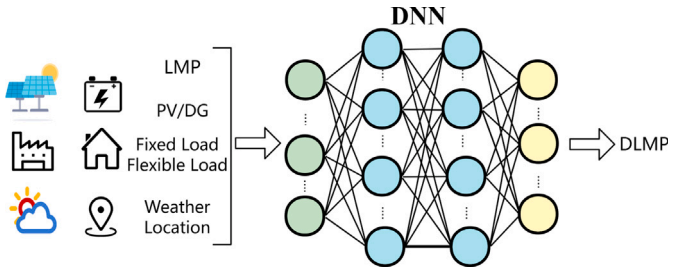


Fig. 2. Schematic of the data-driven approach.

3. Data-driven approach for DLMP

This section describes the proposed data-driven approach. The detailed structure of the LSTM adopted in this study is presented, followed by a modified network topology that enables the feasibility of the proposed method on different distribution systems through transfer learning.

3.1. LSTM-based learning approach

To obtain the day-ahead DLMP for given system conditions, a data-driven approach using LSTM units is proposed. The method can excavate the temporal correlation of the DLMP, and it builds and trains a neural network to accurately predict the day-ahead DLMP with the given distribution system conditions.

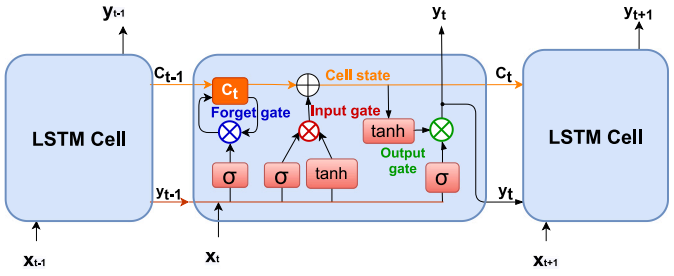


Fig. 3. LSTM cells and their internal structure.

Recurrent neural networks with LSTM units are widely used for time series prediction [33,34]. As shown in Fig. 3, LSTM units take the form of a chain of repeating LSTM cells. The structure of LSTM units are explicitly designed to learn long-term dependencies. Different from standard recurrent neural networks, the repeating cell in the LSTM network contains multiple interacting layers, which enables the LSTM cells to include more information. In this figure, σ denotes a sigmoid function, and tanh denotes a hyperbolic tangent function. Each LSTM cell has three inputs: the input vector x_t, the cell state vector C_{t-1} and the output vector y_{t-1}, and two outputs C_t and y_t. With the help of cell state memory, LSTM units can selectively remember and forget information. The previous output y_{t-1} and the new input the x_t are concatenated and fed into the LSTM cell. The horizontal line in the top of each cell carries information about the cell state which is the key to LSTM units. LSTM can adaptively add and remove information to the cell state by regulating different gates as shown in Fig. 3. These gates are multiplicative units to determine the flow of information into and out of cell activations based on importance. The importance is determined by the weights that are adjusted during the learning process. The forward pass of different gates in the LSTM cell can be described through the following equations:

Forget gate:

$$f_t = \sigma(W_f \cdot [y_{t-1}, x_t] + b_f) \tag{9}$$

Input gate:

$$i_t = \sigma(W_i \cdot [y_{t-1}, x_t] + b_i) \tag{10}$$

Output gate:

$$o_t = \sigma(W_o \cdot [y_{t-1}, x_t] + b_o) \tag{11}$$

Cell input:

$$\widetilde{C}_t = \tanh(W_c \cdot [y_{t-1}, x_t] + b_c) \tag{12}$$

State update:

$$C_t = i_t * \widetilde{C}_t + f_t * C_{t-1} \tag{13}$$

$$y_t = o_t * \tanh\left(C_t\right) \tag{14}$$

where W_i, W_f, W_o, W_c and b_i, b_f, b_o, and b_c are the weight matrices and the bias vectors that will be learned during the training process; i_t, f_t, and o_t are the gating vectors; $\widetilde{C}_t$ is a vector of new candidate state values; and $*$ represents the element-wise product of the vectors. According to the bi-level optimization results from Eqs. (1)–(8), these weights and bias of LSTM units will be trained based on the inputs features of different nodes and the optimal DLMP results.

3.2. Rolling horizon

In practical applications, the uncertainty of the input parameters must be considered. Some of the variables (e.g., the weather forecast of the next 24 h ambient temperature) may not be accurate because of random uncertainties. Therefore, the accuracy of the optimal 24 h DLMP obtained from the LSTM model can be affected. To mitigate the negative effect induced by the uncertainty or prediction error of different input features, a rolling horizon updating method was applied. For a sliding window with the size of k hours, the inputs of the LSTM model were updated at the beginning of every k hours to predict the 24 h-ahead DLMP. To represent the uncertainty of the 24 h input features, random prediction noise was added to several inputs. The noise level was designed to increase with time which represents greater uncertainty (inputs of the twenty-fourth hour usually have larger noise than the previous hours).

3.3. Transfer learning

A major challenge to data-driven learning methods is the long training time and insufficient training data. And majority of learning-based approaches cannot to be easily applied on a system with different topology without retraining. Considering the practical implementation of the proposed method on different distribution systems, transfer learning was applied to the data-driven model. Transfer learning has been used in various topics in the field of power system [35] such as reactive power optimization, load forecasting, renewable resources power prediction [36], carbon-energy combined-flow optimization [29]. Transfer learning can help generalize learning models to different boundary conditions, which significantly enhance the computational efficiency, and the utilization of the transfer learning leads to an initial super-convergence to a relatively low training error [28]. Another benefit of transfer learning is that it functions well with a limited training data set.

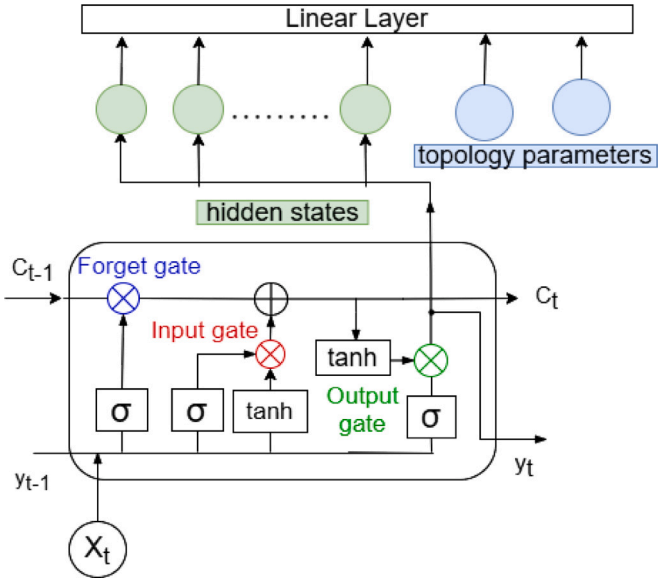


Fig. 4. Detailed network topology.

