World Electric Vehicle Journal

Article

Optimization of Electric Vehicle Charging and Discharging Strategies Considering Battery Health State: A Safe Reinforcement Learning Approach

Shuifu Gu [1], Kejun Qian [1] and Yongbiao Yang [2,*]

[1] State Grid Jiangsu Electric Power Co., Ltd.—Suzhou Power Supply Company, Suzhou 215031, China; seuepri@126.com (S.G.)

[2] School of Electrical Engineering, Southeast University, Nanjing 210096, China

* Correspondence: yyb_seu2025@163.com; Tel.: +86-188-6212-3469

Abstract: With the widespread adoption of electric vehicles (EVs), optimizing their charging and discharging strategies to improve energy efficiency and extend battery life has become a focal point of current research. Traditional charging and discharging strategies often fail to adequately consider the battery's state of health (SOH), resulting in accelerated battery aging and decreased efficiency. In response, this paper proposes a safe reinforcement learning–based optimization method for EV charging and discharging strategies, aimed at minimizing charging and discharging costs while accounting for battery SOH. First, a novel battery health status prediction model based on physics-informed hybrid neural networks (PHNN) is designed. Then, the EV charging and discharging decision-making problem, considering battery health status, is formulated as a constrained Markov decision process, and an interior-point policy optimization (IPO) algorithm based on long short-term memory (LSTM) neural networks is proposed to solve it. The algorithm filters out strategies that violate constraints by introducing a logarithmic barrier function. Finally, the experimental results demonstrate that the proposed method significantly enhances battery life while maintaining maximum economic benefits during the EV charging and discharging process. This research provides a novel solution for intelligent and personalized charging strategies for EVs, which is of great significance for promoting the sustainable development of new energy vehicles.

Keywords: electric vehicle; safe reinforcement learning; battery state of health; physics-informed hybrid neural networks; interior-point policy optimization

check for updates

Academic Editor: Joeri Van Mierlo

Received: 16 February 2025
Revised: 4 May 2025
Accepted: 15 May 2025
Published: 20 May 2025

Citation: Gu, S.; Qian, K.; Yang, Y. Optimization of Electric Vehicle Charging and Discharging Strategies Considering Battery Health State: A Safe Reinforcement Learning Approach. *World Electr. Veh. J.* **2025**, *16*, 286. https://doi.org/10.3390/wevj16050286

Correction Statement: This article has been republished with a minor change. The change does not affect the scientific content of the article and further details are available within the backmatter of the website version of this article.

1. Introduction

Electric vehicles are regarded as a vital component in the construction of a new energy system due to their energy-saving and environmental protection advantages, and they are increasingly becoming the primary choice for transportation. According to statistics from the International Energy Agency, global EV sales reached a record high of 14 million units in 2023 [1]. However, the uncoordinated charging load of large-scale EVs, when superimposed on the basic load of the power grid, can easily lead to peak electricity consumption, increasing the burden on the power system and causing problems such as a decline in power quality [2].

Vehicle-to-Grid (V2G) and Grid-to-Vehicle (G2V) technologies have been regarded as effective means to enhance the stability of power distribution networks [3]. Optimization of EV charging and discharging strategies is a typical stochastic optimization problem, with uncertainties arising from factors such as user travel behaviors, dynamic electricity prices, and battery health status [4]. To address these uncertainties, ref. [5] proposed a scenario-based two-stage stochastic linear programming model for scheduling EV charging under varying grid demands. In [6], a robust optimization model was developed to balance cost efficiency and fast-charging requirements by introducing L2-norm uncertainty to improve robustness against cost fluctuations. To further mitigate the negative impact of EV loads on distribution networks while satisfying user demand, ref. [7] introduced a spatiotemporal-aware scheduling strategy that considers uncertainties in V2G operations. Ref. [8] proposed an optimal charging/discharging strategy that incorporates voltage stability constraints under V2G scenarios. Additionally, ref. [9] constructed a two-stage day-ahead and real-time multi-objective optimization framework to reconstruct travel chains under multi-modal travel scenarios. Reference [10] proposed a model predictive control (MPC)-based energy management method, applying constraints on battery state of charge (SOC), power, and thermal conditions. Moreover, ref. [11] explored the modeling of dynamic user preferences in charging parameters, using a stochastic game–theoretic approach to address uncertainties. A generalized Nash equilibrium approach was further adopted in [12] to optimize EV charging strategies, considering temporal uncertainties. However, these model-based optimization methods heavily rely on accurate system modeling, which is often impractical due to the complexity and uncertainties of real-world environments. Furthermore, for high-dimensional or nonlinear problems, such methods are computationally intensive and prone to convergence to local optima, making it difficult to achieve globally optimal solutions.

As a model-free approach, deep reinforcement learning (DRL) enables agents to learn optimal strategies through interactions with the environment without requiring pre-modeled system knowledge. DRL has been widely applied to sequential decision-making problems. In [13], a hybrid DRL model was integrated into a battery management system framework to manage EV lithium-ion battery charging and discharging processes. Ref. [14] established an EV adjustability identification model based on user characteristics and applied DRL for real-time scheduling of charging and swapping loads. A real-time scheduling strategy based on the baseline Monte Carlo policy gradient method was proposed in [15] to optimize swap station operation. Reference [16] utilized LSTM networks to predict electricity price trends and applied the predicted prices to the Deep Deterministic Policy Gradient (DDPG) algorithm. In [17], driver anxiety was mathematically modeled using statistical principles, and the Soft Actor-Critic (SAC) algorithm was employed to learn optimal charging strategies. Although DRL-based methods have demonstrated good performance in EV charging optimization, operational safety constraints such as voltage and current limits are crucial in practice. These constraints are typically enforced through penalty terms in the reward function. However, setting appropriate penalty coefficients often relies on expert knowledge and tedious manual tuning, and improper parameter choices may lead to significant performance degradation. To overcome these limitations, safe DRL methods have been developed. For example, ref. [18] proposed a constrained policy optimization (CPO) approach for EV charging management, while [19] introduced a primal-dual RL framework that adaptively adjusts Lagrangian multipliers to manage constraints more effectively.

Furthermore, most current EV charging optimization methods primarily focus on maximizing users' economic benefits, while giving limited consideration to battery degradation costs. Frequent charging and discharging cycles can significantly shorten battery lifespan, leading to increased hidden costs for users [20]. Thus, simultaneously optimizing both charging profits and battery health loss has become an emerging research focus. Accurate evaluation of battery degradation costs hinges critically on precise prediction of battery SOH. Accordingly, developing high-accuracy SOH prediction models has become a necessary prerequisite for integrating battery degradation cost into EV charging optimization strategies.

Existing SOH prediction methods can be broadly categorized into model-based and data-driven approaches. Model-based methods rely on physical or electrochemical modeling of battery behaviors, such as equivalent circuit models [21] and electrochemical models [22]. Ref. [23] developed a degradation-aware electrochemical thermal model for SOH estimation, while [24] combined electrochemical impedance spectroscopy features with an extreme learning machine to predict SOH. On the other hand, data-driven methods leverage historical operational data to learn complex feature-performance relationships without requiring explicit physical modeling. Among them, LSTM networks have been widely used for their outstanding performance in long-sequence forecasting. For instance, ref. [25] proposed a hybrid approach combining improved convolutional neural networks with LSTM for battery remaining useful life prediction. Ref. [26] developed a transferable multi-stage LSTM model for cross-battery SOH estimation, and [27] enhanced bidirectional LSTM models with attention mechanisms to improve accuracy and adaptability under temperature variations. However, existing studies have rarely integrated battery SOH prediction effectively into EV charging optimization. Only a few works attempt to incorporate simplified degradation models. For example, ref. [28] combined particle swarm optimization (PSO) and MPC to jointly optimize charging strategies while considering battery aging, but the SOH modeling accuracy remains limited. Similarly, ref. [29] proposed a joint scheduling and charging optimization method for electric buses that simultaneously considers explicit charging costs and implicit battery degradation costs.

To address the limitations of existing DRL methods in modeling battery health status, this paper proposes a physics-informed hybrid neural network (PHNN), which integrates physical degradation mechanisms with a LSTM structure. The PHNN enhances the model's ability to capture complex degradation processes while maintaining high prediction accuracy, thereby providing more effective guidance for optimizing EV charging and discharging strategies. Compared to existing works, the main contributions of this paper are summarized as follows:

(1) A novel battery health state prediction model based on a PHNN is designed, which combines physical prior knowledge with the LSTM's capability to capture temporal features, significantly improving the accuracy of SOH predictions.

