An Optimized Distribution Model for Energy System in Virtual Power Plants Integrating Electric Vehicles Based on TD3 and DQN

Wei Hu, Shuo Wang, and Puliang Du

***Abstract*—To enhance the deployment capability and low-carbon degree of virtual power plants (VPPs), a novel optimized scheduling model is proposed in this paper for a multi-energy VPP. To explore the distribution potential of the VPP and bolster its multi-energy complementarity, an architecture integrated with electric vehicle (EV) charging stations is introduced, and a battery health degradation mechanism is constructed. To address the uncertainty exhibited by EV behaviors, a feature extraction method based on deep Q-network and maximum relevance-minimum redundancy (mRMR) is then proposed. This method optimizes the applicability of mRMR in large datasets, thereby improving the accuracy of charge behavior prediction. Next, to achieve a complex optimization dispatch, a twin delayed deep deterministic policy gradient algorithm is employed. The twin Q-value truncation mechanism and smooth regularization effectively suppress the issue of policy overestimation biases. Furthermore, to validate the performance of the proposed model and algorithm, four different cases are designed, and the scheduling effects achieved for EVs are compared with those of the traditional battery energy storage system framework. The simulation results show that the proposed model significantly reduces both the operational cost and carbon emission level while slowing the battery health degradation process.**

***Index Terms*—Energy distribution, twin deterministic policy gradient, deep Q-network, electric vehicles, virtual power plant.**

I. Introduction

With the accelerated transition toward a low-carbon energy structure, the installed capacity of renewable energy sources, represented by wind and solar power, continues to surge [1]. However, the high volatility of renewable energy substantially increases the challenge of maintaining dynamic balance in power systems, particularly under high-penetration grid scenarios [2]. Consequently, traditional operation models face severe challenges [3]. Therefore, achieving dynamic coordination through flexible resource aggregation technologies has become a core issue [4].

As a pivotal dispatch technology in new electricity systems, virtual power plants (VPPs) integrate distributed energy resources (DERs) into a unified virtual entity that can be centrally controlled through an information-physical fusion system [5]. VPPs employ distributed optimization algorithms to decouple renewable energy output fluctuations and load demand uncertainties, thereby effectively addressing the power balance risk [6]. Moreover, VPPs directly reduce the dependence of the system on fossil fuels by enhancing its spatiotemporal matching ability [7]. However, with the increasing integration of DERs, the bidirectional power flow characteristics within VPPs gradually intensify [8]. The traditional VPP resource framework struggles to cope with the coupling effects of high-frequency power fluctuations and indeterminacy across multiple time scales [9].

To address these challenges, reference [10] proposes the integration of a power-to-gas (P2G) device to improve the energy transfer ability of their system. However, P2G is limited by a multistage process, resulting in delayed responses and relatively low conversion efficiency. In addition, reference [11] suggests incorporating pumped storage into a VPP to increase its dispatch frequency. However, the construction process is heavily dependent on specific geographical conditions and imposes energy density limits.

To address the above issues, a VPP architecture that incorporates collaboration between an electric vehicle (EV) charging station and the P2G model is constructed herein. This VPP resource framework leverages the short-term power regulation competence of EVs to compensate for the response limitations of P2G and other methods while optimizing their economy.

The existing studies on the model of EVs in a VPP often use linear state-of-charge (SOC) constraints to

Received: January 10, 2025
Accepted: September 20, 2025
Published Online: November 1, 2025
Wei Hu (corresponding author), Shuo Wang, and Puliang Du are with the School of Economics and Management, Shanghai University of Electric Power, Shanghai 201306, China (e-mail: hwshiep@shiep.edu.cn; wangshuobh@163.com; Dupuliang@163.com).
DOI: 10.23919/PCMP.2025.000017

represent EV changes. Reference [12] proposes an EV scheduling model based on linear SOC constraints, where charge and discharge power thresholds are set to participate in power balancing [13]. However, this model neglects the degradation of the state of health (SOH) of the examined battery. Reference [14] introduces a SOH degradation model based on the cycle count, but it overlooks the impacts of the charge and discharge rates. Therefore, a bidirectional mathematical charge and discharge model that accounts for both SOH degradation and the charge rate is established, achieving a balance between battery degradation and deployment economy.

The spatiotemporal heterogeneity of EV behaviors and the randomness of user responses require the VPP to possess strong behavioral analysis capabilities [15]. Reference [16] generates EV demand scenarios via Monte Carlo simulations, in which a random sample is used to characterize the inherent uncertainty. However, when addressing high-dimensional features, this method faces significant computational complexity [17]. Reference [18] establishes a behavioral random probability model and used sample- and cluster-based methods to generate deterministic representative cases. However, this approach struggles to capture the nonlinear coupling effects between features. Reference [19] proposes a deterministic method by constructing an analytical model of behavioral parameter variables and corresponding constraints. However, this approach is unable to search for the correlations between latent variables. Additionally, the prediction technique has been widely applied in the field of behavior analysis. Reference [20] employs an autoregressive integrated moving average model (ARIMA) to forecast uncertain behaviors, while reference [21] proposes a charge demand forecasting model based on the long short-term memory (LSTM), which accurately identifies behavioral demand through time series feature extraction. However, these methods rely on high-quality prior data. To resolve these problems, an uncertainty analysis framework based on a deep Q-network with maximum relevance-minimum redundancy (DQN-mRMR) and a genetic algorithm-temporal convolutional network-LSTM (GA-TCN-LSTM) forecasting model is proposed in this paper [22]. By improving the quality of prior data, the algorithm is optimized to enhance its ability to quantitatively analyze the uncertainty.

The dispatch of a VPP depends not only on the accuracy of the underlying uncertainty model but also on the optimization performance achieved for complex decision-making tasks [23], [24]. Reference [25] proposes a scheduling method based on discrete fuzzy logic (DFL), where decisions are made using a predefined fuzzy rule base. However, the discretization feature of this method leads to a rough division scheme, making it difficult to accurately represent continuous features. Reference [26] adopts a mixed-integer linear programming (MILP) to construct an optimization model. However, its computational complexity increases exponentially with the number of resources. In [27], a model predictive control (MPC) framework is introduced, which attains enhanced dynamic adaptability through rolling horizon optimization. However, its performance is highly dependent on the accuracy of the utilized prediction model. Reference [28] proposes a model-free algorithm that is based on Q-learning, though its discrete action space design is not suitable for continuous optimization environments. Reference [29] further employs a deep Q-network (DQN) to achieve a continuous action output. However, its exploration strategy is relatively simple, leading to convergence to a local optimum. Reference [30] applies a deep deterministic policy gradient (DDPG) algorithm, which uses an actor-critic architecture to improve the continuity of the developed policy. However, its issue of overestimating the Q value leads to a policy bias. To address the aforementioned deficiencies, the twin delayed DDPG (TD3) algorithm can effectively mitigate the overestimation of the value function through the truncation of the twin Q-learning process and the policy delay update mechanism. Additionally, the algorithm uses smooth target policy regularization to increase its robustness.

Therefore, to achieve safe and accurate VPP deployments, an optimization framework that integrates EVs and P2G is constructed, and the TD3 algorithm is employed to realize a rational resource schedule. The proposed model addresses the limitations of neglecting battery health, an insufficient high-dimensional behavioral feature model and policy overestimation in the traditional strategies. The main contributions of this paper can be summarized as follows.

1) A VPP-EV dispatch model considering battery health degradation is proposed. In this model, EVs serve as bidirectional power units, where their battery health deteriorates with increasing numbers of cycles and energy changes. Additionally, P2G is incorporated into the VPP to absorb carbon emissions, further enhancing the resilience of the system to fluctuations. Moreover, this integration improves the low-carbon performance of the model.

2) To address the complex uncertainty exhibited by EVs, a DQN-mRMR model is designed for EV behavior analysis purposes. The heuristic mRMR metric is used as the reward function so that the DQN selects the most representative features. Based on the selected features, GA automatically chooses the optimal parameters for TCN and LSTM, and then evaluates the weights between TCN and LSTM, thereby enabling accurate future demand predictions for EVs.

3) The VPP resource framework proposed in this paper is an extremely complex model, making the solution algorithm prone to becoming stuck in local

optima because of the strict imposed constraints. Therefore, the TD3 algorithm is introduced and applied to the continuous optimization of the VPP. By finding appropriate hyperparameters, TD3 leverages twin Q truncation and a noise mechanism to achieve a balance between economic efficiency and low-carbon performance.

The remainder of this paper is organized as follows. Section II presents the VPP-EV framework and the mathematical models of interior interactions. Section III proposes a dispatch model that is mainly based on the DQN and TD3 algorithms. Section IV provides different case studies and a comparison among four algorithms, and finally, Section V presents the conclusions of this work.

II. Virtual Power Plant Model with EVs

The VPP framework proposed in this paper is shown in Fig. 1. It includes three flows, which are divided into power and gas components. Among them, the power grid, natural gas, P2G mechanism, microturbine (MT), EVs and battery have a bidirectional interaction stream with the VPP. Notably, the MT outputs electricity to the VPP, the VPP transmits CH_4 to the MT, and the interaction between P2G and the VPP is reversed. In addition, beyond fulfilling mandatory daily travel demands, EVs engage in supplemental charging to enable their participation in extra dispatch schedules. It is not necessary for the battery to meet the basic demand, but it also participates in an extra schedule.

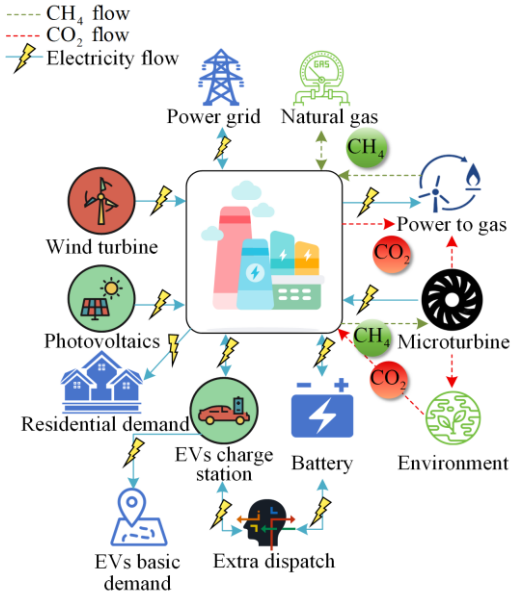


Fig. 1. Framework of the VPP with EVs.

It is assumed that the interactions between DERs, the power grid, and natural gas all occur through the VPP and that no direct relationship is present among them. Additionally, for the purpose of carbon minimization, an interaction is set up with the external carbon environment, where the VPP output includes the carbon emissions, and the input from the environment consists of the carbon emissions to be absorbed in the VPP. All device units in the VPP are standardized to specific power (MWh) and gas (kg) levels. The configuration in Fig. 1 serves as the main framework for constructing the mathematical model of the VPP examined in this paper.

A. Uncertainty Analysis Model for DER Outputs and Residential Demand

The uncertainty of the wind turbine (WT), photovoltaics (PV) outputs and residential load demand in the VPP introduces certain security risks. Therefore, probability density functions (PDFs) are calculated based on the Beta distribution and Weibull distribution for the historical light intensity and wind speed data, respectively. According to the corresponding PDF, four cases are generated from the samples. Then, the normal distribution parameters of the residential load are calculated through the historical data, and four load demand scenarios are generated from the samples. Notably, 365 days of 24-hour data are generated, yielding a total of 8760 points for training TD3. The four cases with 24-hour data generated are then used with the first 24 hours of the 8760-hour data for testing. The specific uncertainty analysis model is as follows.

The PDF of the wind speed obeying the Weibull distribution is as follows:

$$f(v)=\frac{z}{c}\left(\frac{v}{c}\right)^{z-1}\mathrm{e}^{-\left(\frac{v}{c}\right)^{z}} \tag{1}$$

where v denotes the wind speed; c denotes the characteristic parameter; and z denotes the shape parameter.

Based on the basis of the PDF of the wind speed, the z and c are calculated via maximum likelihood estimation (MLE), and the process is as follows:

$$F(z,c)=\prod_{i=1}^{k}\left(\frac{z}{c}\left(\frac{v_i}{c}\right)^{z-1}\mathrm{e}^{-(v_i/c)^{z}}\right) \tag{2}$$

where $F(z,c)$ denotes the MLE function of z and c; k denotes the number of data points; while v_i denotes the speed of ith wind sample.

The logarithm of $F(z,c)$ is as follows:

$$\ln F(z,c)=k\ln z-kz\ln c+(z-1)\sum_{i=1}^{k}\ln v_i-\sum_{i=1}^{k}\left(\frac{v_i}{c}\right)^{z} \tag{3}$$

The respective partial derivatives with respect to the MLE are shown as (4). By setting these partial derivatives equal to zero, parameters can be calculated.

$$\begin{cases}\dfrac{\partial \ln F(z,c)}{\partial c}=-\dfrac{kz}{c}+\dfrac{z}{c^{z+1}}\displaystyle\sum_{i=1}^{k}v_i^{z}\\[2ex] \dfrac{\partial \ln F(z,c)}{\partial z}=\dfrac{k}{z}-k\ln\hat{c}+\displaystyle\sum_{i=1}^{k}\ln v_i-\left(\frac{v_i}{\hat{c}}\right)^{z}\ln\left(\frac{v_i}{\hat{c}}\right)\end{cases} \tag{4}$$

where $\hat{c}$ represents the analytical solution of c.

Finally, z and c are calculated via (4), and the data (subject to the Weibull distribution) are generated.