Although the basic LSTM framework can capture the time-varying impacts of some inputs, the impact of spatial information needs to be embedded so that the learning model can be adopted on a different system model while sharing some common relations learned from the other systems. The accuracy of the learning result will be affected if these spatial relationships are ignored [37]. Such information includes the locations of nodes and DGs. Because of the flexibility of the neural networks, spatial and temporal correlation for the optimal DLMP, we modify the commonly used LSTM to a topology-aware LSTM which is suitable for transfer learning. In the proposed training model, an additional layer was added to the LSTM model. The input features, which are related to the network topology of the distribution system, were directly fed into this new layer as shown in Fig. 4. When the target system provides only limited data, the pretrained LSTM network from a source distribution system can be used as a transferable model that captures the time series behavior. Combined with the new topology parameters from the target system, the pretrained weights of the LSTM can be leveraged on the training of the target system through transfer learning. This proposed neural network topology not only reduces the amount of data required for implementing on a new system, but it also reduce the training time by sharing the pretrained weights from a source system.

The overview of the transfer learning approach is shown in Fig. 5. The LSTM layers capture the relation of the time series inputs with the output DLMP, and the linear layer captures the effect of topology parameters of different systems on DLMP. When fine-tuning the weights

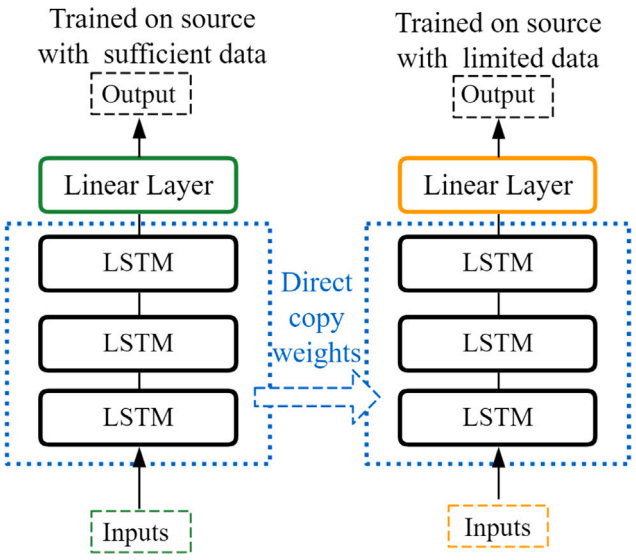


Fig. 5. Overview of the transfer learning approach.

on a new distribution system without sufficient data, the weights inside the blue dashed line are directly transferable. During transfer learning for the target system, the model on the right freezes the weights of the LSTM layers, which are directly copied from the model on the left as shown in Fig. 5, keeping the information learned from the source system. Then, the weights of the last layer (orange box) are trained with limited data from a target system. The pretrained LSTM helps the model to transfer the temporal features from the source system; the fine-tuning of last layer helps the model to learn the spatial relation between the topology features and the DLMP of the target system.

3.4. Performance metrics

The proposed approach to predicting temperature is a regression method. Therefore, to evaluate the error, simulation results were used as benchmark, The Root Mean Square Error (RMSE) between the predicted value and the optimization result is calculated as shown in Eq. (15).

$$A^{RMSE} = \sqrt{\frac{1}{n}\sum_{t=0}^{n}\left(A_t^{predict} - A_t^{optimization}\right)^2} \tag{15}$$

where A is the metric of interest, and n is the total number of samples.

4. Case studies

The overall training and testing workflow of the proposed data-driven approach is shown in Fig. 6. DLMP results from the optimization problem were passed to the DNN model for training and testing. The performance of the proposed data-driven method was first tested on the IEEE 33-bus distribution system. Next, the neural network model with pretrained parameters was trained and tested on the IEEE 123-bus distribution system through transfer learning. Considering the uncertainty of some inputs (e.g., LMP and outdoor temperature), the performance of a rolling horizon method was examined. The improvement in terms of training time and prediction accuracy through the proposed transfer learning approach was also presented. Model training and validation were preformed on an Intel i7-6700 CPU @ 3.60 GHz with 16 GB memory. The neural network model was developed in Python 3.6 with Pytorch.

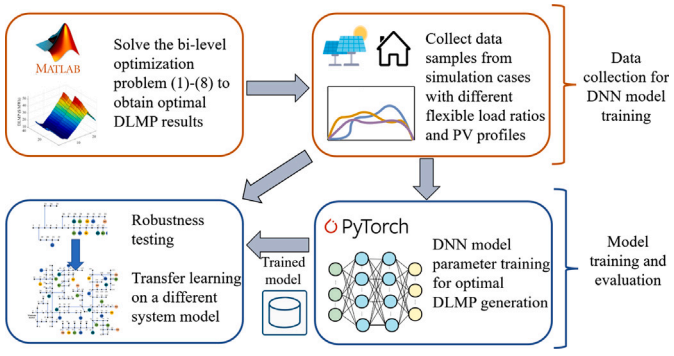


Fig. 6. The training and testing workflow.

Table 1
Inputs for different nodes.

Fixed load
PV/DG max power
PV/DG bidding price
Base voltage
Temperature
Flexible load ratio
Distance to the substation
Distance to DG
LMP

4.1. Testing on the IEEE 33-bus distribution system

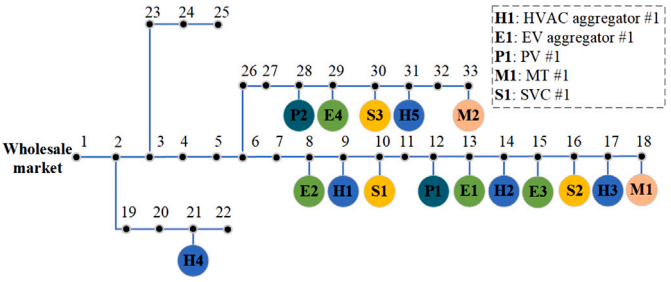


Fig. 7. Modified IEEE 33-bus system.