(2) Considering the battery health state, the EV charging and discharging decision-making problem is formulated as a constrained Markov decision process. A reward function and a cost function that incorporate battery health state are designed, precisely quantifying the cost of battery degradation in V2G operations.

(3) An LSTM-based IPO algorithm is proposed for solving the problem, using a logarithmic barrier function as an extension to the cost function to handle constraints, which effectively reduces computational costs and enhances policy performance.

(4) Based on public datasets, several simulation experiments are designed to validate the algorithm's effectiveness. The experimental results show that the proposed EV charging and discharging strategy optimization method can significantly mitigate the decline in battery health during the V2G process and maximize the charging and discharging benefits for EV owners, demonstrating strong practical significance.

2. Optimization Model of EV Charging and Discharging Strategies Considering Battery Health Constraints, and Its Constrained Markov Decision Process Formulation

2.1. Mathematical Model of EV Charging and Discharging Strategies Considering Battery Health Constraints

To minimize battery degradation and reduce the stress imposed on the battery, this study adopts a slow charging strategy for EVs. It is assumed that EVs begin charging or discharging immediately upon arrival and connection to the grid, and that the charging/discharging behavior is implemented at a constant power level within each decision interval. The specific control model is as follows:

$$\begin{cases} SOC_{t+1} = \begin{cases} SOC_t + p_t \times \eta_{\text{ch}} \times \Delta t, p_t > 0 \\ SOC_t + \frac{p_t \times \Delta t}{\eta_{\text{dis}}}, p_t < 0 \end{cases} & , t_a < t \leq t_d - \Delta t \\ SOC_t = SOC_0 & , t = t_a \\ -p_{\text{dis}}^{\max} \leq p_t \leq p_{\text{ch}}^{\max} \end{cases} \tag{1}$$

In the formula, t_a and t_d represent the arrival and departure times of the EV, respectively, with both arrival and departure occurring at the beginning of the time step; SOC_0 is the remaining electricity of the EV at the start of charging or discharging; SOC_t is the SOC value at time t, η_{ch} and η_{dis} are the energy conversion efficiencies of the EV battery during charging and discharging, respectively; $p_{\text{ch}}^{\max}$ and $p_{\text{dis}}^{\max}$ are the maximum charging and discharging power, respectively; and p_t is the EV charging and discharging power.

This paper aims to minimize the charging cost as the objective function and constructs an optimization model with the constraints of battery energy limits and user travel electricity demand. The specific formulation is as follows:

$$\begin{aligned} & \min \textstyle\sum_{t=t_a}^{t_d} p_t \cdot \xi_t \\ & s.t. E_{\min} \leq E_t + \eta p_t \leq E_{\max} \\ & \left| E_{\text{target}} - E_{t_d} \right| \leq d \end{aligned} \tag{2}$$

In the formula, ξ_t represents the dynamic electricity price at that moment. The product being greater than zero indicates the cost of the EV user purchasing electricity at that time, and, vice versa, represents the revenue from selling electricity. $E_{\min}$ and $E_{\max}$ represent the minimum and maximum capacity of the EV battery, respectively. E_{target} represents the user's travel electricity demand target. d is a small constraint tolerance used to limit the deviation between the EV's electricity level and the charging target.

For EV batteries, there are typically two important parameters that describe their current state: the SOC and the SOH. Among these, SOH measures the ratio between the actual capacity of the battery and its initial capacity, used to represent the degree of battery aging and performance degradation. To consider the cost overhead due to battery degradation in EV charging and discharging, this paper introduces a new parameter: the Initial State of Charge (ISOC), which represents the ratio between the current stored electricity in the battery and its initial capacity, expressed as a percentage. Specifically, the ISOC at time t is defined as follows:

$$ISOC_t = SOC_t \cdot SOH_t \tag{3}$$

Once connected to a charging pile, the EV can charge within a unit of time to meet charging needs or discharge during periods of high electricity prices to gain benefits from price fluctuations. Assuming the EV connects to the charging pile at time t_a, and the user plans to depart at time t_d, the change in battery charge within a unit of time is as follows:

$$ISOC_{t+1} = \begin{cases} ISOC_t & t < t_a, t \geq t_d \\ ISOC_t + \eta \cdot a_t & t_a \leq t < t_d \end{cases} \tag{4}$$

where a_t represents the rate at which the EV charges or discharges within a unit of time.

2.2. Modeling of EV Charging and Discharging Strategy Optimization Based on CMDP

This section further constructs the EV charging and discharging decision optimization problem into a constrained Markov decision process (CMDP) model. As a mathematical framework, CMDP offers an effective means to describe the randomness and uncertainty associated with EV charging and discharging behaviors. This model permits the making of charging and discharging decisions based on electricity prices and battery health states at different time intervals, aiming to optimize the economic benefits for users. To depict the continuous charging and discharging process of the EV, the continuous state transitions are divided into sufficiently small time slots. The subsequent s_t represents the state of the vehicle within the time slot t, as detailed below.

(1) State Space: In this paper, the states obtained from the environment are used as inputs for the charging and discharging control strategy to generate charging and discharging rates. Specifically, within time period t, the state s_t is defined as follows:

$$s_t = (\xi_{t-23}, \xi_{t-22}, \cdots, \xi_t, t_d, ISOC_t, ISOC_d) \tag{5}$$

where $\xi_{t-23}, \xi_{t-22}, \ldots, \xi_t$ represents the electricity pricing from the power company over the past 24 time slots and $ISOC_t$ denotes the amount of electricity in the battery at time t.

(2) Action Space: During the vehicle's charging and discharging process, within the time interval [t, t + 1), the vehicle's charging and discharging behavior is represented by the action a_t, where $a_t = p_t$.

(3) Reward Function: Considering both the charging cost and the discharge degradation cost during the EV charging and discharging process, the reward function is defined as follows:

$$r_t = R(s_t, a_t, s_{t+1}) = -\sigma_1 \times a_t \times \xi_t - \sigma_2 \times c \times (SOH_t - SOH_{t+1}) \tag{6}$$

where $a_t \times \xi_t$ represents the revenue from charging and discharging, $SOH_t - SOH_{t+1}$ denotes the loss of battery capacity caused by the current charging and discharging action in the time period, and c represents the cost for the user to replace the battery. The second term of the reward mentioned above represents the economic loss caused by the battery degradation per unit of time; σ_1 and σ_2 are the reward weights.

(4) Constraint Function: To ensure both the safety of the EV battery during the charging/discharging process and the fulfillment of user travel energy requirements, this study formulates the optimization problem as a CMDP. A set of auxiliary constraint functions is defined as: $C : S \times A \times S \to R^C, C = (c_t^1, c_t^2)$, where c_t^1 and c_t^2 represent the deviation between the final state of charge and the target value, and the safety constraint based on the physical limits of the battery, respectively. The specific forms are given by:

$$\begin{aligned} c_t^1 &= \left| ISOC_t - ISOC_{\text{target}} \right|, if\ t = t_d \\ c_t^2 &= \begin{cases} ISOC_t - ISOC_{\max}, if\ ISOC_t > ISOC_{\max}, t < t_d \\ ISOC_{\min} - ISOC_t, if\ ISOC_t < ISOC_{\min}, t < t_d \end{cases} \end{aligned} \tag{7}$$

The constraint function c_t^1 penalizes the deviation of the final ISOC from the targe $ISOC_{\text{target}}$ at the departure time t_d, reflecting the user's expected residual energy or travel mileage. The constraint c_t^2 incorporates practical physical constraints and battery safety considerations and penalizes situations where the SOC exceeds the allowable range $[ISOC_{\min}, ISOC_{\max}]$, thereby ensuring safe operation of the battery.