With the wind speed data, the resulting wind power output model is as follows:

$$G_t^{\mathrm{WT}}(v)=\begin{cases}G_{\mathrm{rated}}^{\mathrm{WT}}, & 25\leqslant v\leqslant 35\\ G_{\mathrm{rated}}^{\mathrm{WT}}\left(\dfrac{v-3}{v_{\mathrm{rated}}-3}\right)^3, & 3\leqslant v\leqslant 25\\ 0, & \text{otherwise}\end{cases} \tag{5}$$

where G_t^{WT} denotes the output of the WT at time t; $G_{\mathrm{rated}}^{\mathrm{WT}}$ denotes the rated power of the WT; and v_{rated} denotes the rated wind speed.

The PDF of the radiation intensity obeying a beta distribution is as follows:

$$f(\theta_{\mathrm{PV}})=\frac{(\theta_{\mathrm{PV}})^{\alpha_{\mathrm{PV}}-1}(1-\theta_{\mathrm{PV}})^{\beta_{\mathrm{PV}}-1}}{\int_0^1(\theta_{\mathrm{PV}})^{\alpha_{\mathrm{PV}}-1}(1-\theta_{\mathrm{PV}})^{\beta_{\mathrm{PV}}-1}\mathrm{d}(\theta_{\mathrm{PV}})} \tag{6}$$

$$\begin{cases}\alpha_{\mathrm{PV}}=\dfrac{\mu_{\mathrm{PV}}\beta_{\mathrm{PV}}}{1-\mu_{\mathrm{PV}}}\\ \beta_{\mathrm{PV}}=(1-\mu_{\mathrm{PV}})\left(\dfrac{\mu_{\mathrm{PV}}(1-\mu_{\mathrm{PV}})}{\sigma_{\mathrm{PV}}^2}\right)\end{cases} \tag{7}$$

where θ_{PV} denotes the solar radiation; μ_{PV} denotes the mean data value per hour for 365 days; σ_{PV}^2 denotes the variance per hour of data for 365 days; while α_{PV} and β_{PV} denote the shape parameters of Beta distribution.

After calculating α_{PV} and β_{PV}, the radiation intensities at 24 hours are sampled from the historical hourly data. The PV outputs are then acquired as follows:

$$G_t^{\mathrm{PV}}=\theta_{\mathrm{PV}}\times S_{\mathrm{PV}}\times N_{\mathrm{PV}}\times\eta_{\mathrm{PV}} \tag{8}$$

where G_t^{PV} is the PV output at time t; while S_{PV}, N_{PV} and η_{PV} denote the square, amount, and conversion efficiency of the PV power, respectively.

In addition, the residential demand is generated by a normal distribution as follows:

$$f(r_{\mathrm{Re}})=\frac{1}{\sigma_{\mathrm{Re}}\sqrt{2\pi}}\times\mathrm{e}^{-\left(\frac{r_{\mathrm{Re}}-\mu_{\mathrm{Re}}}{2\sigma_{\mathrm{Re}}^2}\right)^2} \tag{9}$$

where r_{Re} denotes the residential demand; while σ_{Re}^2 and μ_{Re} denote the variance and mean values of the actual residential demand for every hour in the whole daily data, respectively.

This method calculates σ_{Re}^2 and μ_{Re} for every hour of each day. Finally, the hourly demand in each case is created by sampling from historical data obeying a normal distribution.

B. MT Model

MT is a power generation device that uses $\mathrm{CH_4}$ as a fuel source and generates electric power by converting chemical energy into mechanical energy. It produces $\mathrm{CH_4}$ from natural gas and outputs power with a corresponding level of carbon emissions. The details of the output model for the MT are as follows:

$$G_t^{\mathrm{MT}}=\eta_{\mathrm{MT}}\times\kappa_{\mathrm{CH_4}}\times V_{\mathrm{CH_4},t}^{\mathrm{MT}} \tag{10}$$

$$V_{\mathrm{CO_2},t}^{\mathrm{MT}}=G_t^{\mathrm{MT}}\times\eta_{\mathrm{CO_2}}^{\mathrm{MT}} \tag{11}$$

where G_t^{MT} denotes the output of the MT at time t; η_{MT} denotes the energy conversion rate; $\kappa_{\mathrm{CH_4}}$ denotes the calorific value of $\mathrm{CH_4}$; $V_{\mathrm{CH_4},t}^{\mathrm{MT}}$ denotes the $\mathrm{CH_4}$ consumption volume at time t; $V_{\mathrm{CO_2},t}^{\mathrm{MT}}$ denotes the total amount of carbon emissions produced by the MT at time t; and $\eta_{\mathrm{CO_2}}^{\mathrm{MT}}$ denotes the $\mathrm{CO_2}$ emissions per unit of power output from the MT.

To ensure the stability of the MT, G_t^{MT} satisfies the following constraints:

$$G_{\mathrm{MT,min}}\leqslant G_t^{\mathrm{MT}}\leqslant G_{\mathrm{MT,max}} \tag{12}$$

$$\left|G_t^{\mathrm{MT}}-G_{t-1}^{\mathrm{MT}}\right|\leqslant\Delta G_{\mathrm{ramp}}^{\mathrm{MT}} \tag{13}$$

where $G_{\mathrm{MT,max}}$ and $G_{\mathrm{MT,min}}$ denote the maximum and minimum output power levels of the MT, respectively; G_{t-1}^{MT} denotes the output of the MT at time $t-1$; and $\Delta G_{\mathrm{ramp}}^{\mathrm{MT}}$ denotes the maximum increase or decrease in the power output per hour.

C. P2G Model

The P2G model is mainly adopted in the form of power to gas. The main roles of P2G are to absorb the carbon emissions generated by the MT and to transfer the generated $\mathrm{CH_4}$ to the MT or sell it as natural gas to reduce the cost of the carbon emissions and provide increased income. The interactions between electricity, $\mathrm{CH_4}$ and $\mathrm{CO_2}$ are as follows:

$$V_{\mathrm{CO_2},t}^{\mathrm{P2G}}=P_t^{\mathrm{P2G}}\times\eta_{\mathrm{CO_2}}^{\mathrm{P2G}} \tag{14}$$

$$V_{\mathrm{CH_4},t}^{\mathrm{P2G}}=\frac{3.6\eta_{\mathrm{CH_4}}^{\mathrm{P2G}}}{\kappa_{\mathrm{CH_4}}}\times P_t^{\mathrm{P2G}} \tag{15}$$

where $V_{\mathrm{CO_2},t}^{\mathrm{P2G}}$ denotes the $\mathrm{CO_2}$ consumption level at time t; P_t^{P2G} denotes the electricity power consumption level of the P2G process at time t; $\eta_{\mathrm{CO_2}}^{\mathrm{P2G}}$ denotes the $\mathrm{CO_2}$ volume consumed per unit of power; $V_{\mathrm{CH_4},t}^{\mathrm{P2G}}$ denotes the production volume of $\mathrm{CH_4}$ at time t; and $\eta_{\mathrm{CH_4}}^{\mathrm{P2G}}$ denotes the $\mathrm{CH_4}$ volume to generate with a unit of power.

To ensure the stable operation of the P2G device, P_t^{P2G} mentioned earlier must obey the following constraints:

$$0\leqslant P_t^{\mathrm{P2G}}\leqslant P_{\mathrm{P2G,max}} \tag{16}$$

where $P_{\mathrm{P2G,max}}$ denotes the maximum power that is transferred from the VPP at once.

D. EV Charging Station Model

As flexible resources within the VPP that can realize charging and discharging schedules, EVs have great potential for improving the dispatch ability of the VPP. In this work, because of the independence between charge piles, the EVs exhibit charging and discharging behaviors at the same time. The coupling relationships between the SOC and the VPP are as follows:

$$O_t^{\mathrm{EVs}} = O_{t-1}^{\mathrm{EVs}} + \frac{E_{t-1}^{\mathrm{EVs}} + G_t^{\mathrm{EVs}} \times \eta_{\mathrm{VPP}}^{\mathrm{EVs}}}{E_{\mathrm{cap}}^{\mathrm{EVs}}} \tag{17}$$

$$G_t^{\mathrm{EVs}} = G_{\mathrm{char},t}^{\mathrm{EVs}} - G_{\mathrm{dis},t}^{\mathrm{EVs}} \tag{18}$$

$$G_{\mathrm{char},t}^{\mathrm{EVs}} = G_{\mathrm{char},t}^{\mathrm{base}} + G_{\mathrm{char},t}^{\mathrm{extra}} \tag{19}$$

where O_t^{EVs} and O_{t-1}^{EVs} denote the SOCs at times t and $t-1$, respectively; $E_{t\text{-}1}^{\mathrm{EVs}}$ denotes the capacity of the EVs at time $t-1$; G_t^{EVs} denotes the energy output by the EVs at time t; $\eta_{\mathrm{VPP}}^{\mathrm{EVs}}$ denotes the energy conversion efficiency; $E_{\mathrm{cap}}^{\mathrm{EVs}}$ denotes the maximum capacity of the EVs; $G_{\mathrm{char},t}^{\mathrm{EVs}}$ and $G_{\mathrm{dis},t}^{\mathrm{EVs}}$ denote the charge derived from the VPP and the discharge from the EVs to the VPP at time t, respectively; $G_{\mathrm{char},t}^{\mathrm{base}}$ denotes the charging power required for the basic demand of EV travel; and $G_{\mathrm{char},t}^{\mathrm{extra}}$ denotes the extra charging power needed for dispatch.

The logic used in this paper for $G_{\mathrm{char},t}^{\mathrm{EVs}}$ is that when SOC at time $t-1$ is less than the demand at time t, the EVs must perform $G_{\mathrm{char},t}^{\mathrm{base}}$ and then choose whether to use $G_{\mathrm{char},t}^{\mathrm{extra}}$. When SOC at time $t-1$ is greater than the demand at time t, the EVs will be charged or discharged on the basis of removing the demand at time t.

Moreover, an EV battery health mechanism is introduced in this paper. Specifically, battery health will attenuate due to the actual volume and frequency of charging and discharging. The health reduction model is as follows:

$$J_{\mathrm{doe},t}^{\mathrm{EVs}} = (G_{\mathrm{char},t}^{\mathrm{EVs}} + G_{\mathrm{dis},t}^{\mathrm{EVs}}) / E_{\mathrm{cap}}^{\mathrm{EVs}} \tag{20}$$

$$H_{\mathrm{circle},t}^{\mathrm{EVs}} = N_{\mathrm{circle},t}^{\mathrm{EVs}} \times \eta_{\mathrm{circle}}^{\mathrm{EVs}} \tag{21}$$

$$H_{\mathrm{doe},t}^{\mathrm{EVs}} = \eta_{\mathrm{doe}}^{\mathrm{EVs}} \times (J_{\mathrm{doe},t}^{\mathrm{EVs}})^2 \tag{22}$$

$$H_t^{\mathrm{EVs}} = H_{\mathrm{circle},t}^{\mathrm{EVs}} + H_{\mathrm{doe},t}^{\mathrm{EVs}} \tag{23}$$

where $J_{\mathrm{doe},t}^{\mathrm{EVs}}$ denotes the depth of the energy (DoE) change; $H_{\mathrm{circle},t}^{\mathrm{EVs}}$ denotes the health loss caused by the frequency of charging and discharging at time t; $N_{\mathrm{circle},t}^{\mathrm{EVs}}$ denotes the cumulative amount of $G_{\mathrm{char},t}^{\mathrm{EVs}}$ or $G_{\mathrm{dis},t}^{\mathrm{EVs}}$ that is greater than or equal to 0.1 MWh until time t; $\eta_{\mathrm{circle}}^{\mathrm{EVs}}$ denotes the cycle degradation rate; $H_{\mathrm{doe},t}^{\mathrm{EVs}}$ denotes the health loss caused by the DoE at time t; $\eta_{\mathrm{doe}}^{\mathrm{EVs}}$ denotes the DoE degradation rate; and H_t^{EVs} denotes the total health loss at time t.

To ensure the normal operation of the EVs, the EV deployment scheme should obey the following constraints:

$$O_{\mathrm{EVs,min}} \leqslant O_t^{\mathrm{EVs}} \leqslant O_{\mathrm{EVs,max}} \tag{24}$$

$$0 \leqslant G_t^{\mathrm{EVs}},\ G_{\mathrm{char},t}^{\mathrm{EVs}},\ G_{\mathrm{dis},t}^{\mathrm{EVs}},\ G_{\mathrm{char},t}^{\mathrm{base}},\ G_{\mathrm{char},t}^{\mathrm{extra}} \leqslant G_{\mathrm{EVs,max}} \tag{25}$$

where $O_{\mathrm{EVs,min}}$ and $O_{\mathrm{EVs,max}}$ denote the minimum and maximum SOCs of the EVs, respectively; and $G_{\mathrm{EVs,max}}$ denotes the maximum power output.

E. Battery Energy Storage System Model

In this work, a battery energy storage system (BESS) is utilized in another framework without EVs for comparison purposes. Unlike the EV charging station, the battery as a whole cannot be charged and discharged simultaneously. The energy change model of the battery is as follows:

$$O_t^{\mathrm{BESS}} = O_{t-1}^{\mathrm{BESS}} + \frac{E_{t-1}^{\mathrm{BESS}} + G_t^{\mathrm{BESS}} \times \eta_{\mathrm{VPP}}^{\mathrm{BESS}}}{E_{\mathrm{cap}}^{\mathrm{BESS}}} \tag{26}$$

where O_t^{BESS} and O_{t-1}^{BESS} denote the SOC at times t and $t-1$, respectively; E_{t-1}^{BESS} denotes the capacity of the BESS at time $t-1$; G_t^{BESS} denotes the energy output of the BESS at time t; $\eta_{\mathrm{VPP}}^{\mathrm{BESS}}$ denotes the energy conversion efficiency; and $E_{\mathrm{cap}}^{\mathrm{BESS}}$ denotes the maximum BESS capacity.