The topology of the IEEE 33-bus system is shown in Fig. 7. There are two 500 kW PVs installed at node 12 and 28, and 2 500 kW micro turbines located at nodes 18 and 33. The DGs are under the same level of the aggregator, this work assumes that residential loads are the main source of flexible load, behind-the-meter solar and DERs are not included in the flexible load. In this case, the data were collected from the optimization result of the simulation described in Section 2. There are four sets of 7-day simulation data from four cases with different flexible load ratios. Two sets of data were used for training the LSTM neural network, and another two sets were used for testing. To prevent overfitting, 20% of the training data were used for validation. The input features are summarized in Table 1. The correlation map based on the inputs data and DLMP is shown in Fig. 8. In general, we can claim that the DLMP is mostly positive related with the base load, PV generation, outdoor temperature, and LMP, while negatively related with the voltage distribution in the baseline case. For the parameter configurations of LSTM, grid search was applied for batch size and learning rate selection. The mini batch size was set to 64, the learning rate was set to 0.001, the LSTM had 128 hidden units, and the number of fully connected layers was 3.

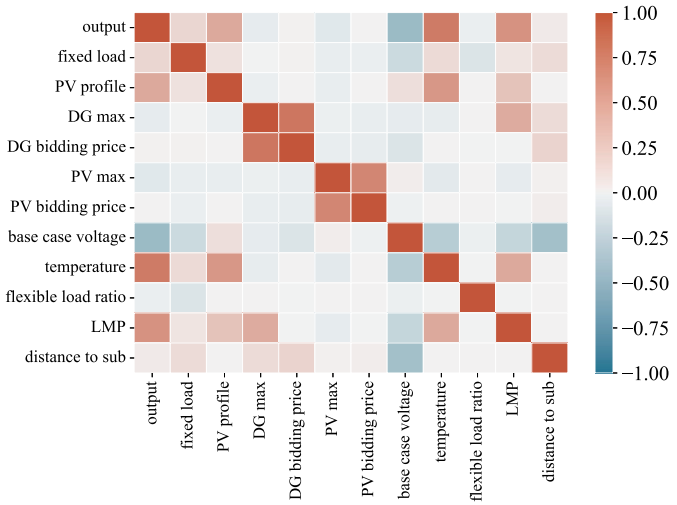


Fig. 8. Correlation Map of inputs and DLMP.

4.2. LSTM testing results under different system load levels

The comparison of DLMP generated from the LSTM and original optimization results is shown in Fig. 9. Each sub-figure shows the DLMP for a different day during a 24 h time period. The DLMPs of different days have similar durations of peak hours with different peak values. The results show that the LSTM units can generate DLMPs for different days that are very close to the OPF results from the optimization solution.

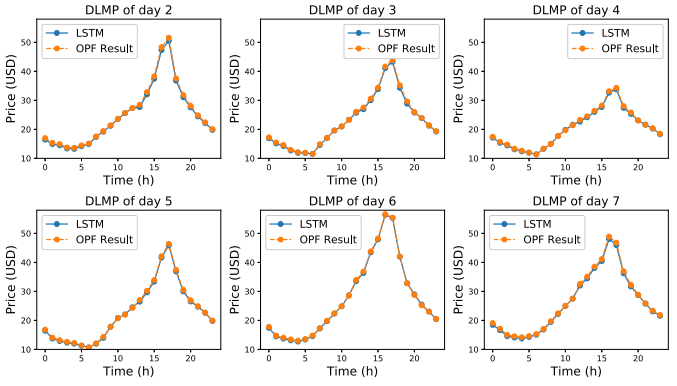


Fig. 9. DLMP of different days generated from the LSTM model compared with the OPF.

The profile of the difference between the DLMP generated from the LSTM and optimization results is shown in Figs. 10 and 11. These profiles show the prediction error of the LSTM model across all 33 nodes (buses) for 24 h of two test cases with PV generation and flexible load ratios of 20% and 30%, respectively. All the errors are within 3%, which demonstrates the accuracy of the proposed data-driven method. The results also illustrate that the proposed method generalized well on cases with different flexible load ratios.

4.3. Robustness testing results against uncertainty

As explained in Section 3.2, a rolling horizon updating method was proposed to enhance the robustness of the learning-based algorithm, where the inputs of the LSTM model are recursively updated at the beginning of every k hours. Many inputs features can be affected by uncertainties such as weather and different PV output profiles as shown in Fig. 12. To represent the uncertainties of different scenarios, random

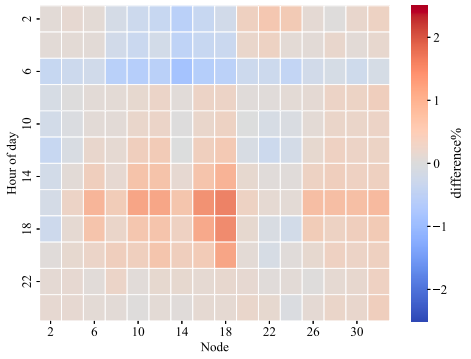


Fig. 10. DLMP difference between LSTM and optimization result (20% flexible load).

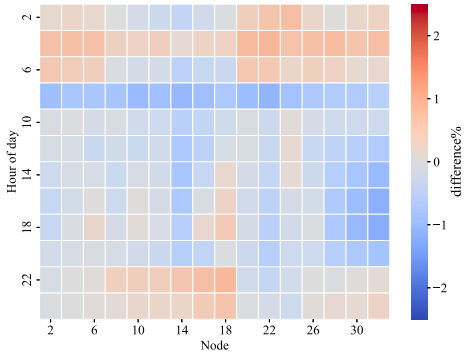


Fig. 11. DLMP difference between LSTM and optimization result (30% flexible load).

prediction noise was added to several inputs such as fixed load, base voltage, ambient temperature, and LMP.

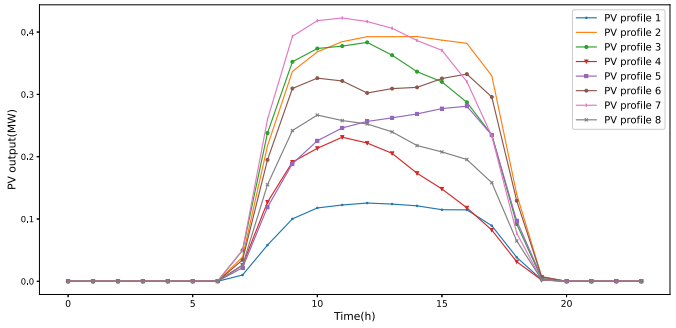


Fig. 12. Different PV profiles.