Based on the above constraint functions, we construct a constraint-aware objective function that captures both travel demand and safety requirements: $J_C(\pi) = E_{\tau\sim\pi}\left[\sum_{t=0}^{T-1}\gamma^t c_t\right]$, where γ is the discount factor that penalizes constraint violations at different time steps. By combining this with the EV charging control conditions, the set of feasible policies can be defined as:

$$\prod\nolimits_C = \left\{\pi \middle| J_{c_t^1}(\pi) \le d_1, c_t^2(s_t, a_t) \le d_2, \forall t\right\} \tag{8}$$

In Equation (8), d_1 and d_2 are acceptable thresholds for the two constraint functions. A policy π is considered feasible if it ensures that the final ISOC meets the user's target while maintaining the SOC within the safety range throughout the charging process. The optimization problem of EV charging and discharging can then be formulated as a CMDP:

$$\begin{gathered} \max_{\pi\in\prod} J_R(\pi) = E_{\tau\sim\pi}\left[\sum_{t=0}^{T-1}\gamma^t r_t\right] \\ s.t.\ J_{C^i}(\pi) \le d_i \end{gathered} \tag{9}$$

where $J_R(\pi)$ represents the expected cumulative reward under policy π, and r_t is the instantaneous reward at time step t. The constraint term $J_{C^i}(\pi)$ is incorporated to ensure safety and reliability. Finally, the optimal policy that maximizes the cumulative reward while satisfying all constraints is given by:

$$\pi^* = \arg\max_{\pi\in\prod_C} J(\pi) \tag{10}$$

Through this CMDP-based formulation, the EV's ISOC target and safe operational range are unified into a decision-making framework, which ensures the satisfaction of user travel needs while maintaining battery operation within a safe and acceptable range. This approach enables safe and reliable decision-making in uncertain real-world environments.

(5) Transition Probability: The transition probability $P(s|s', a)$ is influenced by the charging and discharging power and the material characteristics of the EV's battery. To align with real-world scenarios, the actual state update process of the EV's SOC after accounting for energy losses during charging and discharging is characterized by the product of the charging and discharging power and the energy conversion efficiency in the EV charging and discharging control model.

3. Prediction Model of Battery SOH Degradation Based on PHNN

3.1. Principle of LSTM Temporal Sensitivity Enhancement

LSTM protects and controls the cell state through a "gate" mechanism to selectively pass memory information, including the forget gate, input gate, and output gate. Each gate control device is composed of activation functions and pointwise multiplication functions. Among them, the forget gate determines the timing of low-value information to be discarded, the input gate determines the effective information at the current moment and loads it into the memory cell through sigmoid and tanh activation functions, and the output gate decides the output feature information y_t based on the current input x_t and the memory information c_t. The specific calculation methods for each gate of LSTM are as follows:

$$\begin{aligned} &O_{forget} = \sigma\left(W_f[y_{t-1}, x_t] + r_f\right) \\ &O_{input} = \sigma(W_i[y_{t-1}, x_t] + r_i) * tanh(W_{i'}[y_{t-1}, x_t] + r_{i'}) \\ &O_{out} = \sigma(W_o[y_{t-1}, x_t] + r_o) \end{aligned} \tag{11}$$

In the above formula, $W_f, W_i, W_{i'}, W_0$ represent the weight coefficients of the forget gate, input gate, and output gate, respectively. Since the input gate O_{input} has both sigmoid and tanh activation functions, it has two weight coefficients; similarly, $r_f, r_i, r_{i'}, r_o$ are the biases for the three gates, respectively. Through the aforementioned three gate control structures, LSTM achieves spanning of multiple time steps of temporal information. Therefore, the LSTM neural network can perform feature extraction on historical data, which helps to mine the future trends of historical time series data and enhances the ability to cope with uncertain factors. This paper, with the help of LSTM's temporal feature extraction capability, enhances the perception of historical SOH and electricity prices and then reasonably infers future trends from historical SOH and electricity price data.

3.2. Battery SOH Degradation Prediction Based on PHNN

The proposed PHNN enhances the accuracy and robustness of SOH degradation prediction by combining model-based methods with neural networks and integrating physical and chemical prior knowledge. In our approach, the SOH of a lithium-ion battery is characterized by two variables: the non-forced dynamic Dt before the discharge cycle t (which represents the inherent trend of the system to change without external interference) and the estimated capacity loss Lt that occurs during the discharge cycle. At the end of discharge cycle t, the estimated SOH can be calculated as follows:

$$SOH_t = D_t - L_t \tag{12}$$

According to existing studies, one of the primary chemical mechanisms of capacity degradation is progressive growth of the solid electrolyte interphase (SEI) layer on the surface of the anode particles. This process intensifies with charge-discharge cycling and has been widely adopted for modeling the trend of battery aging. Given that the Arrhenius equation effectively characterizes the relationship between reaction rate and temperature [30], this study incorporates temperature as a key factor when estimating capacity loss based on the SEI growth mechanism. It is important to note, however, that while SEI growth represents a dominant degradation pathway, it does not account for all mechanisms—such as lithium plating, electrode cracking, and electrolyte decomposition. Therefore, we introduce a neural network-based compensation component into the model to learn from these latent or unobservable degradation processes, thereby enhancing the accuracy and generalization capability of capacity loss estimation. Therefore, the cumulative capacity loss can be estimated using the following formula:

$$\omega = A \cdot e^{\frac{-E_a}{RT}} \cdot t^z \tag{13}$$

In the formula, ω represents the cumulative capacity loss, A is a constant, Ea represents the activation energy of the reaction, R is the universal gas constant, T represents the operating temperature of the lithium battery, and z represents the power exponent coefficient.

In single-step prediction, the capacity loss Lt generated during the t-th discharge period can be calculated as follows [30]:

$$L_t = \omega_t - \omega_{t-1} \approx A^{1/z} \cdot e^{\frac{-E_a}{RT}} \cdot z \cdot \omega_{t-1}^{\frac{z-1}{z}} \tag{14}$$

In this equation, z is used to describe the power-law relationship between capacity degradation and cycle number or time. A is a constant related to the characteristics of the electrochemical reaction system. The above degradation model only considers the effects of temperature T and cycle number t on battery capacity.

However, at high discharge rates, the battery capacity fade is more rapid than at standard discharge rates, which may further affect the accuracy of capacity loss estimation [31]. Therefore, we introduce an additional neural network $M_\rho(U_t, I_t)$ composed of fully connected layers to estimate the loss due to the discharge rate effect, where ρ represents the neural network parameters, U_t represents the average voltage for the entire cycle, and I_t represents the average current. Since the network is trained end-to-end, the parameters ρ can be learned simultaneously with the other parts of the network. Then, Equation (14) can be further expanded as follows:

$$L_t = k_1 \cdot e^{\frac{k_2}{T}} \cdot \omega_{t-1}^{k_3} \cdot M_\rho(U_t, I_t) \tag{15}$$

In the formula, $k_1 = zA^{1/z}, k_2 = -E_a/Rz, k_3 = z - 1/z$, ω_{t-1} denotes the capacity degradation factor estimated in the previous cycle, reflecting the influence of historical cumulative degradation on the current prediction, U_t represents the average voltage for the entire cycle, and I_t represents the average current. Based on the above, we have constructed a neural network with a special structure, where Exponential Linear Unit (ELU) nodes are used for exponential calculations, and the remaining nodes represent the three factors proposed, denoting the corresponding coefficients. By leveraging the inherent learning and parameter updating capabilities of neural networks, the model can autonomously learn appropriate parameters based on the characteristics of the current battery. Therefore, guided by prior knowledge, the neural network can calculate Lt based on physical priors according to Equation (15). In addition to this, we use an LSTM network to capture the temporal information of battery health, with its input being an SOH sequence SOH_{seq}, representing the battery capacity values over several past cycles, hence $D_t = LSTM(SOH_{\text{seq}})$.

4. Constrained EV Charging and Discharging Strategy Solution Based on the IPO Algorithm

This chapter first introduces the principles of the IPO algorithm and the LSTM network. Then, it presents the optimization procedure for the EV charging and discharging strategy based on IPO and PHNN. Figure 1 shows the optimization framework for EV charging and discharging strategies based on IPO and PHNN.

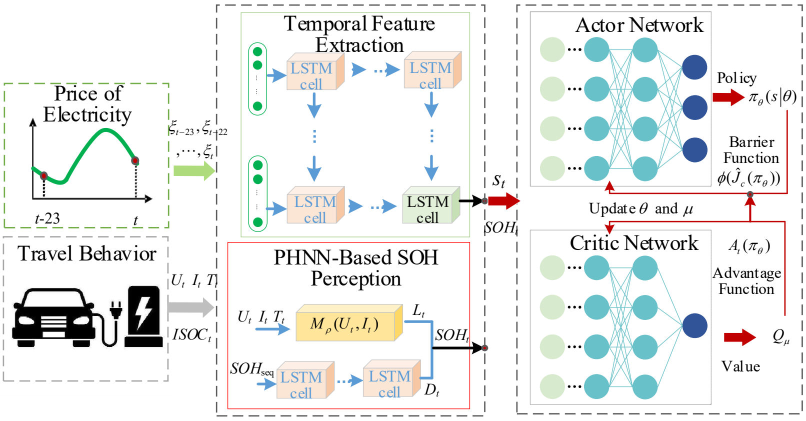

Figure 1. Optimization framework for EV charging and discharging strategies based on IPO and PHNN.