The constraints for normal battery operations are as follows:

$$O_{\mathrm{BESS,min}} \leqslant O_t^{\mathrm{BESS}} \leqslant O_{\mathrm{BESS,max}} \tag{27}$$

$$-G_{\mathrm{BESS,max}} \leqslant G_t^{\mathrm{BESS}} \leqslant G_{\mathrm{BESS,max}} \tag{28}$$

where $O_{\mathrm{BESS,min}}$ and $O_{\mathrm{BESS,max}}$ denote the minimum and maximum SOCs of the battery, respectively; and $G_{\mathrm{BESS,max}}$ denotes the maximum energy output by the battery.

F. Objective Function

In this paper, from the perspectives of economics and carbon minimization, a mathematical VPP model considering EVs is established, and the deployment solution is implemented on the basis of the TD3 algorithm to minimize its operational costs and carbon emissions. Therefore, the objective function at time t consists of the MT cost C_t^{MT}, the EVs cost C_t^{EVs}, the P2G cost C_t^{P2G}, the grid interaction cost C_t^{grid}, and the carbon emission cost C_t^{carbon}. Thus, the objective function is established as follows:

$$f_{\min}(C) = C_t^{\mathrm{total}} = C_t^{\mathrm{MT}} + C_t^{\mathrm{EVs}} + C_t^{\mathrm{P2G}} + C_t^{\mathrm{grid}} + C_t^{\mathrm{carbon}} \tag{29}$$

where $f_{\min}(C)$ denotes the minimum energy optimization cost of TD3 and C_t^{total} denotes the total cost at time t.

The cost of the MT is as follows:

$$C_t^{\mathrm{MT}} = V_{\mathrm{buy},t}^{\mathrm{MT}} \times \pi_{\mathrm{CH_4}} \tag{30}$$

where $V_{\text{buy},t}^{\text{MT}}$ denotes the CH_4 volume purchased by the VPP for the MT at time t and π_{CH_4} denotes the unit price of CH_4.

The cost incurred when EVs interact with the VPP is as follows:

$$C_t^{\text{EVs}} = \lambda_t^{\text{price}} \times G_t^{\text{EVs}} + \eta_{\text{health}}^{\text{EVs}} \times H_t^{\text{EVs}} \tag{31}$$

where λ_t^{price} denotes the electricity price at time t; and $\eta_{\text{health}}^{\text{EVs}}$ denotes the penalty cost per unit of health degradation, which is set to 10 000 ¥/MWh.

The cost of P2G is as follows:

$$C_t^{\text{P2G}} = \pi_{CH_4}(V_{\text{MT},t}^{\text{P2G}} - V_{CH_4,t}^{\text{P2G}}) + \pi_{CO_2}(V_{CO_2,t}^{\text{MT}} - V_{CO_2,t}^{\text{P2G}}) \tag{32}$$

where $V_{\text{MT},t}^{\text{P2G}}$ denotes the volume of CH_4 that is transferred to the MT to generate power and π_{CO_2} denotes the purchase and sale price for a unit of CO_2.

The grid interaction cost is as follows:

$$C_t^{\text{grid}} = \begin{cases} -\lambda_t^{\text{price}} \min(E_{\text{unbalance},t}^{\text{grid}},\ P_{\max}^{\text{grid}}), & E_{\text{unbalance},t}^{\text{grid}} \geqslant 0 \\ \lambda_t^{\text{price}} \min(\left|E_{\text{unbalance},t}^{\text{grid}}\right|,\ P_{\max}^{\text{grid}}), & E_{\text{unbalance},t}^{\text{grid}} < 0 \end{cases} \tag{33}$$

where $E_{\text{unbalance},t}^{\text{grid}}$ denotes the power imbalance observed after the devices in the VPP output and consume power at time t; and $P_{\max}^{\text{grid}}$ denotes the maximum interaction power between the VPP and the grid.

G. Power Balance

In this paper, power and gas flows are present in the VPP. Therefore, it is necessary to balance the constraints imposed on power and gas. The equilibrium constraints corresponding to the power and the two gases are as follows:

$$\begin{aligned} &G_t^{\text{WT}} + G_t^{\text{PV}} + G_t^{\text{MT}} + P_t^{\text{P2G}} - G_t^{\text{EVs}} = \\ &E_{\text{unbalance},t}^{\text{grid}} + L_t^{\text{Re}} + L_t^{\text{EVs}} \end{aligned} \tag{34}$$

where L_t^{Re} and L_t^{EVs} denote the demand loads of residents and EVs traveling at time t, respectively.

$$V_{\text{buy},t}^{\text{MT}} + V_{CH_4,t}^{\text{P2G}} = V_{CH_4,t}^{\text{MT}} + V_{\text{sell},t}^{\text{P2G}} \tag{35}$$

where $V_{\text{sell},t}^{\text{P2G}}$ denotes the volume of CH_4 that is sold by the VPP at time t.

$$V_{CO_2,t}^{\text{VPP}} = V_{CO_2,t}^{\text{MT}} - V_{CO_2,t}^{\text{P2G}} \tag{36}$$

where $V_{CO_2,t}^{\text{VPP}}$ denotes the volume of CO_2 that is emitted from the VPP to the environment at time t.

III. Optimization Algorithm for VPP-EVs

In this section, first, DQN and mRMR are used to select the most representative features for the historical charging data of the EVs, and then a GA-TCN-LSTM model is employed to forecast the demand. On the basis of the prediction module, the TD3 algorithm for VPP deployment is proposed. The framework of the whole optimization process is shown in Fig. 2.

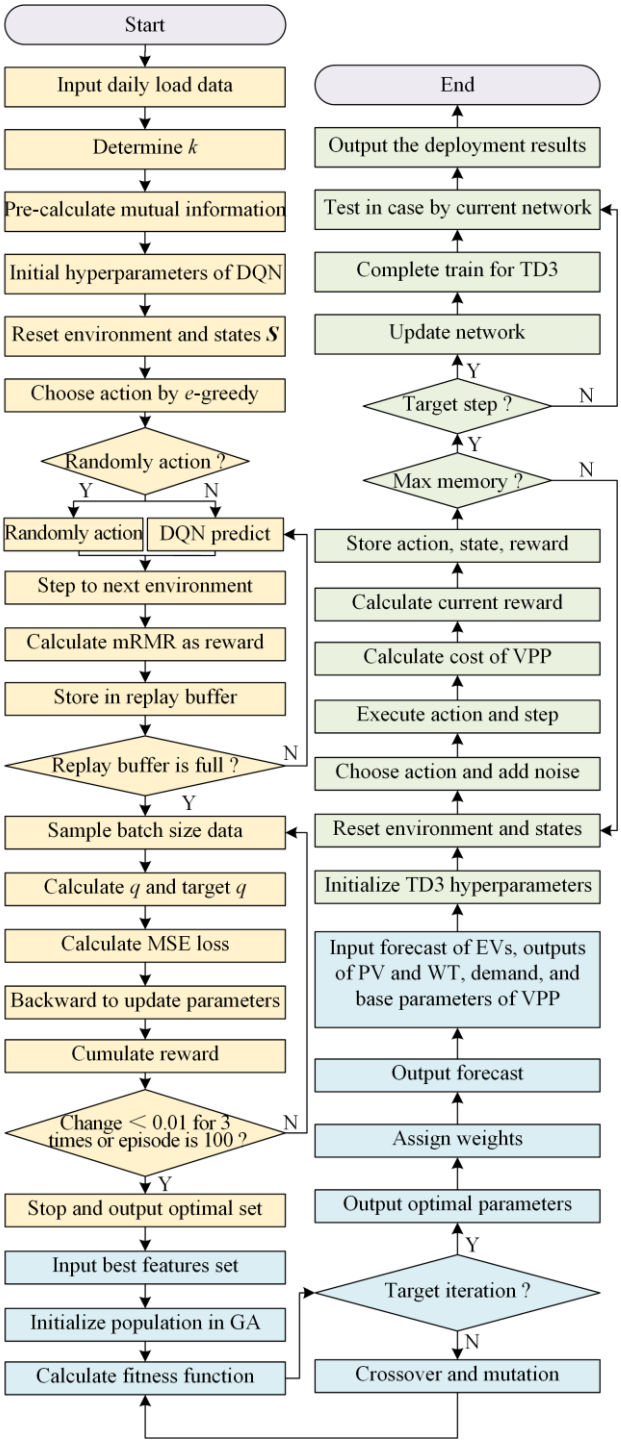


Fig. 2. Framework of the proposed optimization algorithm.

A. Model for Predicting EV Charging Station Demand

The flexibility of EVs results in some volatility risks while increasing their scheduling potential. Therefore, an accurate EV charging load prediction scheme is an important foundation for ensuring the stability of the VPP.

DQN is a deep reinforcement learning method that finds the optimal solution for the state through q value learning. Additionally, mRMR is a heuristic index that

combines the maximum relevance and minimum redundancy of information evaluation indicators to maximize the comprehensive feature evaluation. Based on the selected optimal features, the EV charging station demand prediction model is adopted in combination with TCN-LSTM on the basis of the GA parameter optimization method. TCN-LSTM improves the ability of LSTM to process feature information and increases the ability of TCN to resist nonlinear interference.

First, DQN is combined with mRMR to extract features, by utilizing mRMR as the reward function inside DQN. The mRMR calculation process is as follows:

$$M(X,Y) = -\sum_{x\in X, y\in Y} p(x,y)\log_2 p(x,y) \tag{37}$$

$$M(X) = -\sum_{x\in X} p(x)\log_2 p(x) \tag{38}$$

$$M(Y) = -\sum_{y\in Y} p(y)\log_2 p(y) \tag{39}$$

$$I(X;Y) = M(X) + M(Y) - M(X,Y) \tag{40}$$

where $M(X,Y)$ denotes the joint information entropy of X and Y; $M(X)$ and $M(Y)$ denote the information entropies of X and Y, respectively; X and Y denote the collections of features and labels, respectively; $p(x)$ denotes the probability of one feature; $p(y)$ denotes the probability of one label; $p(x,y)$ denotes the joint probability; and $I(X;Y)$ denotes the mutual information between X and Y.

Then, based on the Pearson correlation coefficient, the information redundancy is calculated as follows:

$$r_{\mathrm{XY}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(X_i-\bar{X})(Y_j-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{j=1}^{m}(Y_j-\bar{Y})^2}} \tag{41}$$

$$\phi = \frac{\sum_{i\neq j}|r_{\mathrm{XY}}|}{n\times m} \tag{42}$$

where r_{XY} denotes the correlation coefficient; $\bar{X}$ and $\bar{Y}$ denote the mean values of X and Y, respectively; n and m denote the amounts of X and Y, respectively; and ϕ denotes the average redundancy.

Finally, mRMR is presented as follows:

$$I_t^{\mathrm{mRMR}} = I(X;Y) - \phi \tag{43}$$

where I_t^{mRMR} denotes the mRMR value at time t.

In addition, DQN is employed to construct an extraction model for EV data. The state $\boldsymbol{S}_t^{\mathrm{DQN}}$ and action $\boldsymbol{A}_t^{\mathrm{DQN}}$ spaces of DQN are shown as follows:

$$\boldsymbol{S}_t^{\mathrm{DQN}} = (\boldsymbol{F}_{\mathrm{sel},t}^{\mathrm{DQN}},\ \boldsymbol{F}_{\mathrm{unsel},t}^{\mathrm{DQN}}) \tag{44}$$

$$\boldsymbol{A}_t^{\mathrm{DQN}} = (a_{1,t},\ a_{2,t},\ \cdots,\ a_{l,t}) \tag{45}$$

where $\boldsymbol{F}_{\mathrm{sel},t}^{\mathrm{DQN}}$ denotes the selected feature vector prior to time t; $\boldsymbol{F}_{\mathrm{unsel},t}^{\mathrm{DQN}}$ denotes the unselected feature vector prior to time t; and $a_{l,t}$ denotes the chosen variable for the lth feature, $a_{l,t}=1$, the corresponding feature is chosen, and $a_{l,t}=0$ is the opposite case.

The iteration process of DQN-mRMR is as follows:

$$Q_t^{-,\mathrm{DQN}} = I_t^{\mathrm{mRMR}} + \gamma^{\mathrm{DQN}}\max_{\boldsymbol{a}} Q_{t+1}^{-,\mathrm{DQN}}(\boldsymbol{s}_{t+1},\boldsymbol{a}_t;\boldsymbol{\theta}_{\mathrm{DQN,t}}^{-}) \tag{46}$$

$$\delta_t^{\mathrm{DQN}} = Q_t^{\mathrm{DQN}}(\boldsymbol{s}_t,\boldsymbol{a}_t;\boldsymbol{\theta}_{\mathrm{DQN},t}) - Q_{t+1}^{-,\mathrm{DQN}} \tag{47}$$

$$\boldsymbol{\theta}_{\mathrm{DQN},t+1} = \boldsymbol{\theta}_{\mathrm{DQN},t} - \alpha^{\mathrm{DQN}}\delta_t^{\mathrm{DQN}} \times \frac{\partial Q(\boldsymbol{s}_t,\boldsymbol{a}_t;\boldsymbol{\theta}_{\mathrm{DQN},t})}{\partial \boldsymbol{\theta}_{\mathrm{DQN},t}} \tag{48}$$

where $Q_t^{-,\mathrm{DQN}}$ denotes the target q value at time t; γ^{DQN} denotes the discount rate; $\max_a Q_{t+1}^{-,\mathrm{DQN}}$ denotes the maximum q value at time $t+1$; $\boldsymbol{s}_t$ and $\boldsymbol{s}_{t+1}$ denote the states at times t and $t+1$, respectively; $\boldsymbol{a}_t$ denotes the action at time t; $\boldsymbol{\theta}_{\mathrm{DQN,t}}^{-}$ denotes the parameters of target network at time t; δ_t^{DQN} denotes the difference between Q_t^{DQN} and $Q_{t+1}^{-,\mathrm{DQN}}$ at time t; Q_t^{DQN} denotes the q value at time t; $\boldsymbol{\theta}_{\mathrm{DQN},t}$ and $\boldsymbol{\theta}_{\mathrm{DQN},t+1}$ denote the parameters of the target network at times t and $t+1$, respectively; and α^{DQN} denotes the learning rate.