To compare the performance with and without sliding window, Monte Carlo simulations with random prediction noise were run to input features. The prediction regions of the DLMP by the LSTM model with different rolling horizon window sizes are shown in Figs. 13 and 14; the maximum applied noise on each input was 5% and 15%, respectively. The blue lines are the optimal DLMP values from the OPF solution. For different window sizes k, the DLMP prediction region of different cases were formed by collecting the maximum and minimum value across different trials at each hour. The results were also compared with the method without a sliding window. The effect of different levels of uncertainty on the DLMP is summarized in Table 2. The mean absolute percentage error introduced by the uncertainty can be significantly reduced by using a rolling horizon method. As expected, the moving window with shorter window length (higher update frequency) more effectively mitigated the effect of uncertainty on prediction errors.

Table 2
Effect of uncertainty with different sliding windows (mean absolute percentage error).

Maximum noise	3%	6%	9%	12%	15%
Without RH	2.6%	3.8%	5.0%	6.1%	7.2%
12 h RH	1.5%	2.0%	2.5%	3.0%	3.6%
8 h RH	1%	1.4%	1.7%	2.0%	2.4%
6 h RH	0.9%	1.1%	1.4%	1.6%	1.8%
4 h RH	0.7%	0.8%	1.0%	1.1%	1.3%

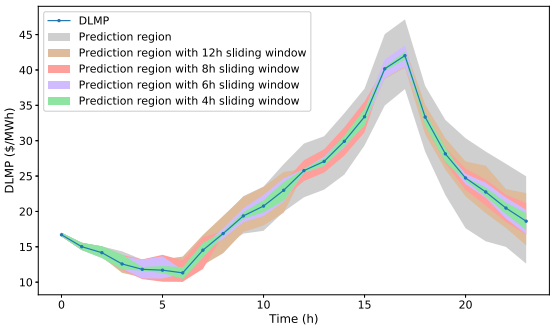


Fig. 13. Case 2: A simple 24 h prediction with different sliding window lengths with 5% maximum input noise.

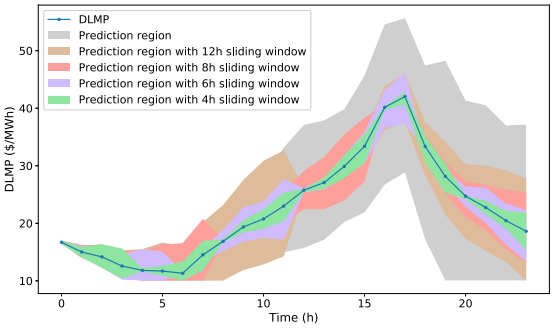


Fig. 14. Case 3: A simple 24 h prediction with different sliding window lengths with 15% maximum input noise.

The gray regions in Figs. 13 and 14 are the possible regions of DLMP values generated from the LSTM model without a sliding window. The prediction result without the sliding window has obvious deviation from the optimal DLMP value. The prediction region becomes wider as the magnitude of input noise increases. With a smaller k value, the predicted DLMP values are more stable and are closer to the optimal value. When k is 4, the DLMP values stay in a narrow band around the optimal value since the inputs are updated more frequently with better accuracy. The results demonstrate the effectiveness of the proposed rolling horizon method under the condition with uncertainty. The average DLMP error across different nodes with different sliding window sizes is shown in Fig. 15. Similarly, the error was maintained at lower levels with smaller k values.

4.4. Transfer learning on the IEEE 123-bus distribution system

The proposed method was further tested on a modified IEEE 123-bus system. The network topology of the system is shown in Fig. 16. Eight HVAC aggregators (H1–H8) and 8 electric vehicle aggregators (E1–E8) were located at different nodes (buses) in the system as shown in Fig. 16. Five simulation cases with different flexible load ratios are described in Table 3.

The pretrained network using the IEEE 33-bus model was trained on the IEEE 123-bus system with limited data as shown in Fig. 5. The transfer learning model froze the weights of the other layers and only trained the weights of the last layer, which are related to the contribution of the spatial (network topology) features. The network was tested on different simulation cases with PV generation and flexible load ratios of 0%, 20%, and 40%. The difference of the DLMP generated by the LSTM model compared with the optimization results is shown in Figs. 17–20 with the various respective ratios. The distribution of DLMP difference of different cases is demonstrated in Fig. 21. Based on the simulation results in these different cases, the difference between the DLMP generated by the neural network and the optimization results is less than 2% most of the time. The results demonstrate the robustness of the proposed neural network model. The mean absolute error of the DLMP is as low as 0.15 $/MWh.

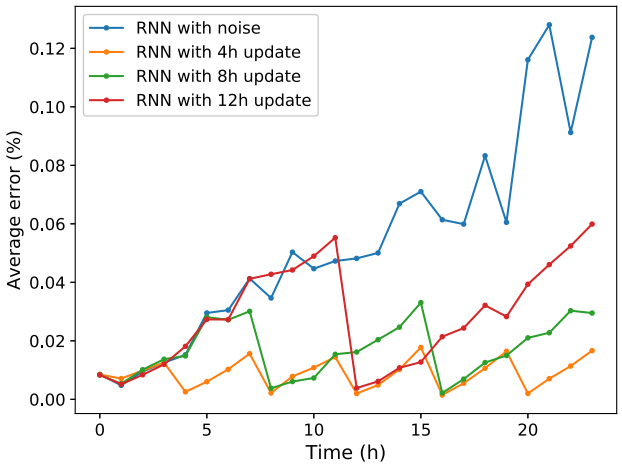


Fig. 15. The average DLMP error across different nodes with different sliding window sizes k.

Table 3
Different cases.

Case	System flexible load ratio	Load composition		
		Flex	Inflex	Other
0	0%	None	H1–H8, E1–E8	Fixed load
1	10%	H1–H2, E1–E2	H3–H8, E3–E8	Fixed load
2	20%	H1–H4, E1–E4	H5–H8, E5–E8	Fixed load
3	30%	H1–H6, E1–E6	H7–H8, E7–E8	Fixed load
4	40%	H1–H8, E1–E8	None	Fixed load

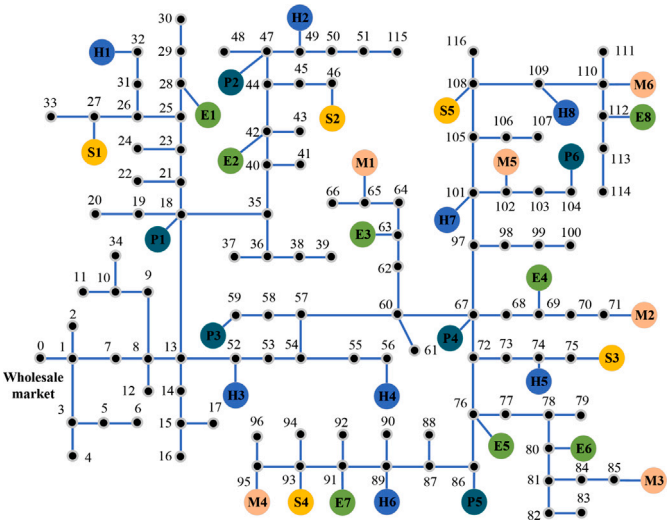


Fig. 16. Modified IEEE 123-bus system.