4.1. Principle of IPO Algorithm and LSTM Temporal Data Feature Extraction

This section further proposes the IPO algorithm to solve the CMDP problem. The IPO algorithm inherits the clipping objective from the Proximal Policy Optimization (PPO) algorithm and uses a logarithmic barrier function to quantify behavior that violates the constraints. The ultimate objective function of the PPO algorithm is given by the following equation:

$$\underset{\theta}{max}\, L^{\text{clip}}(\theta) = \underset{\substack{s \sim \rho_{\pi_\theta} \\ a \sim \pi_\theta}}{E} \left[\min\{r(\theta), \text{clip}(r(\theta), 1-\varepsilon, 1+\varepsilon)\}\hat{A}_t\right] \tag{16}$$

where $r(\theta) = \pi_\theta(a_t|s_t)/\pi_{\theta_{old}}(a_t|s_t)$ and $\text{clip}(\cdot)$ are clipping functions that constrain θ within a specified interval. The PPO algorithm regulates the extent of policy updates through this clipping function. Specifically, when the estimated value of the advantage function $\hat{A}_t > 0$, it indicates that the action chosen based on the current policy $\pi_\theta(a_t|s_t)$ is relatively optimal; hence, the policy's probability distribution needs to be adjusted to increase the sampling probability of this action, thereby optimizing the policy. Conversely, when the estimated value of the advantage function is negative, it suggests that the mapping from state to action under the current policy is unsatisfactory, necessitating an improvement in the policy to decrease the probability of selecting this action.

In dealing with CMDP issues, IPO has two significant advantages: (1) by adding logarithmic barrier functions to expand constraints, when handling complex and diverse safety constraints in real-world application scenarios, it is sufficient to simply add more logarithmic barrier functions to set reasonable safety constraints; (2) by using logarithmic barrier functions, the constrained optimization problem is transformed into an unconstrained first-order optimization problem, which reduces the computational load and improves the convergence speed of training. The objective function of the Interior Point Policy Optimization algorithm is shown in Equation (17):

$$\begin{aligned} &\max_{\theta} L^{\text{clip}}(\theta) \\ &s.t.\ \hat{J}_{C^i}^{\pi_\theta} \leq 0 \end{aligned} \tag{17}$$

In the equation, $\hat{J}_{C^i}^{\pi_\theta} = J_{C^i}^{\pi_\theta} - d_i$. For each constraint satisfaction problem, there is an indicator function $I\left(\hat{J}_{C^i}^{\pi_\theta}\right)$ that satisfies

$$I\left(\hat{J}_{C^i}^{\pi_\theta}\right) = \begin{cases} 0\,,\ \hat{J}_{C^i}^{\pi_\theta} \leq 0 \\ -\infty\,,\ \hat{J}_{C^i}^{\pi_\theta} > 0 \end{cases} \tag{18}$$

That is, when the constraint conditions are met, the problem is treated as an unconstrained policy optimization problem that only considers rewards; when any constraint is violated, the strategy needs to be adjusted first to meet the constraints. In this way, by expanding the objective with an indicator function, the original CMDP problem can be simplified to an unconstrained optimization problem. The logarithmic barrier function is a differentiable approximation of the indicator function, and its definition is as follows:

$$\phi\left(\hat{J}_{C^i}^{\pi_\theta}\right) = \frac{\log\left(-\hat{J}_{C^i}^{\pi_\theta}\right)}{k} \tag{19}$$

where k is a hyperparameter. At this point, by expanding the objective with the logarithmic barrier function, the final objective function of the IPO algorithm is given by Equation (20).

$$\max_{\theta} L^{\text{IPO}}\,(\theta) = L^{\text{clip}}(\theta) + \sum \phi\left(\hat{J}_C^{\pi_\theta}\right) \tag{20}$$

In the problem of EV charging scheduling, the LSTM network fully reflects the historical process of the input long-time series data, better extracting the price trend of time-varying electricity prices. The network structure proposed in this paper takes the past 24 h of time-varying electricity price data as input, and, through the LSTM network, outputs features containing information about future price trends. This output is used as part of the input for the deep reinforcement learning network to facilitate policy learning.

4.2. EV Charging and Discharging Strategy Optimization Process Based on IPO and PHNN

When an EV is connected to a charging pile, the IPO-PHNN agent takes over the charging and discharging control. Within each charging and discharging action cycle, the agent generates the corresponding charging and discharging power, which is then converted into input for the PHNN to obtain the battery health status. This information, along with electricity prices and the SoC, is input into the IPO algorithm for parameter updates.

The IPO algorithm adds logarithmic barrier functions during the network parameter update process to impose safety constraints on the existing policy, thereby ensuring that the agent explores the optimal EV charging and discharging strategy within the safety constraint space. The specific process is as follows. First, initialize the actor network parameters θ, the critic network parameters μ, and the experience pool, set the maximum number of experience accumulations D and the termination time T, and randomly select a day's array from the training data as the initial state s_1. At each control time period t, the agent outputs an action a_t based on the observed state s_t, receives a reward r_t, and stores the experience $< s_t, a_t, r_t, s_{t+1} >$ in the experience replay pool. Then, the experience pool repeatedly collects experience data until the maximum number of experience accumulations D is reached, randomly samples data from the experience pool, and begins batch network updates. During the actor network update process, the clipped target function is obtained based on the ratio of old to new policies and the advantage function, which is then combined with the logarithmic barrier function to update the actor network parameters θ to obtain a new policy. During the critic network update process, the temporal difference error is calculated based on the state-value function obtained from the network to update the critic network parameters μ, continuously correcting the evaluation error. The training process of the IPO algorithm is shown as Algorithm 1.

Algorithm 1: Network Training Process for the IPO Algorithm

```
(1) Input: Initialize actor network parameters $\theta$, critic network parameters $\mu$, and
experience replay buffer;
        set the maximum number of episodes $K$, learning rates for the actor–critic
network, the capacity of
        the experience buffer $D$, and the termination time $T$.
(2) For $k$ = 0 to $K$ do:
    For $t$ = o to $D$ do:
        1. Set the duration time counter $t$;
        2. Acquire the initial state $s_0$ of the EV and the current SOH.
        3. While $t$ < $T$:
```

Algorithm 1: *Cont.*

```
                Enter the electricity price information at time $t$, price trend information, the ISOC value $s_t$ of the
                EV, and the SOH value;
                Based on the observed state, sample the charging and discharging action $a_t$ using the
                policy $\pi_\theta(a|s_t)$;
                Update the EV battery's ISOC and SOH at time $t + 1$; Calculate the real-time reward $r_t$;
                $t = t + 1$;
            End while
            4. Store the EV charging and discharging interaction trajectory
$(s_0, a_0, r_0, s_1, a_1, r_1, \ldots)$;
            Sample charging and discharging decision samples from the experience replay buffer;
            Calculate the ratio of new to old policies $r(\theta)$ and the advantage function $A_t$;
        Assign the new policy network parameters θ to the old network parameters to update the old network;
        Update the actor new network parameters $\theta$ by combining the logarithmic barrier function;
        Calculate the TD error and update the critic network parameters $\mu$;
(3) Training is complete after the maximum number of iterations.
(4) Output: The optimal real-time charging and discharging strategy for EV $\pi_\theta^*$.
```

5. Case Study

5.1. Basic Parameters of the Algorithm

The experimental software environment consists of Python 3.7 and PyTorch 1.6, while the hardware environment is a computer equipped with an Intel Core i9-13900KF CPU (Intel Corporation, Santa Clara, CA, USA) and an NVIDIA GeForce RTX 3080 Ti GPU (NVIDIA Corporation, Santa Clara, CA, USA). The simulation test uses hourly electricity price data collected by the New England ISO [32]. Users' travel habits are referenced from EV travel records in California. The state transition and charging and discharging control of the vehicle in this paper are based on one-hour time slots. The vehicle's initial SOH is randomly sampled from [0.7, 1], the ISOC is randomly sampled from [0, *SOH*], and the user's desired ISOC at departure ($ISOC_\mathrm{d}$) is randomly sampled from [0.5, *SOH*]. The battery capacity C is set to 24 kWh, and the charging and discharging efficiency η_ch and η_dis are set to 0.98. EV batteries are typically composed of multiple small cells connected in series and parallel. In this paper, we assume the cell type is 18650, the charging and discharging current I_t is fixed at 2 A, the temperature T_t is fixed at 298.15 K (25 °C), and the battery consists of 4300 18650 cells. The input for the PHNN is the predicted SOH of the past eight cycles. The weight coefficients of the reward function are set to 10 and 50, respectively. The hyperparameters of the IPO algorithm are shown in Table 1.