The process of DQN-mRMR is shown in Algorithm 1, where e denotes the probability of a random action, which starts at e_{s} and decreases at a rate of e_{d} to a minimum of e_{f}.

Algorithm 1 Feature Selection Model of DQN-mRMR

Input: Initial hyperparameters γ^{DQN}, e_{s}, e_{d}, e_{f}, α^{DQN}

1. $I(X;Y)$, ϕ, replay buffer $\boldsymbol{D}$
2. **for** $t=1$ **to** 2000 **do**
3. Obtain transition $(\boldsymbol{s}_t,\boldsymbol{a}_t,\boldsymbol{s}_{t+1},I_t^{\mathrm{mRMR}})$ via greedy
4. Store transition in $\boldsymbol{D}$
5. **end for**
6. **for** $t=1$ **to** 100 **do**
7. Sample batch $(\boldsymbol{s}_t,\boldsymbol{a}_t,\boldsymbol{s}_{t+1},I_t^{\mathrm{mRMR}})$ from $\boldsymbol{D}$
8. Calculate Q_t^{DQN}
9. Calculate target $Q_{t+1}^{-,\mathrm{DQN}}$
10. Compute TD error δ_t^{DQN}
11. Update $\boldsymbol{\theta}_{\mathrm{DQN},t+1}$
12. **end for**

Output: Optimal transitions $\left\{\left(\boldsymbol{s}_t,\boldsymbol{a}_t,\boldsymbol{s}_{t+1},I_t^{\mathrm{mRMR}}\right)\right\}$

By deeply embedding GA into the architectural design and training process of TCN-LSTM, GA-TCN-LSTM realizes a paradigm shift from a manual design scheme to the automatic evolution of time series prediction models. Therefore, GA is utilized to search for the optimal hyperparameters for TCN and LSTM. The flow chart of the automatic hyperparameter exploration mechanism is shown in Fig. 3.

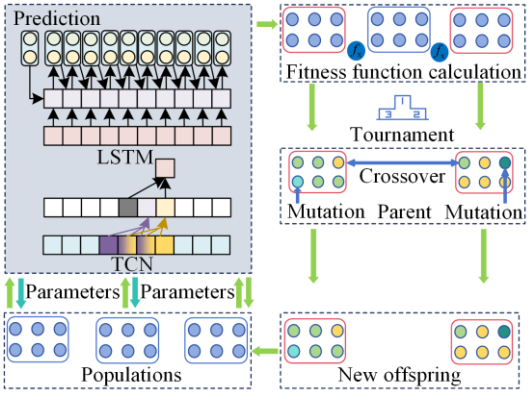


Fig. 3. Framework of GA for searching for the optimal hyperparameters of TCN and LSTM.

Notably, the root mean square error (RMSE) Δ_{RMSE} of the prediction results is chosen to calculate the fitness function for each population, calculated as follows:

$$\Delta_{\mathrm{RMSE}} = \sqrt{1/n_{\mathrm{R}}\left[\sum_{i=1}^{n_{\mathrm{R}}}(y_i - \hat{y}_i)^2\right]} \tag{49}$$

where n_{R} denotes the number of predicted data points; y_i denotes the ith actual value; and $\hat{y}_i$ denotes ith forecasted value.

B. TD3 Deployment Algorithm for the VPP

In this section, the proposed TD3 algorithm is utilized to reach carbon minimization and economic targets. The corresponding framework is shown in Fig. 4.

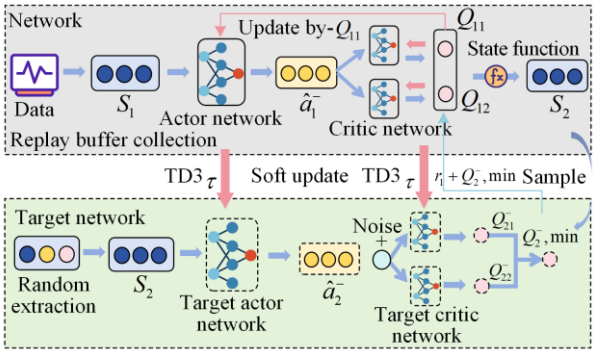


Fig. 4. Framework of the TD3 algorithm for training.

First, the state and action spaces are set as follows:

$$\boldsymbol{S}_t^{\mathrm{TD3}} = \left(t, \lambda_t^{\mathrm{price}}, G_t^{\mathrm{WT}}, G_t^{\mathrm{PV}}, L_t^{\mathrm{Re}}, L_t^{\mathrm{EVs}}, H_t^{\mathrm{EVs}}, O_t^{\mathrm{EVs}}\right) \tag{50}$$

$$\boldsymbol{A}_t^{\mathrm{TD3}} = (V_{\mathrm{CH_4},t}^{\mathrm{MT}}, V_{\mathrm{MT},t}^{\mathrm{P2G}}, G_t^{\mathrm{EVs}}, P_t^{\mathrm{P2G}}) \tag{51}$$

where $\boldsymbol{S}_t^{\mathrm{TD3}}$ denotes the state space of TD3 at time t; and $\boldsymbol{A}_t^{\mathrm{TD3}}$ denotes the action space of TD3 at time t.

Moreover, the reward function in this paper is the total cost, and is calculated as follows:

$$r_t = -C_t^{\mathrm{total}} \tag{52}$$

where r_t denotes the reward at time t.

When the experience in the replay buffer is collected, there is no need to calculate the q value. After the collection step, the TD3 samples $(\boldsymbol{s}_t, \boldsymbol{a}_t, \boldsymbol{s}_{t+1}, r_t)$ acquired from the replay buffer are used to simultaneously train the initial and target networks.

First, in the initial network part, an actor network $\pi_{\boldsymbol{\omega}_t^{\mathrm{TD3}}}(\boldsymbol{s}_t)$ outputs the predicted action $\hat{\boldsymbol{a}}_t$ and then inputs it into two critic networks to calculate two q_t^{TD3} results. In the target network part, TD3 inputs $\boldsymbol{s}_{t+1}$ and forecasts $\hat{\boldsymbol{a}}_{t+1}^-$ for the target actor and then adds noise ε to $\hat{\boldsymbol{a}}_{t+1}^-$. Furthermore, TD3 transfers $\hat{\boldsymbol{a}}_{t+1}^-$ to two target critic networks and obtains two $q_{t+1}^{-,\mathrm{TD3}}$ outputs. Consequently, to overcome the risk of overestimating q_t^{TD3}, TD3 uses a creative mechanism called clipped double Q-learning. The agent chooses the minimum value between two $q_{t+1}^{-,\mathrm{TD3}}$ options and calculates the target q value as follows:

$$q_t^{\mathrm{TD3}} = Q_t^{\mathrm{TD3}}(\boldsymbol{s}_t, \hat{\boldsymbol{a}}_t; \boldsymbol{\theta}_t^{\mathrm{TD3}}) \tag{53}$$

$$q_{t+1}^{-,\mathrm{TD3}} = \min_{i=1,2}\left\{Q_{i,t+1}^{-,\mathrm{TD3}}\left(\boldsymbol{s}_{t+1}^-, \hat{\boldsymbol{a}}_{t+1}^- + \varepsilon; \boldsymbol{\theta}_{i,t}^{-,\mathrm{TD3}}\right)\right\} \tag{54}$$

$$\varepsilon \sim \mathrm{clip}\left(\mathrm{N}(0, \sigma_{\mathrm{TD3}}^2), -1, 1\right) \tag{55}$$

$$\hat{q}_t^{-,\mathrm{TD3}} = r_t + \gamma_{\mathrm{TD3}} \times q_{t+1}^{-,\mathrm{TD3}} \tag{56}$$

where q_t^{TD3} denotes the q value at time t; Q_t^{TD3} denotes the critic network; $\boldsymbol{s}_t$ and $\hat{\boldsymbol{a}}_t$ denote the state and action at time t, respectively; $\boldsymbol{s}_{t+1}^-$ and $\hat{\boldsymbol{a}}_{t+1}^-$ denote the state and action at time $t+1$, respectively; $\boldsymbol{\theta}_t^{\mathrm{TD3}}$ denotes the parameter of the critic network; γ_{TD3} denotes the discount rate; $\hat{q}_t^{-,\mathrm{TD3}}$ denotes the target q value at time t; $q_{t+1}^{-,\mathrm{TD3}}$ denotes the q value derived from the target critic network $Q_{t+1}^{-,\mathrm{TD3}}$ at time $t+1$; ε denotes the noise added to action to provide a smooth regularization process; $Q_{i,t+1}^{-,\mathrm{TD3}}$ denotes the ith target critic network at time $t+1$; $\boldsymbol{\theta}_{i,t}^{-,\mathrm{TD3}}$ denotes the argument of the ith target critic network at time t; $\mathcal{N}$ denotes the normal distribution; and σ_{TD3}^2 denotes the variance, which is 0.01 in this paper.

After the target q value is acquired, TD3 calculates the difference between q_t^{TD3} and $\hat{q}_t^{-,\mathrm{TD3}}$. Therefore, $\boldsymbol{\theta}_t^{\mathrm{TD3}}$ is updated as follows:

$$\delta_t^{\mathrm{TD3}} = \hat{q}_t^{-,\mathrm{TD3}} - q_t^{\mathrm{TD3}} \tag{57}$$

$$L(\boldsymbol{\theta}_t^{\mathrm{TD3}}) = \frac{1}{B}\sum_{i=1}^{B}(\delta_t^{\mathrm{TD3}})^2 \tag{58}$$

$$\boldsymbol{\theta}_{t+1}^{\mathrm{TD3}} = \boldsymbol{\theta}_t^{\mathrm{TD3}} - \alpha^{\mathrm{TD3}} \times \nabla_{\boldsymbol{\theta}_t^{\mathrm{TD3}}} L(\boldsymbol{\theta}_t^{\mathrm{TD3}}) \tag{59}$$

where δ_t^{TD3} denotes the difference at time t; $L(\boldsymbol{\theta}_t^{\mathrm{TD3}})$ is the loss function at time t; $\boldsymbol{\theta}_{t+1}^{\mathrm{TD3}}$ denotes the new argument at time $t+1$; α^{TD3} denotes the learning rate of the two critic networks; and B denotes the batch size.

Regarding the parameter of the actor network, when $\hat{\boldsymbol{a}}_t$ is predicted by the actor network $\pi_{\boldsymbol{\omega}_t^{\mathrm{TD3}}}(\boldsymbol{s}_t)$ for calculating q_t^{TD3}, $\boldsymbol{\omega}_t^{\mathrm{TD3}}$ is updated as follows:

$$L(\boldsymbol{\omega}_t^{\mathrm{TD3}}) = -Q_t^{\mathrm{TD3}}(\boldsymbol{s}_t, \boldsymbol{a}_t; \boldsymbol{\theta}_t^{\mathrm{TD3}}) \tag{60}$$

$$\nabla_{\boldsymbol{\omega}_t^{\mathrm{TD3}}} L(\boldsymbol{\omega}_t^{\mathrm{TD3}}) = \mathbb{E}_{\boldsymbol{s}_t}\begin{bmatrix}\nabla_{\boldsymbol{a}_t} Q_t^{\mathrm{TD3}}(\boldsymbol{s}_t, \boldsymbol{a}_t; \boldsymbol{\theta}_t^{\mathrm{TD3}}) \times \\ \nabla_{\boldsymbol{\omega}_t^{\mathrm{TD3}}} \pi_{\boldsymbol{\omega}_t^{\mathrm{TD3}}}(\boldsymbol{s}_t)\end{bmatrix} \tag{61}$$

$$\boldsymbol{\omega}_{t+1}^{\mathrm{TD3}} = \boldsymbol{\omega}_{t}^{\mathrm{TD3}} + \beta^{\mathrm{TD3}} \times \nabla_{\boldsymbol{\omega}_{t}^{\mathrm{TD3}}} L(\boldsymbol{\omega}_{t}^{\mathrm{TD3}}) \tag{62}$$

where $L(\boldsymbol{\omega}_t^{\mathrm{TD3}})$ denotes the loss function at time t; $\nabla_{\boldsymbol{\omega}_t^{\mathrm{TD3}}} L(\boldsymbol{\omega}_t^{\mathrm{TD3}})$ denotes the gradient with respect to $L(\boldsymbol{\omega}_t^{\mathrm{TD3}})$ at time t; $\mathbb{E}_{\boldsymbol{s}_t}$ denotes the expectation calculated under $\boldsymbol{s}_t$ at time t; $\nabla_{\boldsymbol{a}_t}$ and $\nabla_{\boldsymbol{\omega}_t^{\mathrm{TD3}}}$ denote the gradients of $\boldsymbol{a}_t$ and $\boldsymbol{\omega}_t^{\mathrm{TD3}}$, respectively; $\pi_{\boldsymbol{\omega}_t^{\mathrm{TD3}}}$ denotes the actor network at time t; $\boldsymbol{\omega}_t^{\mathrm{TD3}}$ and $\boldsymbol{\omega}_{t+1}^{\mathrm{TD3}}$ denote the parameters of the critic network at times t and $t+1$, respectively; and β^{TD3} denotes the learning rate of the actor network.