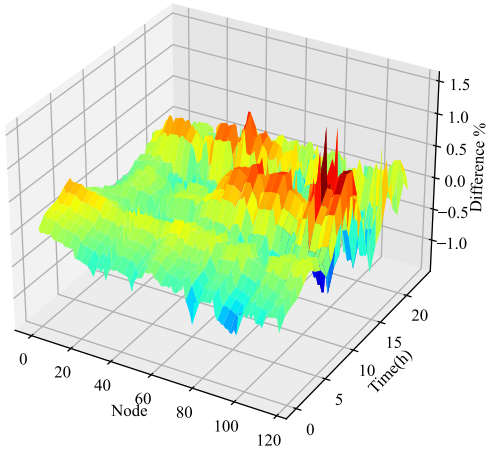


Fig. 17. 3D profile of the DLMP difference between the LSTM and optimization results on IEEE 123-bus distribution system through transfer learning (case with 0% flexible ratio).

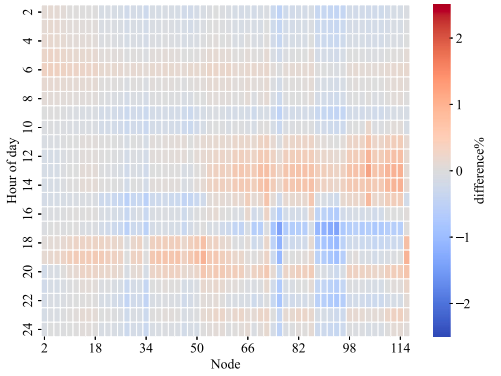


Fig. 18. DLMP difference between the LSTM and optimization results on IEEE 123-bus distribution system through transfer learning, MAE: 0.15$/MWh (case with 0% flexible ratio).

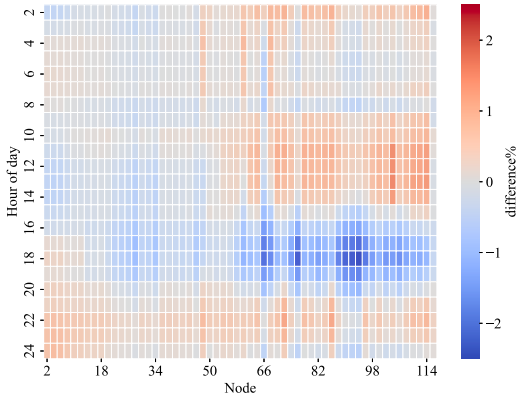


Fig. 19. DLMP difference between the LSTM and optimization results on IEEE 123-bus distribution system through transfer learning, MAE: 0.2$/MWh (case with 20% flexible ratio).

Figs. 22 and 23 show the detailed DLMP results comparison between the learning-based approach and the optimization-based solution for the IEEE 123-bus system under different system flexible load ratios at 10 a.m. and 7 p.m., respectively. The figures show that the DLMP is higher at 7 p.m. for all three cases, and the learning-based approach can predict the different DLMPs at different times of day. The overall increasing trend of the DLMP across different nodes is well captured

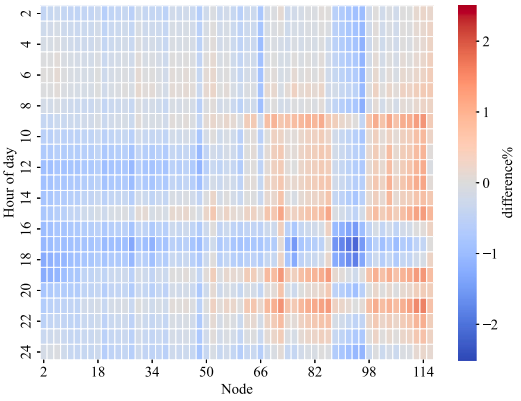


Fig. 20. DLMP difference between the LSTM and optimization results on IEEE 123-bus distribution system through transfer learning, MAE: 0.17$/MWh (case with 40% flexible ratio).

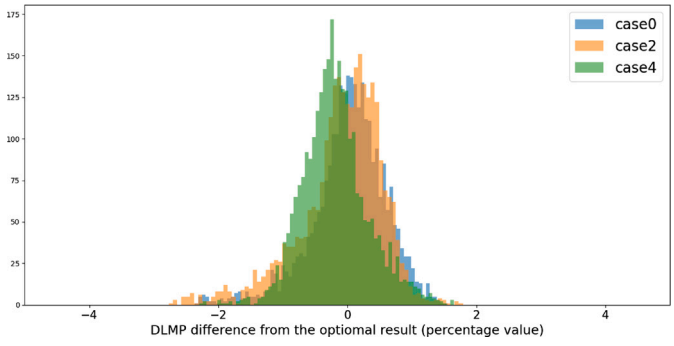


Fig. 21. The DLMP percentage error distribution of different cases.

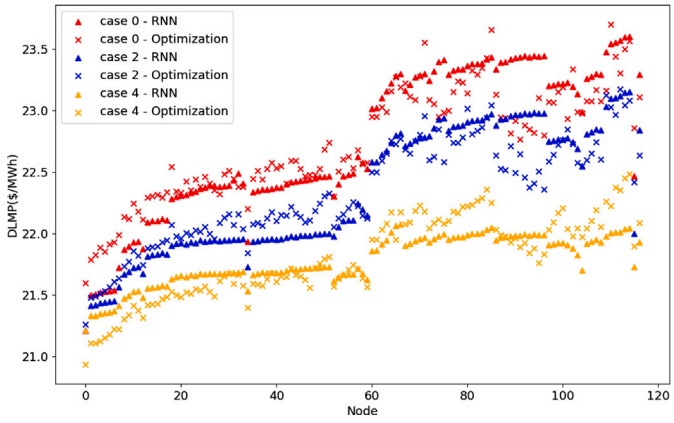


Fig. 22. DLMP of different cases at different nodes at 10 a.m. (RNN represents the proposed transfer learning solution).

by the proposed transfer learning solution. Moreover, with the system flexible load ratio as one of the input features, the proposed neural network can generate DLMPs at different price levels for cases with different flexibility. The RMSEs for 10 a.m., 7 a.m. and the whole 24 h are summarized in Table 4.