Table 1. The hyperparameters of the IPO.

Parameter	Value
MLP Hidden Layer Size	(256, 256)
LSTM Hidden Layer Size	128
Sliding Steps	24

Table 1. *Cont.*

Parameter	Value
MLP Activation Function	Relu
Actor Network Learning Rate	5×10^{-4}
Critic Network Learning Rate	5×10^{-4}
Discount Factor	0.97
Clipping Factor	0.2
IPO-k	20

This study uses a dataset from NASA's Prognostics Center of Excellence [33]. In this dataset, lithium-ion batteries undergo cycles of charging, discharging, and impedance measurement at room temperature (25 °C). Specifically, the battery is first charged at a constant current of 1.5 A until the voltage reaches 4.2 V; then, the battery is charged at a constant voltage until the charging current drops to 20 mA; finally, the battery is discharged at a continuous discharge current of 2 A until the battery voltage reaches a calibration value. These experimental conditions provide strong data support for battery performance prediction and health state assessment. In the PHNN network, the following configuration is used: the machine learning module consists of three fully connected hidden layers with the number of neurons increasing from 8 to 128 layer by layer; this is followed by an LSTM neural network with 128 neurons; and finally, a linear layer produces the final output Dt. The model-driven module $M_\rho(U_t, I_t)$ consists of three fully connected layers with the number of neurons increasing from 8 to 128 and then decreasing to 1.

5.2. PHNN SOH Degradation Prediction Performance Analysis

To evaluate the prediction performance of the proposed PHNN network model in forecasting the SOH degradation of lithium-ion batteries, this experiment selected the first 50% of the cycle data from Battery B05 as the training set and then conducted testing on the remaining 50% of the cycle data. Figure 2 shows the original SOH and capacity of Battery B05 across all cycles. It can be observed that, as the number of cycles increases, both the SOH and the capacity of the battery decrease.

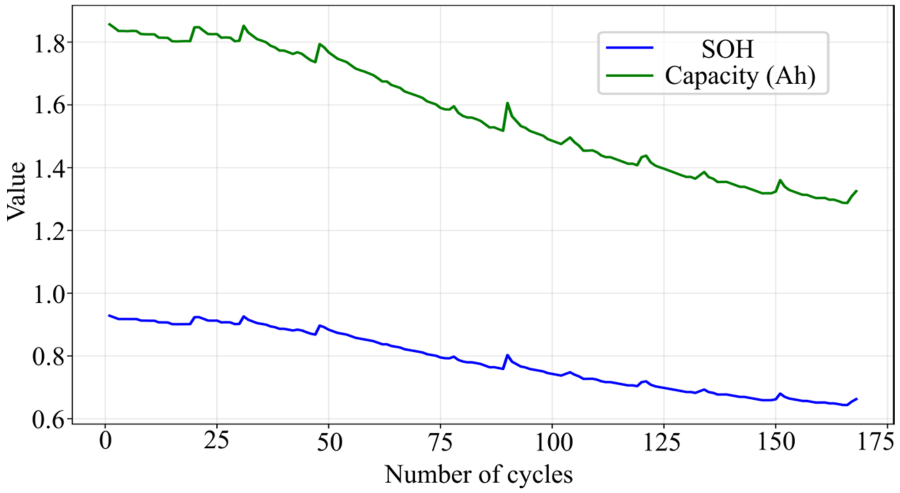

Figure 2. Original SOH and capacity changes of Battery B05 cycle data.

In this experiment, we used the Root Mean Square Error (RMSE) as the loss function to evaluate the training effectiveness of the PHNN neural network model. From the loss curves of the training and test sets, it can be observed that, as the training progresses, the loss of the training set gradually decreases, indicating that the model's fitting effect on the training data is continuously improving. The test set loss, however, shows some fluctuation. In particular, it is higher in the early stages, but as training deepens, the test set loss gradually stabilizes. This fluctuation may reflect the model's overfitting to the test set in the early stages, but as training continues, the model's generalization ability gradually strengthens.

By comparing the loss curves of the training and test sets, it is observed that, in the later stages of training, the test set loss basically trends towards stability and remains close to the training set loss. This indicates that the model has achieved a good level of generalization and exhibits small errors on new data, which is consistent with the ideal training outcome. During the training process, the evolution of loss (as a percentage of total loss) for both the training and test sets, along with the predicted SOH results, are shown in Figure 3. The SOH curve predicted by the PHNN for 63 cycles remains highly consistent with the original SOH curve in terms of overall trend, with an average absolute error of 3.03%. This fully demonstrates the accuracy of the proposed PHNN in predicting the degradation of lithium-ion batteries.

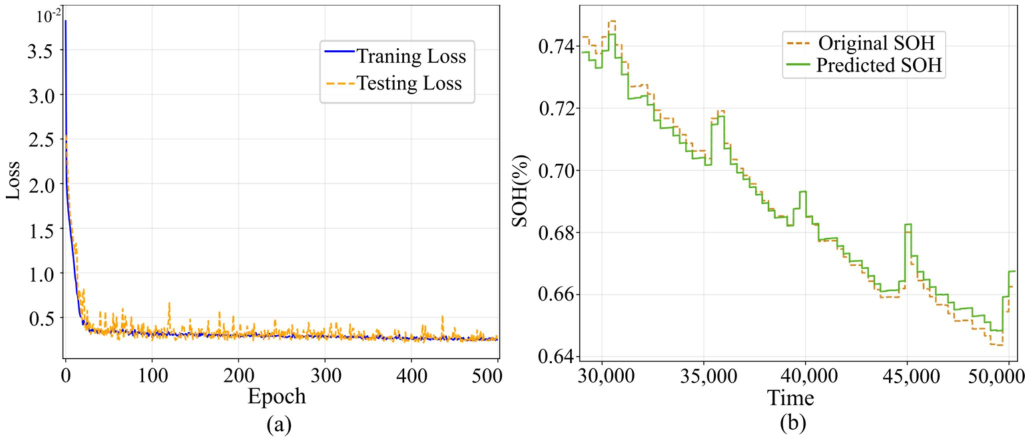

Figure 3. Training progress and SOH prediction results: (**a**) Training and testing loss curves across epochs; (**b**) Comparison between original and predicted SOH values over time.

5.3. Convergence Performance of the IPO Algorithm and Analysis of EV Charging and Discharging Strategies

To validate the effectiveness of the proposed algorithm, this section trains the IPO agent for 1000 episodes and tests the trained agent over 7 test days. Figure 4 presents the convergence of the average cumulative reward value and constraint value per episode during the training process of the proposed method. Due to the randomness of parameters in the reinforcement learning agent training process, we conducted three experiments with different random seeds. The solid lines in the figure represent the mean of multiple experiments, and the shaded areas indicate the variance. As can be observed from Figure 4, both the mean and variance of rewards and constraints show a trend of gradual stability with the progression of training. The IPO algorithm may face significant exploration issues in the early stages of training, leading to higher instability in rewards and constraints. As training continues, the rewards gradually increase and stabilize, and the occurrences of constraint violations decrease and finally approach zero. This demonstrates the performance of the proposed safe reinforcement learning algorithm and its advantage in enforcing constraints.

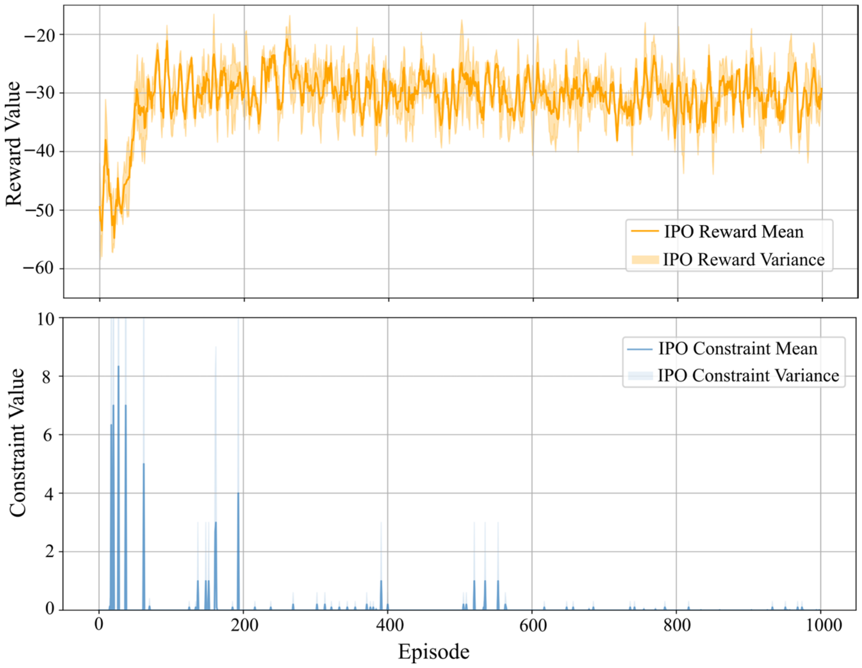

Figure 4. Evolution of rewards and constraints during the training process of the proposed algorithm.