Notably, the parameters of the three target networks undergo a soft update at a predetermined interval during the training process.

The algorithm is elaborated in Algorithm 2, where $\boldsymbol{\omega}_t^{-,\mathrm{TD3}}$ denotes the parameters of the target actor network; τ^{TD3} denotes the soft update rate; and d^{TD3} denotes the soft update target step. After the optimal dispatch is observed at time t, the current information is stored in the historical collection for use by DQN-mRMR to perform feature extraction again.

Algorithm 2 TD3 Optimization Method

Input: t, $\lambda_t^{\mathrm{price}}$, G_t^{WT}, G_t^{PV}, L_t^{Re}, L_t^{EVs}

1. **Initialize:** $\boldsymbol{S}_t^{\mathrm{TD3}}$, $\boldsymbol{A}_t^{\mathrm{TD3}}$, $\boldsymbol{\theta}_t^{\mathrm{TD3}}$, $\boldsymbol{\theta}_{i,t}^{-,\mathrm{TD3}}$, $\boldsymbol{\omega}_t^{\mathrm{TD3}}$, $\boldsymbol{\omega}_t^{-,\mathrm{TD3}}$

Phase 1: data collection

2. **for** $t=1$ **to** 8760 **do**
3. Observe state $\boldsymbol{s}_t$, execute action $\boldsymbol{a}_t$
4. Store transition $(\boldsymbol{s}_t, \boldsymbol{a}_t, \boldsymbol{s}_{t+1}, r)$ in replay buffer
5. **end for**

Phase 2: parameter update

6. **for** $t=1$ **to** 100 **do**
7. **if** t mod $d^{\mathrm{TD3}} = 0$ **then**
8. Soft update target networks:
$\boldsymbol{\theta}_{i,t}^{-,\mathrm{TD3}} \leftarrow \tau^{\mathrm{TD3}} \boldsymbol{\theta}_t^{\mathrm{TD3}} + (1-\tau^{\mathrm{TD3}}) \boldsymbol{\theta}_{i,t}^{-,\mathrm{TD3}}$
$\boldsymbol{\omega}_t^{-,\mathrm{TD3}} \leftarrow \tau^{\mathrm{TD3}} \boldsymbol{\omega}_t^{\mathrm{TD3}} + (1-\tau^{\mathrm{TD3}}) \boldsymbol{\omega}_t^{-,\mathrm{TD3}}$
9. **end if**
10. Sample batch $(\boldsymbol{s}_t, \boldsymbol{a}_t, \boldsymbol{s}_{t+1}, r_t)$ from replay buffer
11. Compute target actions $\hat{\boldsymbol{a}}_{t+1}^{-}$
12. Calculate $\hat{q}_{t+1}^{-,\mathrm{TD3}}$
13. Update critic: $\boldsymbol{\theta}_t^{\mathrm{TD3}}$
14. Update actor: $\boldsymbol{\omega}_t^{\mathrm{TD3}}$
15. **end for**

Phase 3: dispatch execution

16. **for** $t=1$ **to** 24 **do**
17. Observe state $\boldsymbol{s}_t$, select action $\boldsymbol{a}_t$
18. Execute action and receive reward r_t
19. **end for**

Output: Optimal dispatch trajectories $\left\{\left(\boldsymbol{S}_t^{\mathrm{TD3}}, \boldsymbol{A}_t^{\mathrm{TD3}}, \boldsymbol{r}_t\right)\right\}_{t=1}^{24}$

IV. Case Study

In this study, the proposed TD3 optimization algorithm is employed to dispatch DERs in a VPP with EVs. The calculation results are validated by using actual data acquired from the University of California San Diego microgrid (with residential demand and EV charging demand) and the Xinjiang Uygur Autonomous Region (radiation intensity and wind speed data). The selected period is from January 1, 2019, to December 31, 2019. First, the main parameters, which include DERs and methods, are introduced. Second, four cases are set to verify the generalization ability of the TD3 model. These cases are generated based on an uncertainty analysis. In a comparison with three other methods, the superiority and rationality of the proposed TD3 decision-making method are tested and validated. The test is conducted on a personal computer, which is equipped with an 11th-Gen Intel(R) Core(TM) i5-11300H processor @3.10 GHz-3.11 GHz and 16 GB of RAM. The operating system is Windows 11.

A. Basic Arguments

As depicted in Section II, the device environment of the VPP examined in this paper is illustrated. The corresponding equipment arguments, as well as the prices of electricity and gas, the carbon emission penalty and the hyperparameters of the DQN and TD3, are detailed in Table I.

TABLE I
PARAMETERS OF THE VPP MODEL

Parameters	Value	Parameters	Value
$G_{\mathrm{rated}}^{\mathrm{WT}}$ (MWh)	2	$\eta_{\mathrm{doe}}^{\mathrm{EVs}}$	0.002
v_{rated} (m/s)	25	$O_{\mathrm{EVs,max}}$, $O_{\mathrm{BESS,max}}$	0.9
S_{PV} (m^2)	1	$O_{\mathrm{EVs,min}}$, $O_{\mathrm{BESS,min}}$	0.1
N_{PV}	1	$G_{\mathrm{EVs,max}}$, $G_{\mathrm{BESS,max}}$ (MWh)	7
η_{PV}	0.8	$\pi_{\mathrm{CH_4}}$ (¥/kg)	2.4
η_{MT}	0.6/0.3	$\pi_{\mathrm{CO_2}}$ (¥/kg)	0.946
$G_{\mathrm{MT,max}}$ (MWh)	10/20	γ^{DQN}	0.999
$G_{\mathrm{MT,min}}$ (MWh)	0	α^{DQN}	0.01
$\Delta G_{\mathrm{ramp}}^{\mathrm{MT}}$	2/4	e_{s}	0.9
$\eta_{\mathrm{CO_2}}^{\mathrm{MT}}$ (kg/MWh)	550	e_{d}	100
$\kappa_{\mathrm{CH_4}}$ (MJ/kg)	50	e_{f}	0.01
$\eta_{\mathrm{CO_2}}^{\mathrm{P2G}}$ (kg/MWh)	128.7	γ^{TD3}	0.99
$\eta_{\mathrm{CH_4}}^{\mathrm{P2G}}$	0.65	α^{TD3}	0.001
$P_{\mathrm{P2G,max}}$ (MWh)	16	β^{TD3}	0.001
$\eta_{\mathrm{VPP}}^{\mathrm{EVs}}$, $\eta_{\mathrm{VPP}}^{\mathrm{BESS}}$	0.9	ε	2
$E_{\mathrm{cap}}^{\mathrm{EVs}}$, $E_{\mathrm{cap}}^{\mathrm{BESS}}$ (MWh)	10	d^{TD3}	2
$\eta_{\mathrm{circle}}^{\mathrm{EVs}}$	0.001	τ^{TD3}	0.01

B. Case Settings

To prove the generalizability of the proposed TD3 dispatch approach, four cases are set up for calculation

and analysis purposes. The four case environments produced by an uncertainty analysis and forecasting are shown in Fig. 5. In this paper, the electricity price and EV demand are the same in the four cases.

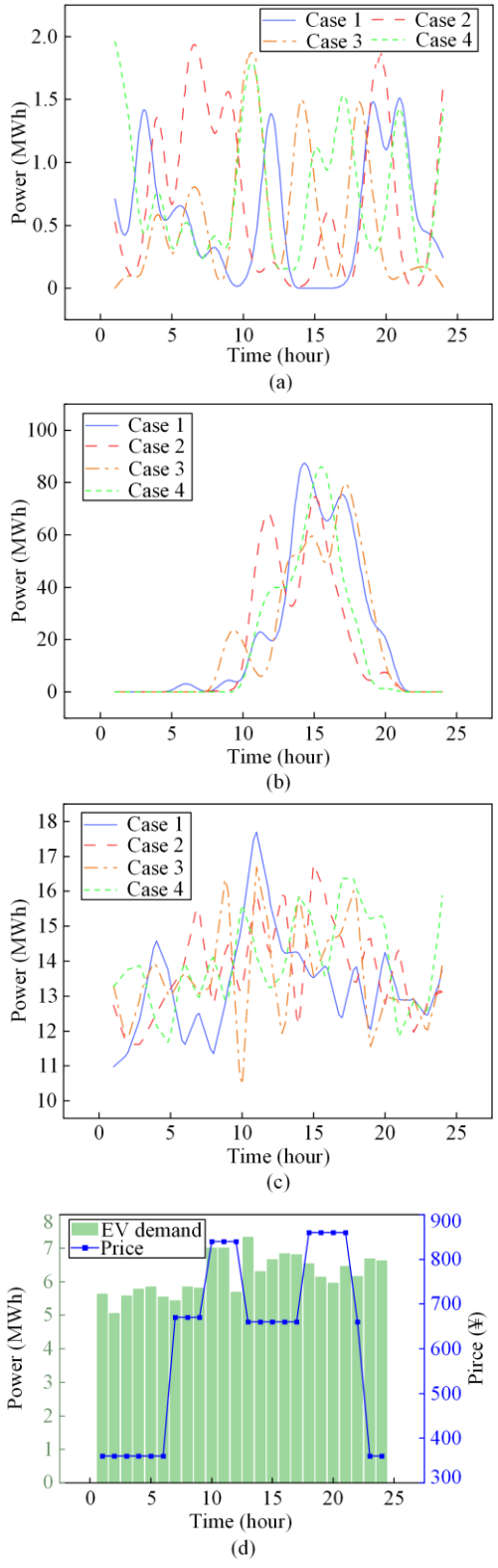


Fig. 5. Environment sets for the four cases. (a) Wind power. (b) Photovoltaic power. (c) Residential demand. (d) Electricity price and EV charging demand.

The basic charging demand of the EVs is forecasted by GA-TCN-LSTM with features selected by DQN-mRMR. The other environments employed in the four cases are generated by an uncertainty analysis. In addition, three algorithms are utilized for Case 1 to examine the dominance of the TD3 deployment model with EVs. The three algorithms are designed in terms of energy optimization performance and EV changes. Among them, Gurobi is a solver for MILP and quadratic problems. In this paper, it is used for the model presented in Section II and implemented via the gurobipy package in Python. Moreover, after the SOC variation results are obtained, the EV health and cost changes are manually incorporated.

For the MPC scene, a battery capacity reduction mechanism is added. This mechanism causes the available battery capacity to decrease as the health of the battery decreases. The details of the above mechanism are as follows:

$$E_{\text{cap},t+1}^{\text{EVs}} = E_{\text{cap},t}^{\text{EVs}} \times (1 - H_t^{\text{EVs}}) \tag{63}$$

where $E_{\text{cap},t}^{\text{EVs}}$ and $E_{\text{cap},t+1}^{\text{EVs}}$ denote the maximum EV volumes at times t and t+1, respectively.

In addition, the prediction horizon of MPC is set to 5. To simulate a more refined model, a prediction error parameter based on the scene data is incorporated, and another mechanism in which the prediction error increases over time is established.

An extra VPP-BESS framework is constructed to compare the total costs of the VPPs in the four cases. The algorithms for testing and configuring the two VPP frameworks are shown in Table II. The simulation results obtained for the above scenes are analyzed from the aspects of economics, EV health and carbon minimization.

TABLE II
ALGORITHMS TESTED UNDER TWO VPPS CONFIGURATIONS

Algorithm	EVs	BESS
TD3	TD3-EVs	TD3-BESS
DDPG	DDPG-EVs	
Gurobi	Gurobi-EVs	
MPC	MPC-EVs	

C. Prediction Framework Verification

Figure 6 shows the training processes of diverse DQN hyperparameter sets. The first set exhibits stable convergence within 100 episodes. To avoid the problem of falling into a local optimal solution, the episode of the first curve is prolonged. This set eventually converges to an average reward of approximately –4.33 for the last 25 episodes. The reward provided by group 1 in the first 100 episodes (–3.73) is 58.2% higher than that of the second group (–8.92), and group 1 is close to converging to a positive reward (–0.12) in the later periods

(100–200), indicating that its stability and convergence speed are better. The standard deviation of the rewards in the first group for periods 50–100 is 7.2, which is much lower than the 12.5 and 14.8 values of the second and fourth groups, respectively, and the volatility is reduced by 42.4% and 51.4%. Thus, the first hyperparameter set is chosen to train the DQN and extract the optimal features for EV demand forecasting.

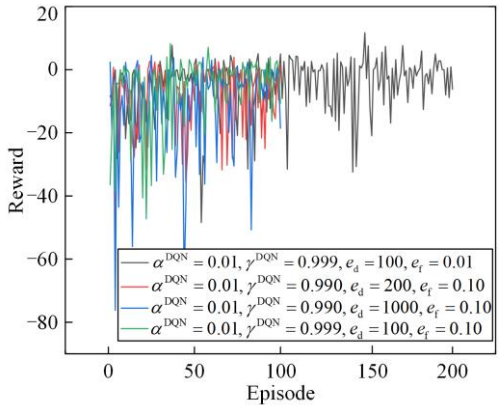


Fig. 6. Rewards curves produced with four hyperparameter sets in the training stage of the DQN.