To further evaluate the proposed learning model, the results of different training scenarios were compared. Figs. 24–26 show the distribution of the error from different training options for the DLMP on the IEEE 123-bus system. The results from Fig. 24 were generated by directly using the model trained on the 33-bus system. This model tended

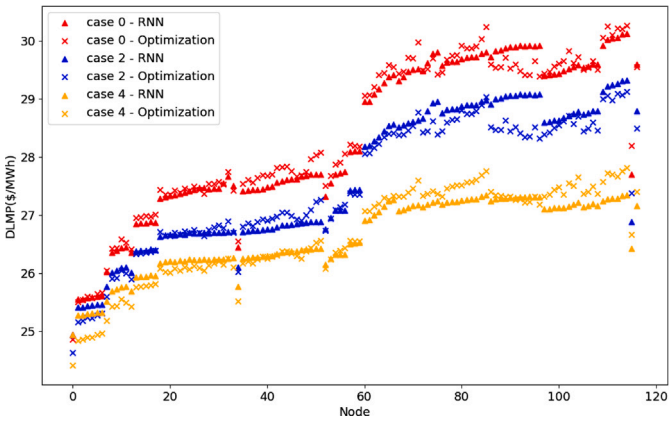


Fig. 23. DLMP of different cases at different nodes at 7 p.m. (RNN represents the proposed transfer learning solution).

Table 4
RMSEs ($/MWh) of different cases.

Case	10 a.m.	7 p.m.	Daily average
Case 0	0.197	0.154	0.177
Case 2	0.212	0.159	0.182
Case 4	0.184	0.169	0.153

Table 5
Comparison of prediction error for different training models (with and without transfer learning).

Training model	RMSE ($/MWh)	R^2
Model trained from the source system	3.253	0.888
Model trained on the target system with limited data	0.246	0.993
Transfer learning on the target system	0.164	0.997

to underestimate the DLMP values; larger deviation can be observed as the DLMP value increases. This underestimation could be caused by the system topology and scale difference from the source system and target system. In Fig. 25, the model was trained with limited data from the target system. Although it shows moderate error overall, a few outliers exist. The performance of the model is limited by insufficient data. Fig. 24 shows the results from the transfer learning model that trained the weights of the last layer with limited data and kept the weights of the other layer from the pretrained model on the source system. This model generated the best results among the three training options. To better understand the performance improvement across these three figures, we added a violin plot Fig. 27, which captures statistics of the prediction errors. As shown in Fig. 27, both the mean and standard deviation are in a descent trend from the left to the right. And the green one (using transfer learning technique) delivers notable accuracy gains, which demonstrates the efficacy of the proposed method. As a summary, detailed numerical numbers of the errors for each training scenario is given in Table 5. It is obvious that, our proposed transfer learning solution has achieved up to 95% RMSE reduction in predicting the DLMP by transferring knowledge learned from another network topology, even though there is only slight improvement on the R^2 when comparing transfer learning and direct training on the target system, the RMSE is reduced significantly.

Finally, we want to illustrate the improvement in learning progress through the proposed transfer learning technique. The convergence process plot in Fig. 28 reveals three key findings: (1) the direct training comes with much larger oscillation (red shaded area) than the proposed transfer learning approach; (2) the direct training and transfer learning approaches can generate decent results given the gap between two methods vanishes at the end, and (3) the transfer learning method delivers more than 5 times acceleration during the training process.

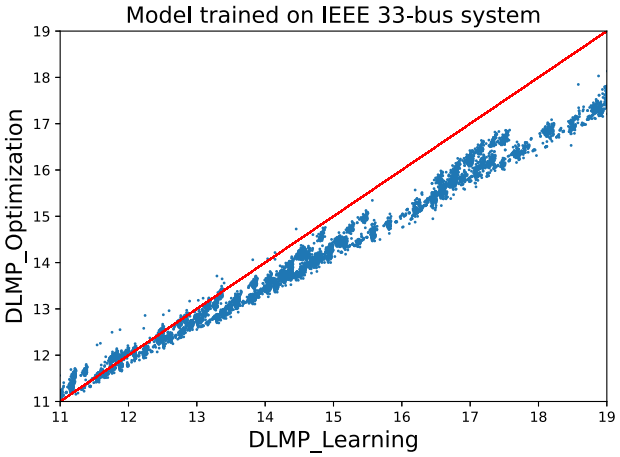


Fig. 24. Model trained from the source system.

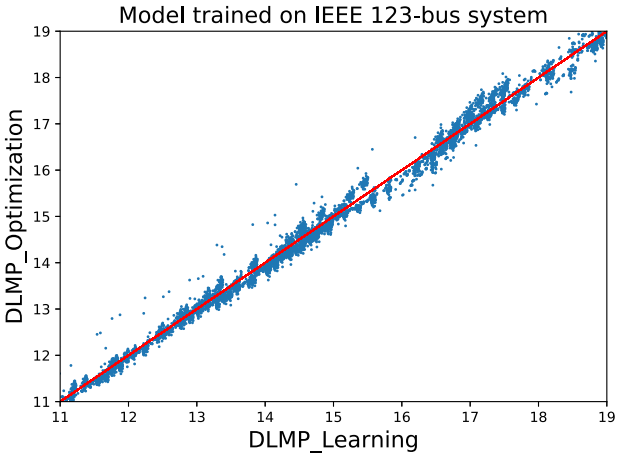


Fig. 25. Model trained on the target system with limited data.

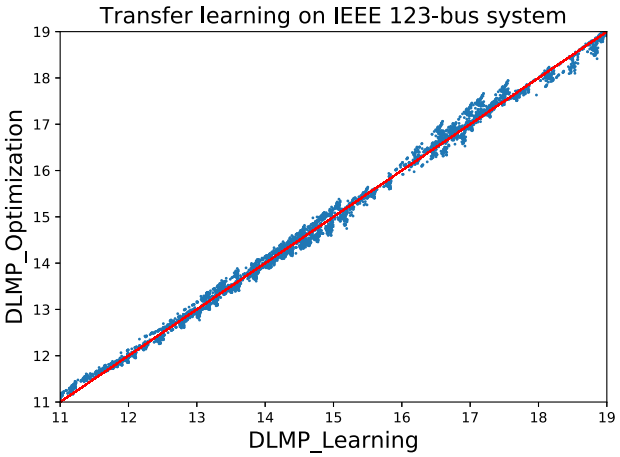


Fig. 26. Transfer learning on the target system.