To evaluate the effectiveness of the proposed algorithm in generating EV charging and discharging control strategies, this paper validates the trained EV charging and discharging strategies on a dataset of consecutive 7 test days. Figure 5 presents the hourly electricity prices, the EV charging and discharging schedules, and the corresponding ISOC values on the test days. As can be observed from the figure, the EV discharges rapidly when the electricity price is high, and when the price drops, the EV quickly purchases electricity for charging. During the EV charging and discharging process, the battery's state of charge is maintained within the constraint boundaries, and when the EV departs from home, the EV battery's state of charge reaches the charging target. These results verify the effectiveness of the method in optimizing real-time EV charging and discharging strategies under the constraint of charging demand satisfaction using the IPO algorithm.

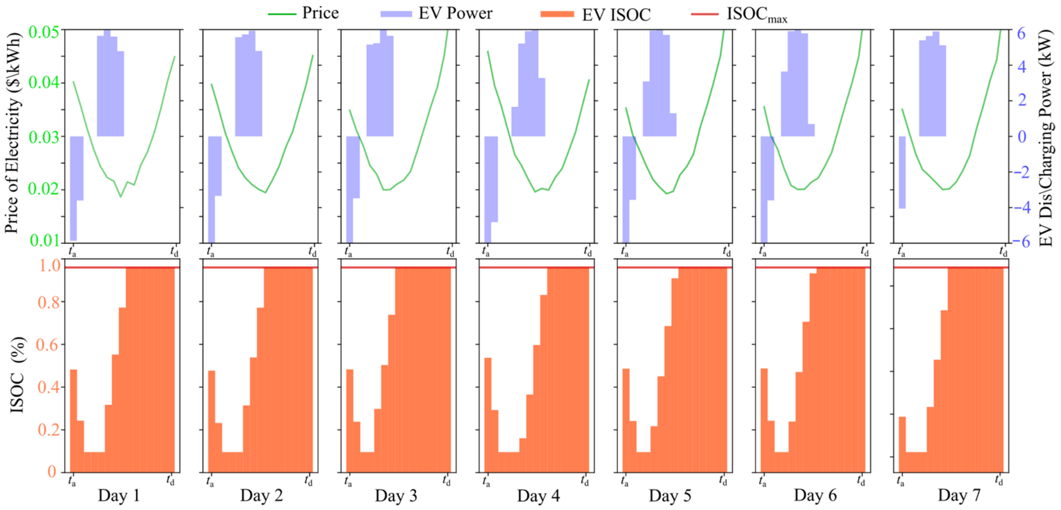

Figure 5. Display of EV charging and discharging strategies and ISOC changes over 7 test days.

To further validate the stability and robustness of the proposed method under different hyperparameter configurations, a sensitivity analysis was conducted on two key parameters in the IPO algorithm: the sliding window length in the LSTM architecture (sliding steps) and the logarithmic barrier function parameter k. The sliding step determines the length of the historical state sequence observed by the agent during each decision-making process, while the parameter k controls the penalty intensity for constraint violations in the policy. We designed three configurations for each parameter and recorded the agent's average cumulative reward and cumulative constraint violation over the final 200 training episodes. The experimental results are summarized in the table.

As shown in Table 2, when the sliding step is set to 24, the agent achieves a favorable trade-off between reward performance and constraint satisfaction. Increasing the step size to 96 leads to marginal improvement in constraint adherence but introduces significant computational overhead, with limited gains in reward. This indicates that an appropriately sized observation window enhances policy quality, whereas overly long sequences may result in training inefficiency. On the other hand, setting the logarithmic barrier parameter k to 20 yields the best balance between reward and constraint satisfaction. Both overly small and excessively large values of k result in performance degradation. In summary, the proposed IPO algorithm demonstrates strong robustness to these two critical hyperparameters within reasonable ranges, achieving a desirable balance between performance and computational efficiency through appropriate tuning.

Table 2. Hyperparameter sensitivity analysis of the proposed algorithm.

Algorithm Hyperparameters		**Cumulative Reward**	**Cumulative Constraint Violation**
Sliding Steps	4	-43.73 ± 3.39	$9.31 \times 10^{-2} \pm 4.6 \times 10^{-3}$
	24	-40.56 ± 3.7	$2.95 \times 10^{-3} \pm 5.1 \times 10^{-3}$
	96	-40.01 ± 2.98	$2.06 \times 10^{-3} \pm 7.7 \times 10^{-3}$
IPO-k	10	-44.93 ± 2.68	$1.96 \times 10^{-1} \pm 6.4 \times 10^{-2}$
	20	-40.56 ± 3.7	$2.95 \times 10^{-3} \pm 5.1 \times 10^{-3}$
	40	-41.17 ± 3.09	$6.89 \times 10^{-2} \pm 3.7 \times 10^{-3}$

5.4. Comparative Advantages of the Proposed Method

To highlight the effectiveness of the proposed IPO algorithm in optimizing charging costs while satisfying energy constraints, a comparative analysis is conducted over a 30-day simulation period between the IPO algorithm and two baseline reinforcement learning methods, CPO [34] and PPO [35], in terms of cumulative constraint violation and cumulative charging reward for EVs. In addition, two traditional model-based approaches are introduced for further comparison. (1) First, an offline optimal scheduling method, where all uncertainties are assumed to be known in advance was introduced. The EV charging/discharging problem is formulated as a deterministic optimization model and solved using the Gurobi solver. The resulting optimal solution (OS), serving as a performance benchmark, reflects an ideal scenario. (2) Second, an MPC method [10], which forecasts EV departure time and future electricity prices at each decision step, and performs rolling horizon optimization to minimize the cost during each time interval, was introduced. The calculation formula for the constraint violation index is $\sum_{N=1}^{30} \sum_{t=0}^{T-1} \gamma^t c_t / d \times 100\%$, where $d = 0.1$ kWh is the maximum constraint tolerance. Finally, the total computational time across all decision steps within a single day (from vehicle arrival to departure) was compared among the five optimization methods.

As shown in Table 3, the proposed IPO algorithm achieves the highest cumulative reward while maintaining a low constraint violation rate of only 46.19%. Compared to the CPO algorithm, which is also based on safe reinforcement learning, the IPO algorithm improves constraint satisfaction performance by 22.54%. In contrast, the PPO algorithm exhibits a significantly higher constraint violation rate of 1318.34%. These results demonstrate that the proposed method outperforms other DRL algorithms in satisfying the upper and lower bounds of EV battery capacity, as well as the energy demand of users' travel schedules. The MPC approach and the solver-based global optimum both perform well in terms of constraint compliance and economic efficiency. However, they require solving a predictive optimization problem at every decision step, resulting in high computational costs, particularly in multi-period optimization scenarios. Specifically, the MPC method takes 87.36 s to complete all decision steps in a single day, while the direct use of an optimization solver requires as much as 1360.1 s.

Table 3. Cumulative rewards, constraint violation rates, and decision time of five optimization methods.

Methods	Cumulative Reward	Cumulative Constraint Violation	Total Computation Time per Day
PPO	−1550.94	1318.34	0.24 s
CPO	−1163.87	68.73	0.24 s
IPO	−963.25	46.19	0.24 s
MPC	−1039.36	236.9	87.36 s
OS	−901.78	0	1360.1 s

In contrast, DRL-based methods such as IPO, PPO, and CPO rely only on forward inference of trained neural networks during the online phase, allowing for real-time decision-making within milliseconds. Although reinforcement learning methods require substantial time and data investment during the offline training phase, their high efficiency during execution makes them well-suited for practical applications involving time-of-use scheduling and dynamic demand response. In summary, the proposed IPO algorithm achieves a balanced trade-off among constraint satisfaction, economic performance, and computational efficiency, demonstrating strong potential for real-world engineering applications.