After training, the features are selected by the mRMR calculation and DQN exploration processes. The mRMR and recursive feature elimination (RFE) approaches are then employed to test the superiority of the proposed DQN-mRMR model. On the basis of the same datasets and test size, a random forest classifier is utilized to present the performance of the three algorithms. The results are shown in Table III. DQN-mRMR has better selection accuracy (0.982) than those of mRMR and RFE, but it is slower than the other algorithms because of its computational complexity.

TABLE III
PERFORMANCE COMPARISON AMONG THREE ALGORITHMS

Algorithm	Accuracy	Time (s)	Features
DQN-mRMR	0.982	8.053	7
mRMR	0.873	0.063	6
RFE	0.673	0.022	6

Figure 7 illustrates the comparison between the forecasting conditions (time, demand, longitude, and latitude) based on the DQN-mRMR and manual feature selection methods. According to the GA evaluation of the optimal hyperparameters yielded by the TCN and LSTM, the window size is set to 5. Hence, the predicted load has a delay of 5 hours with respect to the actual load. As shown in Fig. 7, the RMSE, mean absolute error (MAE) and R^2 in Fig. 7(a) are all greater than those in Fig. 7(b), which means that the forecasting performance of the features derived from DQN-mRMR is better than that of manual selection. Consequently, the demand curve forecasted by the features extracted from DQN-mRMR are used as the EV demand scene for the four cases.

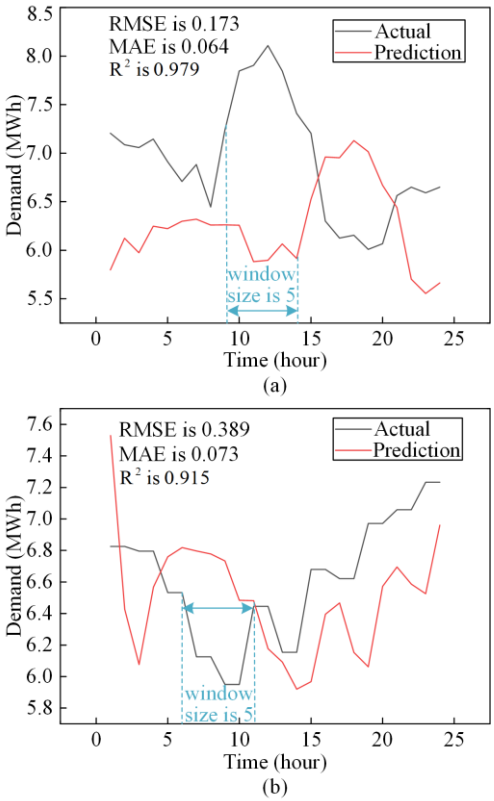


Fig. 7. Comparison between the actual and prediction demands based on the DQN-mRMR and manual methods. (a) DQN-mRMR. (b) Manual method.

D. Comparison Among the Dispatch Performances of the Four Algorithms with EVs

For Case 1, the TD3 training reward conditions obtained with different hyperparameters are shown in Fig. 8. The optimization results produced by the four algorithms are shown in Fig. 9, where the portion below 0 indicates the power sale behavior with respect to the grid. In contrast, the part above 0 includes the MT output and the energy change exhibited by the EVs, as well as the power purchased from the grid for P2G or other processes.

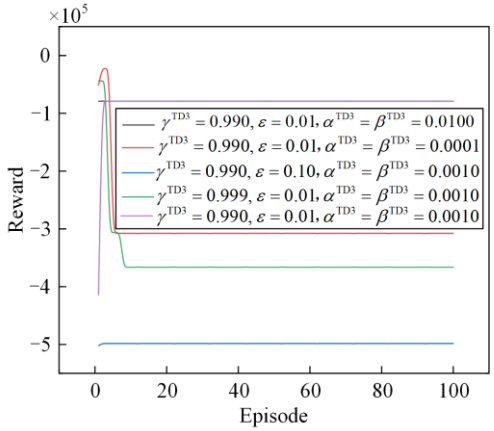


Fig. 8. Rewards yielded by the TD3 and DDPG algorithms under different hyperparameters within 100 episodes.

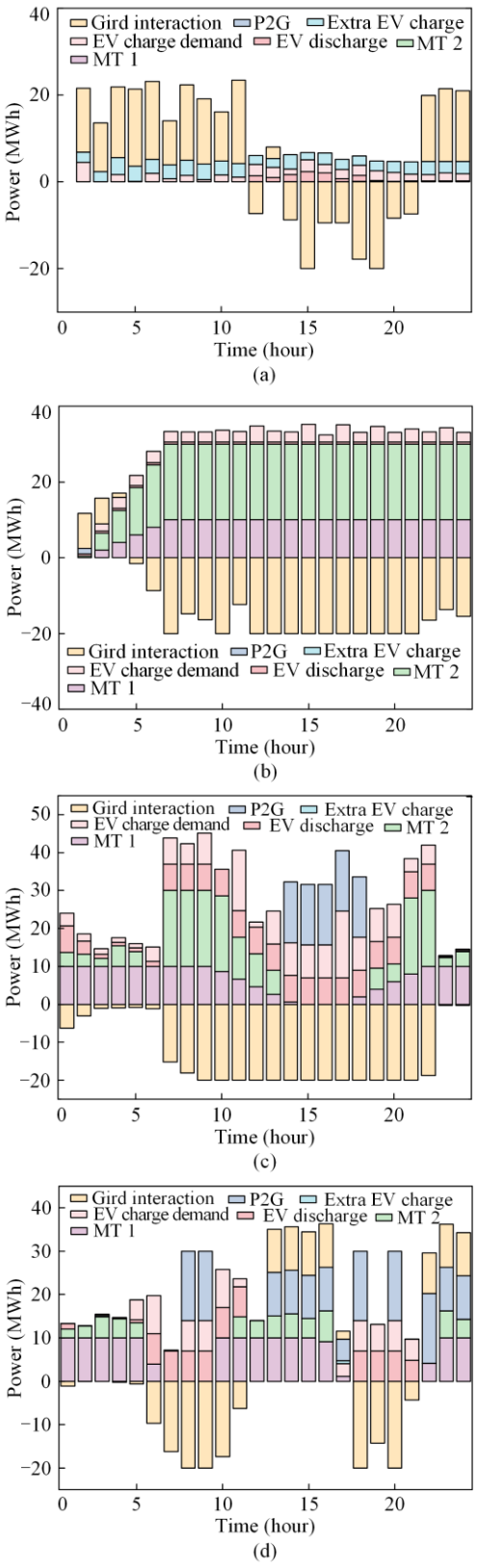


Fig. 9. Energy optimization results obtained with the proposed TD3 algorithm and the other three algorithms. (a) TD3. (b) DDPG. (c) Gurobi. (d) MPC.

P2G refers to the electricity purchased from the VPP for achieving a CO_2 reduction. The EV charging demand refers to the charging power needed to meet the imposed demand. The extra EV charge is the additional power charge derived from the VPP after the demand is satisfied. As shown in Fig. 8, five hyperparameter combinations are used by TD3 for parameter training. The hyperparameter set of the DDPG approach is the same as the fifth combination of TD3, which is the set chosen in this paper. The reward demonstrates the impacts of diverse hyperparameters.

Initially, various learning rates are tested. When α^{TD3} and β^{TD3} are 0.01, which is higher than the values in the chosen set, the reward converges within two episodes to 79 880.968. The reward process of combination 1 is similar to that of the chosen set (0.001). However, in the whole training stage, the reward remains stable after two episodes, and a risk of overfitting is present. Therefore, we examine other combinations. When the learning rate is 0.0001, which is lower than the chosen rate, the reward exceeds both 0.01 and 0.001 in the four episodes and then converges to –307 655.053.

Furthermore, a larger ε is observed, but the corresponding reward converges rapidly to –498 450.605, which is clearly below all other results.

Compared with that of the chosen set, the reward obtained with $\gamma^{\mathrm{TD3}} = 0.999$ is higher in three episodes; then, it decreases to a lower reward, –366 441.162, which is less than that of combination 2. Finally, the reward obtained with the DDPG algorithm under the same chosen hyperparameters is presented, but it does not converge.

Above all, ε affects the training performance, γ^{TD3} has a smaller effect, whereas α^{TD3} and β^{TD3} have the smallest impacts. In view of the above perspectives, the fifth hyperparameter combination is chosen as the training set for TD3 in this paper.

The energy deployment schemes produced based on the four algorithms are shown in Fig. 9.

In contrast with the other three algorithms, TD3 significantly reduces the power output by the MTs to only 1.163 MWh, which is 0.31% of that of the DDPG algorithm and lower than those of Gurobi and MPC (99.44% and 99.45%, respectively).

Moreover, TD3 has a better performance than other methods in terms of EV energy interactions (except the power charged to satisfy the basic demand). TD3 results in an EV energy change volume of 70.771 MWh, while DDPG algorithm has the smallest volume (12.57 MWh), MPC is similar to TD3 (70.14 MWh), whereas Gurobi has the greatest volume (127.314 MWh).

Notably, for TD3 and MPC, the extra charging powers are 64.025 MWh and 1.02 MWh, respectively. This situation explains the EV energy schedule capacity of the proposed TD3 model.

TD3 demonstrates a strong grid exchange ability (310.719 MWh in total), whereas MPC (221.102 MWh) is inferior. Although the results of the DDPG algorithm

(356.778 MWh) and Gurobi (325.941 MWh) are higher than those of both TD3 and MPC, the entire interaction involves electricity sales, and no obvious dispatch pattern is observed according to the electricity prices. In the peak pricing period, TD3 and MPC present the power sale action, and they choose to buy during the valley period. With respect to P2G power in the purchasing part, TD3 provides a result of 0, while the DDPG algorithm, Gurobi, and MPC yield 1.542 MWh, 80 MWh, and 144.974 MWh, respectively. These results indicate the carbon emission control capabilities of the four algorithms. TD3 does not require additional expenses to avoid the carbon emission penalty, and thus, the total cost of the VPP decreases.

Figure 10 shows the SOC change and EV health degradation process observed under the four algorithms. The results indicate the keen response capabilities of EVs based on TD3 dispatching.

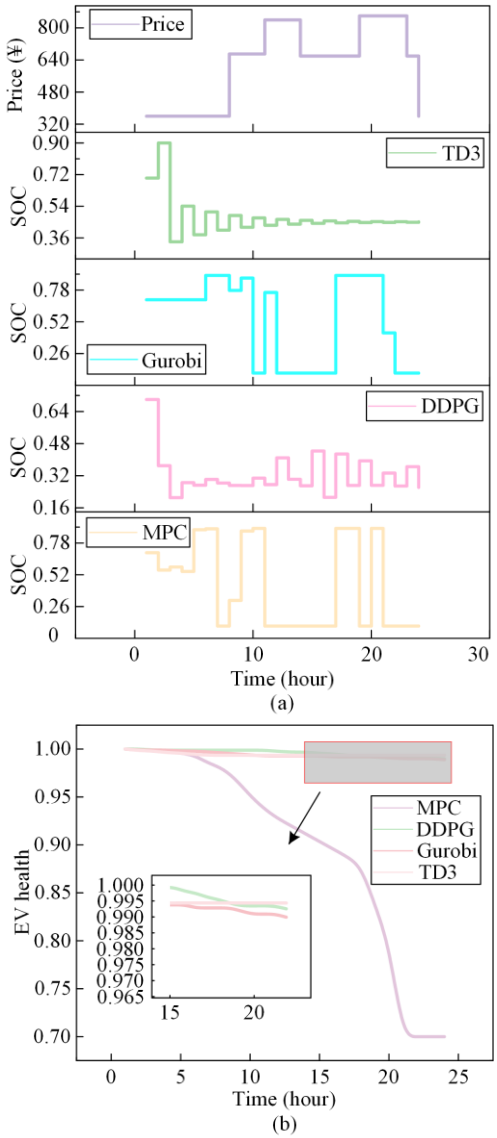


Fig. 10. SOC change and EV health degradation results obtained based on four algorithms with electricity prices. (a) SOC. (b) EV health.

As shown in Fig. 10(a), the SOC curve of TD3 increases to 0.9 and then decreases to 0.34, whereafter it presents increasingly small fluctuations and finally tends to converge to approximately 0.45. The reason for this phenomenon is the EV battery health setting. Owing to this mechanism, TD3 keeps the SOC variation beyond the range that would cause healthy attenuation, thereby reducing the cost of battery depletion. Finally, TD3 explores the balance between protecting the health of the EVs and satisfying the next demand imposed after the EV discharging process.

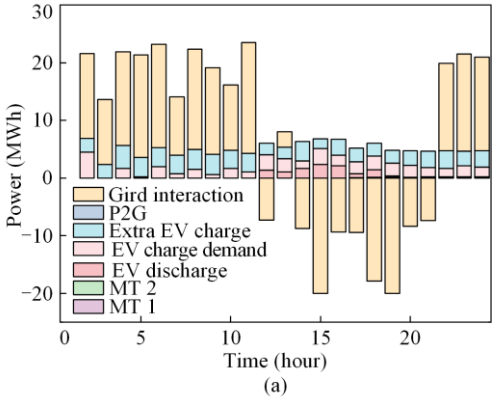
Similar to TD3, the DDPG algorithm also presents certain EV health monitoring and control capabilities. Nevertheless, both Gurobi and MPC experience multiple cases of overcharging and discharging, which lead to rapid declines in EV health.