Compared with direct training on the target system with limited data, the transfer learning method provides a warm-start point for the training, which significantly enhances the convergence speed and prediction accuracy.

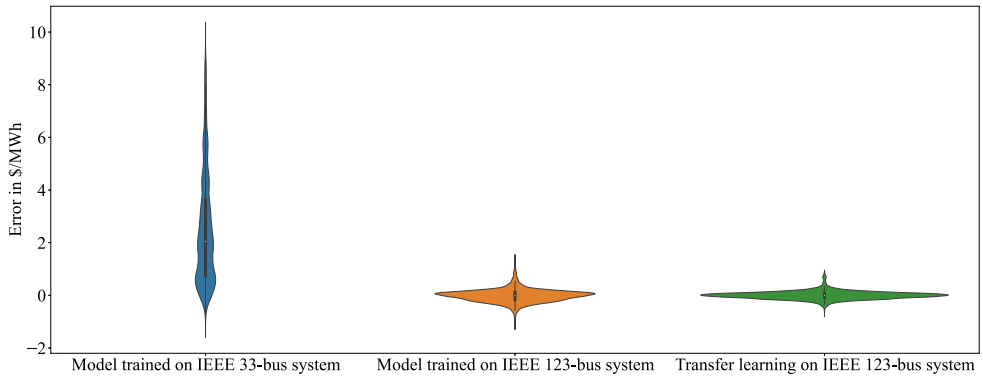


Fig. 27. Violin plot of the errors from different training models.

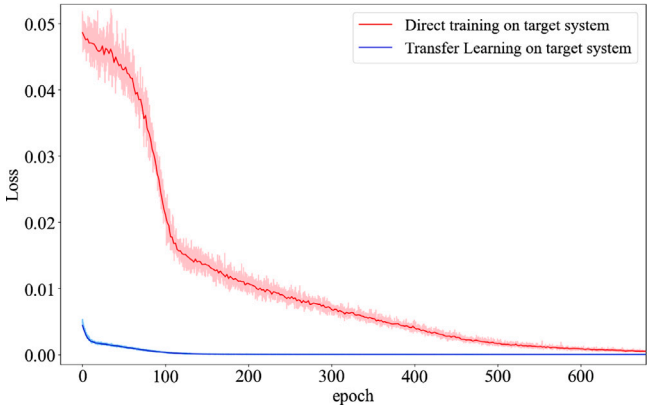


Fig. 28. Comparison of training losses for direct training and transfer learning methods.

Table 6
Computation and training time.

Case	Computation time (s)		Training time (s)	
	Optimization	Neural network	Direct learning	Transfer learning
33-bus system	20	0.124	1390	
123-bus system	60	0.272	3630	640

4.5. Computation time

As shown in Table 6, compared with the benchmark optimization-based method, the data-driven algorithm can significantly reduce the computation time. More than 99% of computation time was reduced for both the IEEE 33-bus and 123-bus systems. Additionally, the data-driven method is more computational efficient in larger systems since the computation time of the optimization-based method will increase exponentially with an increasing number of nodes. The training time for the 33-bus and 123-bus systems are 23 min and 1 h, respectively. With transfer learning, training time can be reduced 82% for the 123-bus system.

5. Conclusions

To address the scalability and practicability challenges in deploying the DLMP in distribution network, we have proposed a learning-based approach that realizes a data-driven distribution market mechanism considering load and PV uncertainties. The framework features the seamless integration of generally available utility data across the distribution network, which makes it possible to deploy DLMP in real-world distribution systems. The proposed approach combines LSTM and transfer learning to develop deep neural network that can capture the temporal dependencies of the time series inputs, as well as the spatial patterns of the distribution network. This has led to dramatic deployment complexity reduction when the real-world network models are not readily available for solving the corresponding optimal power flow problem. Testing results on IEEE 33-bus and 123-bus systems show that the proposed approach can generate a comparable DLMP (less than 5% deviation) against the optimization solutions with up to 99% reduction in computation time. In addition, several other adaptive and robust techniques were integrated to make the data-driven solution ready for more challenging scenarios, which include a sliding window method to mitigate the uncertainty/noise in the prediction profiles, and transfer learning to enable a quick deployment to other distribution networks.

This work points to a number of directions warranting follow-up investigations. It is interesting to extend the current transfer learning technique to distributed multi-agent learning framework, which can further reduce the communication and computation overhead. On the other hand, the performance of learning-based solution can be enhanced by exploring different DNN structures with various combinations of distribution network topology, flexible load ratio, renewable penetration ratio, etc. Last, particularization of the current design to real feeders provides abundance of cross-disciplinary research opportunities, for example, state estimation and feature selection under different system conditions.

CRediT authorship contribution statement

Boming Liu: Conceptualization, Methodology, Software, Result analysis, Writing – original draft. **Jin Dong:** Visualization, Concept development, Result analysis, Writing – original draft. **Jianming Lian:** Concept development, Writing – review & editing. **Teja Kuruganti:** Project administration, Supervision, Writing – review & editing. **Xiaofei Wang:** Data curation, Result analysis, Writing – review & editing. **Fangxing Li:** Result analysis, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This material is based upon work supported by the US Department of Energy (DOE), Grid Modernization Laboratory Consortium (GMLC), DOE Energy Efficiency and Renewable Energy, Building Technology Office, and DOE Office of Electricity.

References

[1] Cain MB, O'neill RP, Castillo A, et al. History of optimal power flow and formulations. Vol. 1. Federal Energy Regulatory Commission; 2012, p. 1–36, Citeseer.

[2] Li F, Bo R. DCOPF-based LMP simulation: Algorithm, comparison with ACOPF, and sensitivity. IEEE Trans Power Syst 2007;22(4):1475–85.

[3] Sotkiewicz PM, Vignolo JM. Nodal pricing for distribution networks: Efficient pricing for efficiency enhancing DG. IEEE Trans Power Syst 2006;21(2):1013–4.

[4] Yuan H, Li F, Wei Y, Zhu J. Novel linearized power flow and linearized OPF models for active distribution networks with application in distribution LMP. IEEE Trans Smart Grid 2016;9(1):438–48.