Furthermore, to verify the improvement of the proposed LSTM data feature extraction for electricity price trends on EV charging and discharging decisions, a comparison of the effects before and after adopting this technology was conducted. The results show that the cumulative rewards for 30 testing days before and after using this technology were −1038.11 and −963.25, respectively. This demonstrates that LSTM feature extraction effectively assists the agent in perceiving changes in time-varying electricity prices within the state space. By discharging during high-price periods to gain revenue and reducing charging costs and quickly recharging the EV during low-price periods, the technique effectively meets the electricity demand for EV charging and discharging. Therefore, by sensing future trends in time-series data, it can more effectively assist the agent's sequential decision-making, thereby reducing charging costs and enhancing robustness in the face of uncertainty.

The lifespan of EV batteries is significantly influenced by frequent charging and discharging cycles, and a battery's SOH is closely related to charging behavior. Excessive charging and discharging accelerates battery aging, leading to early battery replacement and increased overall operational costs. To comprehensively assess the role of the proposed PHNN model in optimizing the trade-off between battery lifespan and economic efficiency, this paper simulates the EV operation process over 30 consecutive working days based on the aforementioned experimental design and compares the overall performance of different algorithms from the perspectives of cumulative reward value, constraint violation rate, and daily decision time. Figure 6 shows the comparison of SOH results. The results indicate that the reinforcement learning method considering the proposed PHNN performs better in terms of maintaining battery SOH, and all methods outperform the other approaches in terms of performance.

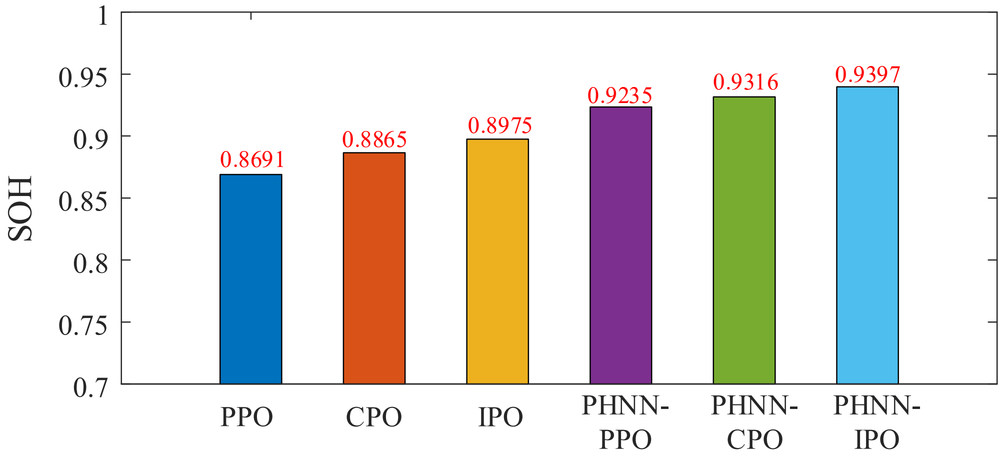

Figure 6. Comparison of SOH optimization results for the proposed method.

Furthermore, we present dynamic estimation of battery SOH under the PHNN-IPO, CPO, and MPC strategies over 15 consecutive workday charge–discharge cycles, plotting their SOH decline curves on the same axes (Figure 7). The results show that PHNN-IPO exhibits the slowest degradation, indicating that it can more effectively retard battery aging while balancing user benefits; both CPO and MPC, lacking explicit modeling of battery health, degrade somewhat faster than PHNN-IPO.

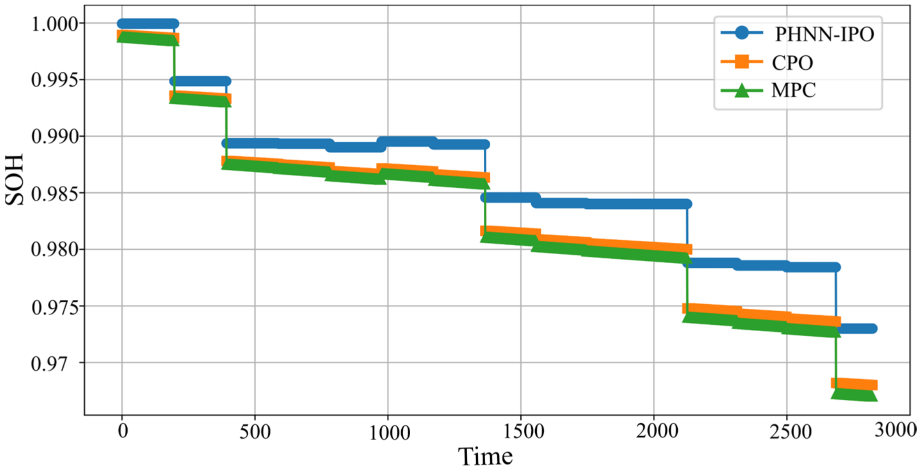

Figure 7. Comparison of SOH degradation rates under three strategies.

6. Conclusions

This paper addresses the optimization problem of EV charging and discharging strategies by proposing a safe reinforcement learning method that considers battery health status. The core innovations of this method include designing a battery health status prediction model based on a PHNN, which combines physical prior knowledge with the LSTM's ability to capture temporal features, significantly improving the accuracy of Battery SOH predictions. The EV charging and discharging decision-making problem is formulated as a constrained Markov decision process considering battery health status, and a reward function and cost function that incorporate battery health status are designed to accurately quantify the cost of battery degradation in V2G. An LSTM-based IPO algorithm is proposed, which introduces a logarithmic barrier function to handle constraints, effectively reducing computational costs and improving policy performance.

The experimental results show that the proposed EV charging and discharging strategy optimization method can significantly slow down the decline in battery health while maximizing the charging and discharging benefits for EV owners. Specifically, the results are as follows:

(1) The PHNN model exhibits high accuracy in predicting lithium battery degradation, with an average absolute error of 3.03%.
(2) The IPO algorithm demonstrates good convergence during training and is capable of optimizing real-time EV charging and discharging strategies while satisfying charging demand constraints.
(3) Compared to other optimization algorithms, the IPO algorithm shows advantages in both cumulative reward values and constraint violation rates, particularly in meeting the power constraint, where the performance improvement is significant.
(4) Additionally, this paper verifies the role of LSTM data feature extraction in perceiving price trend improvements in EV charging and discharging decisions, as well as the superior performance of the reinforcement learning algorithm considering PHNN in terms of balancing battery SOH.

Overall, the research presented in this paper offers a new solution for the intelligent and personalized charging strategies of EVs, which is of great significance for promoting the sustainable development of new energy vehicles.

Author Contributions: Conceptualization, S.G. and Y.Y.; Methodology, S.G. and K.Q.; Software, S.G.; Validation, S.G., K.Q. and Y.Y.; Formal Analysis, S.G. and K.Q.; Investigation, S.G. and K.Q.; Resources, Y.Y.; Data Curation, S.G.; Writing—Original Draft Preparation, S.G.; Writing—Review & Editing, K.Q. and Y.Y.; Visualization, S.G.; Supervision, Y.Y.; Project Administration, Y.Y.; Funding Acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the State Grid Jiangsu Electric Power Co., Ltd. Science and Technology Project: J2020092.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Authors Shuifu Gu and Kejun Qian were employed by the State Grid Jiangsu Electric Power Co., Ltd.—Suzhou Power Supply Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Huang, W.W.; Wu, L.F.; Zeng, X.B.; Sun, Y. Analysis of spatial distribution characteristics of electric vehicle sales and service shops and the corresponding influencing factors. *Trans. GIS* **2023**, *27*, 1357–1390. [CrossRef]
2. Tavakoli, A.; Negnevitsky, M.; Saha, S.; Haque, E.; Arif, M.T.; Contreras, J.; Oo, A. Self-Scheduling of a Generating Company With an EV Load Aggregator Under an Energy Exchange Strategy. *IEEE Trans. Smart Grid* **2019**, *10*, 4253–4264. [CrossRef]
3. Li, Z.; He, Y.L.; Peng, G.; Yin, J. Stochastic Optimal Scheduling of Flexible Traction Power Supply System for Heavy Haul Railway Considering the Online Degradation of Energy Storage. *World Electr. Veh. J.* **2025**, *16*, 206. [CrossRef]
4. Leijon, J. Charging strategies and battery ageing for electric vehicles: A review. *Energy Strategy Rev.* **2025**, *57*, 101641. [CrossRef]
5. Wang, Z.; Jochem, P.; Fichtner, W. A scenario-based stochastic optimization model for charging scheduling of electric vehicles under uncertainties of vehicle availability and charging demand. *J. Clean. Prod.* **2020**, *254*, 119886. [CrossRef]
6. Tran, T.D.; Nguyen, N.-D.; Chu, H.; Ghaoui, L.E.; Ambrosino, L.; Calafiore, G. A robust optimization model for cost-efficient and fast electric vehicle charging with L_2-norm uncertainty. *arXiv* **2025**, arXiv:2502.04024.
7. Wu, S.C.; Pang, A.P. Optimal scheduling strategy for orderly charging and discharging of electric vehicles based on spatio-temporal characteristics. *J. Clean. Prod.* **2023**, *392*, 136318. [CrossRef]
8. Kim, H.Y.; Shin, C.S.; Mahseredjian, J.; Kim, C.-H. Voltage Stability Index(VSI)-Based Optimal Vehicle-to-Grid(V2G) Charging/Discharging Strategy in Radial Distribution System. *J. Electr. Eng. Technol.* **2024**, *19*, 3885–3890. [CrossRef]
9. Zhu, Y.S.; Sun, X.; Xie, X.F.; Ding, T.K.; Wu, F.Z.; Shi, Z.P. Multi-objective collaborative optimal dispatch for electric vehicles in multistate scenarios considering trip chain reconstruction. *Autom. Electrie Power Syst.* **2024**, *48*, 129–141.