As shown in Fig. 10(b), TD3 has a better performance than the other methods in terms of protecting the health of EVs, with the EV health reduced from 1.0 to 0.994 (total attenuation: 0.641%), and the attenuation rate being only 63.3% of that of the DDPG algorithm (1.013%), 58.3% that of Gurobi (1.1%), and 95.9% lower than that of MPC (30.0%). In addition, the standard deviation of the TD3 attenuation process is 0.000 32, which is 89.0% lower than that of the DDPG algorithm (0.003) and 99.7% lower than that of MPC (0.098). Notably, the reason for the faster EV health decline exhibited by MPC is the effect of the declining health on the available EV capacity.

In summary, all the algorithms respond to the electricity price to some extent (charging during the low-price period from 01:00–08:00 and discharging during the high-price periods from 12:00–14:00 and 19:00–22:00). However, TD3 has better performance in terms of maintaining the balance between EV health and the electricity price.

E. Algorithmic Generality Verification Based on Four Scenarios

To verify the generalization ability of the proposed TD3 deployment model, Cases 1–4 are implemented. The case details are provided in Fig. 5. Figure 11 shows the energy deployment conditions of the four cases, while Fig. 12 illustrates the diverse SOC variations observed in the four cases.



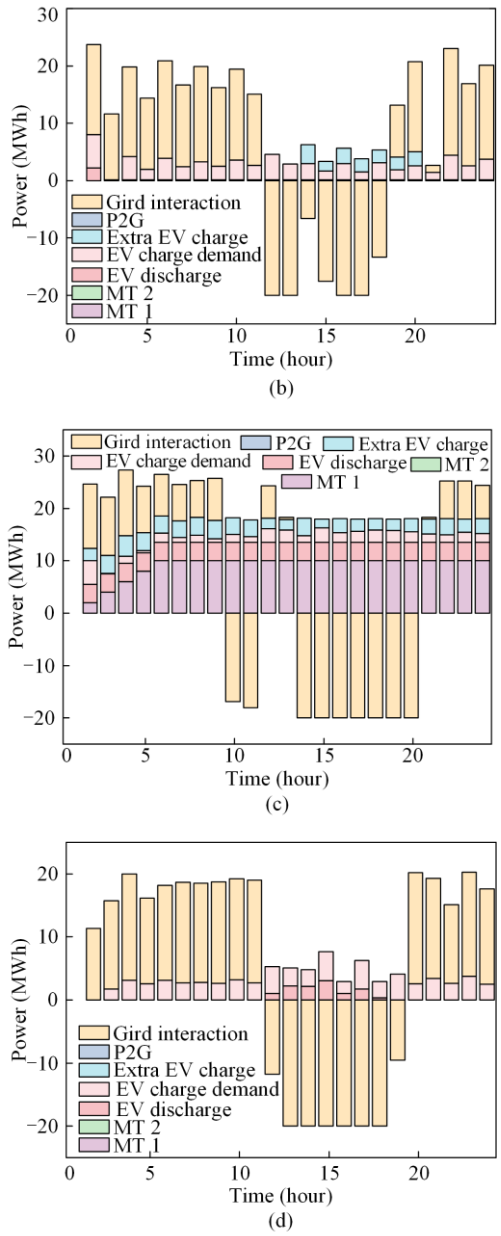


Fig. 11. Energy deployment details of four cases based on TD3. (a) Case 1. (b) Case 2. (c) Case 3. (d) Case 4.

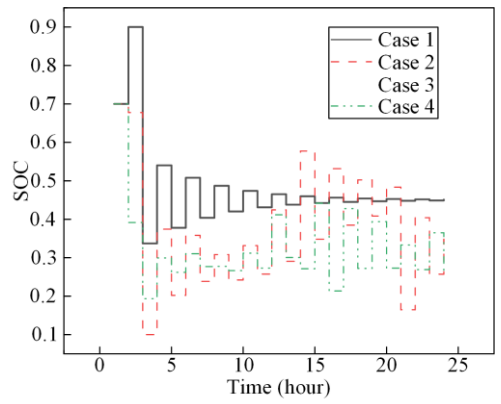


Fig. 12. EV SOC changes observed in the four cases based on TD3.

As shown in Fig. 11, TD3 always exhibits a certain rule for the schedules under the four cases. And the grid interactions always maintain the trend of purchasing power during electricity price valleys and selling power during electricity price peaks. Furthermore, the output of the MTs in the four cases is 211.163 MWh, including 1.163 MWh in Case 1 and 210 MWh in Case 3, as shown in Figs. 11(a) and (c), respectively. The MTs in the other cases provide no outputs. In Cases 1–3, the EVs have an extra charge of 145.988 MWh. In addition, Case 1 and Case 4 have EV discharging during the high-price period, and the total volume is 22.347 MWh in Figs. 11(a) and (d). Case 3 has a consistent EV discharging effect over the whole period, as shown in Fig. 11(c). Additionally, Cases 1–4 have stable performance in terms of grid interactions, with average values of 3.898 MWh, 4.29 MWh, 3.019 MWh and 3.631 MWh, respectively. Notably, in all the cases, the P2G electricity purchase level is 0, indicating great economic ability.

As depicted in Fig. 12, Case 1 has a similar SOC pattern to that of Case 3, and the similarity between the two scenes reaches 98%. Case 2 is similar to Case 4, their SOC trends remain highly consistent (85.4% similarity), but the difference comes at 14:00. This is because the residential demand in Case 2 at 14:00 is higher than that in Case 4, and the EVs need to charge at this moment in Case 2. In the meantime, the fluctuations exhibited by the four cases have stable changes; only in Case 2 does the SOC reach the lower bound of 0.1 (this case has the largest volatility of 57.75%), whereas that in Case 1 is 56.29%, that in Case 3 is 18.11%, and that in Case 4 is 24.84%. Moreover, the EV discharge volume proportions observed during the high-price period are more than 62.9% in Cases 1–4, and the EV charging proportions observed during the valley period are more than 68.5%. The mean standard deviation of the SOC in the four scenarios is 0.062.

In summary, the TD3 model has superior adaptability and reasonably dispatches the resources and requirements in different environments. It has the ability to absorb surplus power, meet different needs and reduce operational costs.

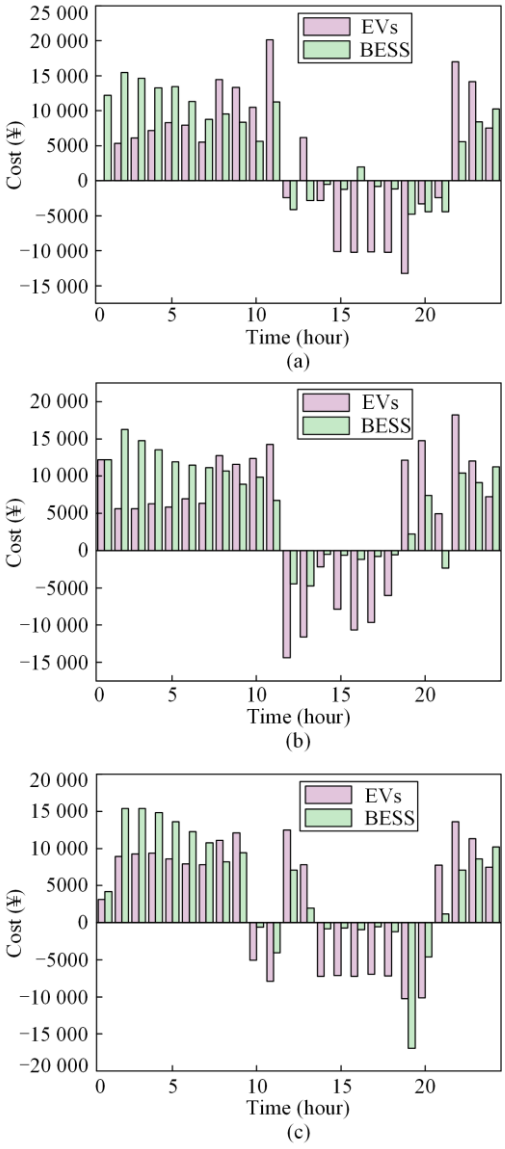
F. Economy Comparison

Economic viability is an essential condition for VPPs to achieve sustainable operations. To analyze the superior economic performance of the proposed TD3 algorithm, the operational costs of four cases and two VPPs frameworks are presented.

As shown in Fig. 13, the EV scenes in Cases 1–4 present certain patterns relative to the BESS scenes. From 01:00–08:00 and 22:00–24:00, the costs of the EVs are higher than those of the BESS During 12:00–21:00, the EVs gain more than the BESS does except for Case 3 in Fig. 13(c). Figures 13(a), (b), and (d)

present more negative costs, which result in more revenue. The total cost of the EV scene is ¥ 78 991, which is less than the ¥ 125 927 cost of the BESS scene, thus saving 37.3% in Case 1. In the other three cases, the total costs of the EV scenes are lower than those of the BESS scenes by 30%, 36.8%, and 52.4%, respectively. The average cost reduction rate is 39.1% in the EV scenarios for Cases 1–4.

In addition, the income of the EV scene is ¥ –64 835, which is 2.67 times greater than that of the BESS scene in Case 1, whereas in Case 4, the income of the EV scene is 4.67 times greater than that of the BESS scene. In addition, the cost of the EV scene is ¥ 54 930 during the low-price period (01:00–08:00), but throughout the high-price period (12:00–21:00), the VPP accrues ¥ 64 835, achieving net income. At the same time, the BESS scene has a greater cost (¥ 89 091) during the low-price period and only earns ¥ 22 343 during the high-price period.



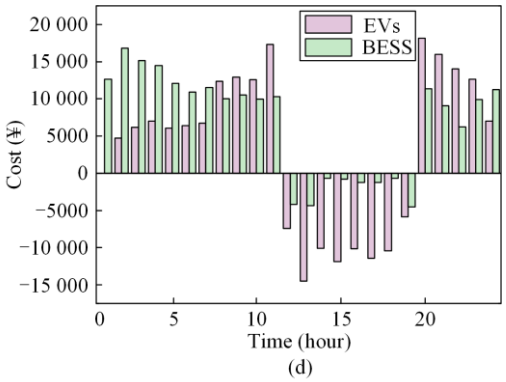


Fig. 13. Operational cost comparisons between EVs and BESS frameworks based on TD3 in Cases 1–4. (a) Case 1. (b) Case 2. (c) Case 3. (d) Case 4.

As shown in Fig. 14, the costs of the four algorithms demonstrate the advantages of the TD3 model. The total cost of TD3 is lower than those of Gurobi, the DDPG algorithm and MPC (41.46%, 57.42%, and 2.13%, respectively). In addition, the total cost of MPC is similar to that of TD3. However, from 06:00–10:00, MPC is outstanding due to its negative cost, whereas TD3 still keeps its positive cost controllable to avoid the drastic fluctuations exhibited by MPC. At the same time, Gurobi has a higher cost, and DDPG has a lower cost. From 13:00–21:00, DDPG has a higher cost than does TD3.

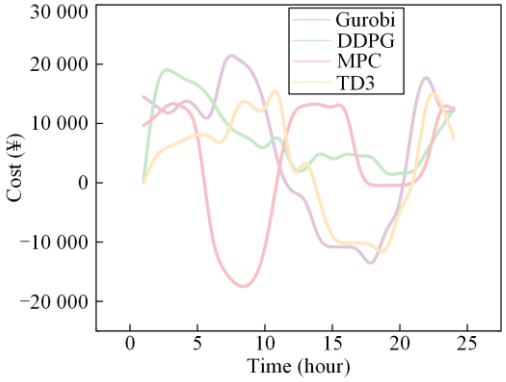


Fig. 14. Operational cost comparison among the four algorithms in Case 1.

In summary, TD3 has significant economy in different cases. The participation of EVs provides more opportunities for the VPP to increase its revenue through deployment. Compared with the other algorithms, TD3 has certain advantages in terms of its total cost and stability in the face of complex and variable EV scheduling scenarios.

G. Carbon Minimization Degree Verification

To test the environmental benefit of the proposed TD3-EV framework, CO_2 emissions, the CH_4 production derived from P2G, and the CO_2 penalty are described, and the results are shown in Table IV.

TABLE IV
CARBON MINIMIZATION ABILITY VALIDATION FOR EVS AND THE BESS BASED ON TD3

Data	Scenes	1	2	3	4
CO_2 (kg)	EVs	−55 720.6	−38.01	60 834.6	−33 611
	BESS	166 314.3	171 983.2	171 949	169 717.5
CH_4 (kg)	EVs	27.97	21.762	26.966	22.285
	BESS	21.47	11.024	17.905	12.222
Penalty (¥)	EVs	−5272.8	−3.598	5756.7	−3180.6
	BESS	15 738.3	16 274.7	16 271.5	16 060.3

In terms of CO_2 emissions, the EVs have stronger carbon minimization capabilities in the four cases. In Case 1, the CO_2 emissions of EVs are −55 720.6 kg (a negative value indicates carbon absorption from the environment), whereas the emissions of the BESS are 166 314.3 kg. Compared with the BESS, the EVs reduce the emissions by 133.50%, achieving a significant decrease in carbon. In Case 2, the EV emissions are −38.01 kg, the BESS emissions are 171 983.2 kg, and the EV emission reduction rate is 100.02%. In Case 4, the EVs further provide negative emissions of −33 611 kg, which are 119.80% lower than those of the BESS. Even in Case 3, which has positive EV emissions, the carbon emissions are still 64.62% lower than those of the BESS.

In terms of CH_4 production, the EV scene produces slightly more than the BESS does. The mean CH_4 production level of the EVs (24.75 kg) is 58.05% greater than that of the BESS (15.66 kg). Additionally, the EVs achieve a negative cost (i.e., earning from selling carbon certification) in Cases 1, 2 and 4. Even in Case 3, where the EVs have to pay a penalty cost (¥ 5 756.7), their cost is still lower than that of the BESS (¥ 16 271.5). Synthesizing the data of the four cases, the average penalty cost incurred by the EVs is ¥ −1074.82, the average BESS penalty is ¥ 16 086.01, and the overall cost of the EVs is reduced by 106.68%.