[5] Renani YK, Ehsan M, Shahidehpour M. Optimal transactive market operations with distribution system operators. IEEE Trans Smart Grid 2017;9(6):6692–701.

[6] Pediaditis P, Ziras C, Hu J, You S, Hatziargyriou N. Decentralized DLMPs with synergetic resource optimization and convergence acceleration. Electr Power Syst Res 2020;187:106467.

[7] Rezvanfar R, Hagh MT, Zare K. Power-based distribution locational marginal pricing under high-penetration of distributed energy resources. Int J Electr Power Energy Syst 2020;123:106303.

[8] Moghadam AZ, Javidi MH. Designing a two-stage transactive energy system for future distribution networks in the presence of prosumers' P2P transactions. Electr Power Syst Res 2022;211:108202.

[9] He Y, Chen Q, Yang J, Cai Y, Wang X. A multi-block ADMM based approach for distribution market clearing with distribution locational marginal price. Int J Electr Power Energy Syst 2021;128:106635.

[10] Wang B, Tang N, Bo R, Li F. Three-phase DLMP model based on linearized power flow for distribution with application to DER benefit studies. Int J Electr Power Energy Syst 2021;130:106884.

[11] Castillo A, O'Neill RP. Survey of approaches to solving the ACOPF. Vol. 11. Tech. Rep, Federal Energy Regulatory Commission; 2013.

[12] Wang H, Murillo-Sanchez CE, Zimmerman RD, Thomas RJ. On computational issues of market-based optimal power flow. IEEE Trans Power Syst 2007;22(3):1185–93. http://dx.doi.org/10.1109/TPWRS.2007.901301.

[13] Liu B, Akcakaya M, McDermott TE. Reduced order model of transactive bidding loads. IEEE Trans Smart Grid 2021.

[14] Livas-García A, Tzuc OM, May EC, Tariq R, Torres MJ, Bassam A. Forecasting of locational marginal price components with artificial intelligence and sensitivity analysis: A study under tropical weather and renewable power for the Mexican Southeast. Electr Power Syst Res 2022;206:107793.

[15] Pan X, Zhao T, Chen M. DeepOPF: Deep neural network for DC optimal power flow. In: 2019 IEEE international conference on communications, control, and computing technologies for smart grids. IEEE; 2019, p. 1–6.

[16] Huang W, Pan X, Chen M, Low SH. DeepOPF-V: Solving AC-OPF problems efficiently. 2021, arXiv preprint arXiv:2103.11793.

[17] Chatzos M, Mak TW, Van Hentenryck P. Spatial network decomposition for fast and scalable AC-OPF learning. 2021, arXiv preprint arXiv:2101.06768.

[18] Owerko D, Gama F, Ribeiro A. Optimal power flow using graph neural networks. In: ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing. IEEE; 2020, p. 5930–4.

[19] Liu S, Wu C, Zhu H. Graph neural networks for learning real-time prices in electricity market. 2021, arXiv preprint arXiv:2106.10529.

[20] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2009;22(10):1345–59.

[21] Shi Z, Yao W, Li Z, Zeng L, Zhao Y, Zhang R, et al. Artificial intelligence techniques for stability analysis and control in smart grids: Methodologies, applications, challenges and future directions. Appl Energy 2020;278:115733.

[22] Huang W, Zhang X, Zheng W. Resilient power network structure for stable operation of energy systems: A transfer learning approach. Appl Energy 2021;296:117065.

[23] Yang H, Schell KR. Real-time electricity price forecasting of wind farms with deep neural network transfer learning and hybrid datasets. Appl Energy 2021;299:117242.

[24] Ibrahim MS, Dong W, Yang Q. Machine learning driven smart electric power systems: Current trends and new perspectives. Appl Energy 2020;272:115237.

[25] Theocharides S, Makrides G, Livera A, Theristis M, Kaimakis P, Georghiou GE. Day-ahead photovoltaic power production forecasting methodology based on machine learning and statistical post-processing. Appl Energy 2020;268:115023.

[26] Perera A, Wickramasinghe P, Nik VM, Scartezzini J-L. Machine learning methods to assist energy system optimization. Appl Energy 2019;243:191–205.

[27] Sun Y, Liu X, Ni Y, Sun Q. Generalized demand-side resource hierarchical control strategy based on multi-agent consensus of historical data online transfer. Electr Power Syst Res 2022;211:108166.

[28] Mohammadian M, Baker K, Fioretto F. Gradient-enhanced physics-informed neural networks for power systems operational support. 2022, arXiv preprint arXiv:2206.10579.

[29] Zhang X, Chen Y, Yu T, Yang B, Qu K, Mao S. Equilibrium-inspired multiagent optimizer with extreme transfer learning for decentralized optimal carbon-energy combined-flow of large-scale power systems. Appl Energy 2017;189:157–76.

[30] Fan L, Zhang J, He Y, Liu Y, Hu T, Zhang H. Optimal scheduling of microgrid based on deep deterministic policy gradient and transfer learning. Energies 2021;14(3):584.

[31] Ren C, Xu Y. Transfer learning-based power system online dynamic security assessment: using one model to assess many unlearned faults. IEEE Trans Power Syst 2019;35(1):821–4.

[32] Wang X, Li F, Dong J, Olama MM, Zhang Q, Shi Q, et al. Tri-level scheduling model considering residential demand flexibility of aggregated HVACs and EVs under distribution LMP. IEEE Trans Smart Grid 2021;12(5):3990–4002.

[33] Laptev N, Yosinski J, Li LE, Smyl S. Time-series extreme event forecasting with neural networks at Uber. In: International conference on machine learning. Vol. 34. 2017, p. 1–5.

[34] Chen J, Zeng G-Q, Zhou W, Du W, Lu K-D. Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. Energy Convers Manage 2018;165:681–95. http://dx.doi.org/10.1016/j.enconman.2018.03.098, URL http://www.sciencedirect.com/science/article/pii/S0196890418303261.

[35] Cheng L, Yu T. A new generation of AI: A review and perspective on machine learning technologies applied to smart energy and electric power systems. Int J Energy Res 2019;43(6):1928–73.

[36] Cai L, Gu J, Ma J, Jin Z. Probabilistic wind power forecasting approach via instance-based transfer learning embedded gradient boosting decision trees. Energies 2019;12(1):159.

[37] Weng Y, Liao Y, Rajagopal R. Distributed energy resources topology identification via graphical modeling. IEEE Trans Power Syst 2016;32(4):2682–94.