10. Hu, Q.H.; Amini, M.R.; Wiese, A.; Seeds, J.B.; Kolmanovsky, I.; Sun, J. Electric Vehicle Enhanced Fast Charging Enabled by Battery Thermal Management and Model Predictive Control. *IFAC-PapersOnLine* **2023**, *12*, 10684–10689. [CrossRef]
11. Chung, H.M.; Maharjan, S.; Zhang, Y.; Eliassen, F. Intelligent charging management of electric vehicles considering dynamic user behavior and renewable energy: A stochastic game approach. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 7760–7771. [CrossRef]
12. Alsabbagh, A.; Wu, B.; Ma, C.B. Distributed electric vehicles charging management considering time anxiety and customer behaviors. *IEEE Trans. Ind. Inform.* **2021**, *17*, 2422–2431. [CrossRef]
13. Yalçın, S.; Herdem, M.S. Optimizing EV Battery Management: Advanced Hybrid Reinforcement Learning Models for Efficient Charging and Discharging. *Energies* **2024**, *17*, 2883. [CrossRef]
14. Liu, D.N.; Wang, L.X.; Wang, W.Y. Optimal scheduling of electric vehicle load for large-scale battery charging and swapping based on deep reinforcement learning. *Autom. Electr. Power Syst.* **2022**, *46*, 36–46.
15. Zhang, W.N.; Li, R.; Zang, X.D.; Yan, J.R.; Zhu, J.Y. Real-time scheduling strategy optimization for electric vehicle battery swapping station based on reinforcement learning. *Electr. Power Autom. Equip.* **2022**, *42*, 134–141.
16. Zhang, F.Y.; Yang, Q.Y.; An, D. CDDPG: A deep-reinforcement-learning-based approach for electric vehicle charging control. *IEEE Internet Things J.* **2021**, *8*, 3075–3087. [CrossRef]
17. Yan, L.F.; Chen, X.; Zhou, J.Y.; Chen, Y.; Wen, J. Deep reinforcement learning for continuous electric vehicles charging control with dynamic user behaviors. *IEEE Trans. Smart Grid* **2021**, *12*, 5124–5134. [CrossRef]
18. Yang, Z.; Li, L.; Yu, L.; Shahidehpour, M. Constrained policy optimization for electric vehicle charging strategy considering grid security constraints. *IEEE Trans. Smart Grid* **2023**, *14*, 1085–1096.
19. Qin, Y.H.; Xing, Z.F.; Li, X.L.; Zhang, Z.S.; Zhang, H.J. Primal-Dual Deep Reinforcement Learning for Periodic Coverage-Assisted UAV Secure Communications. *IEEE Trans. Veh. Technol.* **2024**, *73*, 19641–19652. [CrossRef]
20. Peng, J.C.; Gao, Y.; Cai, L.; Zhang, M.; Sun, C.; Liu, H. State of Health Estimation for Lithium-Ion Batteries Using Electrochemical Impedance Spectroscopy and a Multi-Scale Kernel Extreme Learning Machine. *World Electr. Veh. J.* **2025**, *16*, 224. [CrossRef]
21. Song, Y.C.; Peng, Y.; Liu, D. Model-based health diagnosis for lithium-ion battery pack in space applications. *IEEE Trans. Ind. Electron.* **2021**, *68*, 12375–12384. [CrossRef]
22. Hu, X.S.; Yuan, H.; Zhou, C.F.; Li, Z.; Zhang, L. Co-estimation of state of charge and state of health for lithium-Ion batteries based on fractional-order calculus. *IEEE Trans. Veh. Technol.* **2018**, *67*, 10319–10329. [CrossRef]
23. Li, Y.; Wei, Z.B.; Xiong, B.Y.; Vilathgamuwa, D.M. Adaptive ensemble-based electrochemical-thermal degradation state estimation of lithium-ion batteries. *IEEE Trans. Ind. Electron.* **2022**, *69*, 6984–6996. [CrossRef]
24. Fu, Y.M.; Xu, J.; Shi, M.J.; Mei, X. A fast impedance calculation-based battery state-of-health estimation method. *IEEE Trans. Ind. Electron.* **2022**, *69*, 7019–7028. [CrossRef]
25. Ren, L.; Zhao, L.; Hong, S.; Zhao, S.; Wan, H.; Zhang, L. Remaining useful life prediction for lithium-ion battery: A deep learning approach. *IEEE Access* **2018**, *6*, 50587–50598. [CrossRef]
26. Qin, Y.; Yuen, C.; Yin, X.Y.; Huang, B. A transferable multistage model with cycling discrepancy learning for lithium-ion battery state of health estimation. *IEEE Trans. Ind. Inform.* **2023**, *19*, 1933–1946. [CrossRef]
27. Xu, P.H.; Wang, C.C.; Ye, J.L.; Ouyang, T. State-of-charge estimation and health prognosis for lithium-ion batteries based on temperature-compensated bi-LSTM network and integrated attention mechanism. *IEEE Trans. Ind. Electron.* **2024**, *71*, 5586–5596. [CrossRef]
28. Li, T.T.; Zhang, W.C.; Huang, G.S.; He, H.; Xie, Y.; Zhu, T.; Liu, G. Real-world data-driven charging strategies for incorporating health awareness in electric buses. *J. Energy Storage* **2024**, *92*, 112064. [CrossRef]
29. Li, X.R.; Wang, W.; Jin, K.; Qin, S. Joint optimization of vehicle scheduling and charging strategies for electric buses to reduce battery degradation. *J. Renew. Sustain. Energy* **2024**, *4*, 044704. [CrossRef]
30. Li, K.; Zhou, P.; Lu, Y.F.; Han, X.; Li, X.; Zheng, Y. Battery life estimation based on cloud data for electric vehicles. *J. Power Sources* **2020**, *468*, 228192. [CrossRef]
31. Yue, L.; Qiong, W.; Shun, L.; Liu, S.; Li, Q.T.; Liu, Y.; Wang, H.B. Reinforcement Learning Algorithm for Charging Discharging Control of Electric Vehicles Considering Battery Loss. *Comput. Sci.* **2024**, *51*, 1042–1048.
32. ISO New England. Realtime Maps and Charts [EB/OL]. Available online: https://www.iso-ne.com/isoexpress/ (accessed on 18 March 2025).
33. Saha, B.; Goebel, K. Moffett Field, CA, USA: NASA AmesPrognostics Data Repository [EB/OL]. Available online: http://ti.arc.nasa.gov/project/prognostic-data-repository (accessed on 18 March 2025).

34. Li, H.P.; Wan, Z.Q.; He, H.B. Constrained EV Charging Scheduling Based on Safe Deep Reinforcement Learning. *IEEE Trans. Smart Grid* **2020**, *3*, 2427–2439. [CrossRef]
35. Hong, L.C.; Wu, M.H.; Wang, Y.F.; Shahidehpour, M.; Chen, Z.H.; Yan, Z.H. MADRL-Based DSO-Customer Coordinated Bi-Level Volt/VAR Optimization Method for Power Distribution Networks. *IEEE Trans. Sustain. Energy* **2024**, *15*, 1834–1846. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.