The above results show that the TD3-EV framework based on the TD3 algorithm accounts for both environmental and economic benefits.

V. CONCLUSIONS

A low-carbon collaborative scheduling model is constructed for a VPP integrated with EVs. Through a dynamic battery health degradation mechanism, multi-energy flow coupling constraints and a cross-time-scale optimization framework, a refined DER model is produced. This model significantly enhances the flexibility of the system by coordinating the fast response capabilities of EVs with long-cycle energy storage characteristics, thereby reducing the risk of power imbalance errors within the system.

In addition, to effectively solve the proposed complex VPP-EV model, an optimization framework based on TD3 is designed. The framework achieves efficient data preprocessing through the integration of DQN-mRMR and GA-TCN-LSTM, followed by performing dynamic strategy optimization using the TD3 algorithm. This framework not only handles multi-objective problems in high-dimensional nonlinear state spaces but also effectively mitigates overestimation issues through the truncation of twin Q-learning and a smoothing mechanism. The results show that the developed framework provides a robust and scalable solution model for the sophisticated deployment of VPPs. Moreover, the results of this study are as follows.

1) The participation of EVs reduces the induced operational cost by an average of 39.1% compared with that of the traditional BESS framework. Additionally, through P2G carbon absorption and flexible EV dispatching, negative carbon emissions are achieved in multiple cases, with the carbon emissions reduced by 133.5% compared with those of the BESS. Furthermore, the EV scenario generates additional revenue through carbon trading, with multiday carbon trade costs that are lower than those of the BESS scenario, thus validating the dual advantages of the model in terms of both economic performance and environmental sustainability.

2) In addition, the proposed DQN-mRMR feature extraction model significantly improves the accuracy of the GA-TCN-LSTM demand forecasting method. The RMSE decreases by 63.0%, the MAE decreases by 81.2%, and the R^2 increases by 7.0%. The feature selection accuracy is 98.2%. This model efficiently analyzes high-dimensional nonlinear features, providing reliable data support for the subsequent optimization stage.

3) The proposed TD3 algorithm effectively reduces the dependence on fossil energy, lowering the power output of the MT to 1.163 MWh (a 99.69% reduction compared with that of DDPG). It also optimizes the EV charging and discharging strategy, achieving 70.771 MWh of energy interaction, while ensuring that the battery health degradation rate is controlled at 0.641% (a 63.3% reduction compared with that of DDPG). The standard deviation of the health fluctuations is only 0.000 32 (an 89% reduction relative to the results of DDPG). Furthermore, the total cost of TD3 is lower than those of Gurobi, DDPG, and MPC (by 41.46%, 57.42%, and 2.13%, respectively). These results demonstrate its comprehensive ability to optimize economic performance, environmental sustainability, and EV lifespans in a coordinated manner.

To conclude, the proposed VPP-EV collaborative scheduling model and TD3 optimization algorithm demonstrate significant advantages in terms of both economic and carbon minimization performance. This method effectively reduces the operational cost of the system and decreases its carbon emissions while meeting the imposed demand. It also protects the health of the EVs.

This paper focuses on model construction and simulation verification processes. Future work will explore the VPP framework to develop a novel multi-timescale customer load directrix-based demand response method [31]–[36].

ACKNOWLEDGMENT

Not applicable.

AUTHORS' CONTRIBUTIONS

Wei Hu: formal analysis, methodology, software, and writing-original draft. Shuo Wang: formal analysis, writing-original draft, and writing-review & editing. Puliang Du: conceptualization and methodology. All authors read and approved the final manuscript.

FUNDING

This work is supported by the National Nature Science Foundation of China (No. 72401182).

AVAILABILITY OF DATA AND MATERIALS

Please contact the corresponding author for data material request.

DECLARATIONS

Competing interests: 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 article.

AUTHORS' INFORMATION

Wei Hu received the Ph.D. degree in management science and engineering from Shanghai Jiaotong University, Shanghai, China. He is now a professor and the deputy dean of the School of Economics and Management at Shanghai University of Electric Power. His main research interests include artificial intelligence, smart energy, power engineering management, energy and power optimization and decision making.

Shuo Wang is currently pursuing his master's degree at Shanghai University of Electric Power. His main research interests include power engineering management and smart energy.

Puliang Du received the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China. He is currently teaching at the School of Economics and Management, Shanghai University of Electric Power. His main research interests are the application of decision theory and methods in power systems, the optimal operation of distribution networks, and artificial intelligence.

REFERENCES

[1] S. Hu, Y. Chen, and J. Feng, "A flexible interactive coordination control method of commercial virtual power plant based on WCVAR," *International Journal of Electrical Power & Energy Systems*, vol. 160, pp. 1-14, Sep. 2024.

[2] X. Guo, L. Wang, and D. Ren, "Optimal scheduling model for virtual power plant combining carbon trading and green certificate trading," *Energy*, vol. 318, pp. 1-13, Mar. 2025.

[3] Q. Li, Y. Zhou, and F. Wei *et al.*, "Multi-time scale scheduling for virtual power plants: integrating the flexibility of power generation and multi-user loads while considering the capacity degradation of energy storage systems," *Applied Energy*, vol. 362, pp. 1-18, Mar. 2024.

[4] J. Li, Z. Sun, and X. Niu *et al.*, "Economic optimization scheduling of virtual power plants considering an incentive based tiered carbon price," *Energy*, vol. 305, pp. 1-15, Jun. 2024.

[5] Q. Li, F. Dong, and G. Zhou *et al.*, "Co-optimization of virtual power plants and distribution grids: emphasizing flexible resource aggregation and battery capacity degradation," *Applied Energy*, vol. 377, pp. 1-14, Sep. 2025.

[6] E. Akbari, A. F. Naghibi, and M. Veisi *et al.*, "Multi-objective economic operation of smart distribution network with renewable-flexible virtual power plants considering voltage security index," *Scientific Reports*, vol. 14, no. 1, pp. 1-18, Aug. 2024.

[7] P. Yong, Z. Yang, and N. Zhang *et al.*, "On the self-scheduling of cellular base station-based virtual power plants," *IEEE Internet of Things Journal*, vol. 11, no. 14, pp. 24511-24522, Jul. 2024.

[8] M. Bahloul, L. Breathnach, and S. Khadem, "Residential virtual power plant control: a novel hierarchical multi-objective optimization approach," *IEEE Transactions on Smart Grid*, vol. 16, no. 2, pp. 1301-1313, Mar. 2025.

[9] T. C. Ou, H. Tieng, and T. H. Tsai *et al.*, "Design of green power clouds for intelligent virtual power plants," *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 18051-18062, Jun. 2025.

[10] T. Wang, Z. Huang, and R. Ying *et al.*, "A low-carbon operation optimization method of ETG-RIES based on adaptive optimization spiking neural P systems," *Protection and Control of Modern Power Systems*, vol. 9, no. 5, pp. 162-177, Sep. 2024.

[11] Y. Shang, X. Li, and T. Xu *et al.*, "Uncertainty-output virtual power plant participates in multi-electricity market considering the improved Shapley value distribution method," *International Journal of Electrical Power & Energy Systems*, vol. 165, pp. 1-17, Jan. 2025.

[12] R. Chen, D. Song, and L. Liao *et al.*, "Feedback correction scheduling strategy for electric vehicles based on multi-regional agent master-slave and evolutionary hybrid game," *Energy*, vol. 319, pp. 1-17, Feb. 2025.

[13] Y. Wan, Y. Wang, and H. Yu *et al.*, "Smart mobile power bank: a novel grid-friendly mobile microgrid for power grid with high penetration of renewable vehi-

cles," *IEEE Transactions on Transportation Electrification*, vol. 11, no. 1, pp. 2287-2298, Feb. 2025.
[14] U. Qureshi, I. Andrabi, and M. Manzoor *et al.*, "Optimizing electric vehicle integration in virtual power plants: a stochastic optimization framework with MDNN integration," *IEEE Transactions on Industry Applications*, vol. 60, no. 6, pp. 9227-9236, Dec. 2024.
[15] J. H. Kim, J. S. Hwang, and Y. S. Kim, "An IGDT-WDRCC based optimal bidding strategy of VPP aggregators in new energy market considering multiple uncertainties," *Energy*, vol. 313, pp. 1-13, Nov. 2024.
[16] T. Newbolt, P. Mandal, and H. Wang, "Sustainability of dynamic wireless power transfer roadway for in-motion electric vehicle charging," *IEEE Transactions on Transportation Electrification*, vol. 10, no. 1, pp. 1347-1362, Mar. 2024.
[17] K. V. Sastry, D. G. Taylor, and M. J. Leamy, "Decentralized smart charging of electric vehicles in residential settings: algorithms and predicted grid impact," *IEEE Transactions on Smart Grid*, vol. 15, no. 2, pp. 1926-1938, Mar. 2024.
[18] X. Lin, X. Lin, and W. Zhong *et al.*, "Predictive operation optimization of multi-energy virtual power plant considering behavior uncertainty of diverse stakeholders," *Energy*, vol. 280, pp. 1-15, Jun. 2023.
[19] C. Huang, K. Li, and N. Zhang, "Strategic joint bidding and pricing of load aggregators in day-ahead demand response market," *Applied Energy*, vol. 377, pp. 1-10, Sep. 2025.
[20] J. Zhang, Z. Liu, and Y. Liu, "A scheduling optimization model for a gas-electricity interconnected virtual power plant considering green certificates-carbon joint trading and source-load uncertainties," *Energy*, vol. 315, pp. 1-21, Jan. 2025.
[21] Y. Shang, Y. Shang, and H. Yu, "Achieving efficient and adaptable dispatching for vehicle-to-grid using distributed edge computing and attention-based LSTM," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 10, pp. 6915-6926, Oct. 2021.
[22] Q. Zhang, X. Cai, and Y. Zhong, "Dynamic response prediction of high-speed train on cable-stayed bridge based on genetic algorithm and fused neural networks," *Engineering Structures*, vol. 306, pp. 1-16, Mar. 2024.
[23] Y. Wang, M. Mao, and L. Chang, "Multi-time scales prediction of aggregated schedulable capacity of electric vehicle fleets based on enhanced Prophet-LGBM algorithm," *Applied Energy*, vol. 374, pp. 1-13, Aug. 2024.
[24] H. Meng, H. Jia, and T. Xu *et al.*, "Internal pricing driven dynamic aggregation of virtual power plant with energy storage systems," *Energy*, vol. 321, pp. 1-17, Mar. 2025.
[25] W. M. Nkounga, M. F. Ndiaye, and O. Cisse *et al.*, "Automatic control and dispatching of charging currents to a charging station for power-assisted bikes," *Energy*, vol. 246, pp. 1-22, May 2025.
[26] M. Du and X. Liu, "A cyber-secured preventive dispatch for power systems with high wind penetration," *IEEE Transactions on Power Systems*, vol. 39, no. 2, pp. 2398-2409, Mar. 2024.
[27] K. Pang, C. Wang, and N. D. Hatziargyriou *et al.*, "Dynamic restoration of active distribution networks by coordinated repair crew dispatch and cold load pickup," *IEEE Transactions on Power Systems*, vol. 39, no. 2, pp. 4699-4713, Mar. 2024.
[28] C. Zheng, H. Sang, and L. Xing *et al.*, "A self-adaptive memetic algorithm with Q-learning for solving the multi-AGVs dispatching problem," *Swarm and Evolutionary Computation*, vol. 90, pp. 1-20, Aug. 2024.
[29] X. Zhang, K. Xiong, and W. Chen *et al.*, "Minimizing AoI in high-speed railway mobile networks: DQN-based methods," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 12, pp. 20137-20150, Dec. 2024.
[30] Y. Wang, J. Wu, and H. Sun *et al.*, "Promoting collaborative dispatching in the ride-sourcing market with a third-party integrator," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 7, pp. 6889-6901, Jul. 2024.
[31] Y. Zhang, Y. Meng, and S. Fan *et al.*, "Multi-time scale customer directrix load-based demand response under renewable energy and customer uncertainties," *Applied Energy*, vol. 383, pp. 1-16, Apr. 2025.
[32] J. Elio and R. J. Milcarek, "Assessment of optimal energy storage dispatch control strategies for cost savings in 606 commercial and industrial facilities leveraging utility demand response programs," *Applied Energy*, vol. 384, pp. 1-15, Apr. 2025.
[33] W. Li, Y. Zou, and H. Yang *et al.*, "Two-stage stochastic energy scheduling for multi-energy rural microgrids with irrigation systems and biomass fermentation," *IEEE Transactions on Smart Grid*, vol. 16, no. 2, pp. 1075-1087, Mar. 2025.
[34] R. Sepehrzad, A. S. G. Langeroudi, and A. Al-Durra *et al.*, "Demand response-based multi-layer peer-to-peer energy trading strategy for renewable-powered microgrids with electric vehicles," *Energy*, vol. 320, pp. 1-18, Apr. 2025.
[35] X. Liu, W. Sun, and T. Chen *et al.*, "Energy and environmental performance of iron and steel industry in real-time demand response: a case of hot rolling process," *Applied Energy*, vol. 389, pp. 1-11, Mar. 2025.
[36] L. Li, S. Fan, and J. Xiao *et al.*, "Energy management strategy for community prosumers aggregated VPP participation in the ancillary services market based on P2P trading," *Applied Energy*, vol. 384, pp. 1-19, Apr. 2025.