

Article

A Hybrid Framework of VMD-KPCA and PLO-PINN for Lithium-Ion Battery SOH Estimation

Zhiwei Yang ¹, Qianli Dong ¹, Rui Dong ² and Guangjun Liu ^{1,*}

¹ Hubei Key Laboratory for High-Efficiency Utilization of Solar Energy and Operation Control of Energy Storage System, Hubei University of Technology, Wuhan 430068, China; yzw20020107@163.com (Z.Y.); 13932669704@163.com (Q.D.)

² Motor Soft-Start Hubei Provincial Key Laboratory, Xiangyang 441022, China; dongrui@bigpawer.com

* Correspondence: 20161023@hbut.edu.cn

Abstract

Accurate state of health (SOH) estimation of lithium-ion batteries (LIBs) is critical to ensuring the safety and reliability of battery management system (BMS). To achieve precise estimation, this study proposes a hybrid framework that integrates variational mode decomposition (VMD), kernel principal component analysis (KPCA), polar lights optimizer (PLO), and physics-informed neural network (PINN) for SOH estimation. First, multi-dimensional health features are extracted and decomposed by VMD into intrinsic mode functions (IMFs), which are then compressed into a one-dimensional principal component via KPCA, retaining over 95% of the original information. Subsequently, the PLO algorithm is used to adaptively optimize three key hyperparameters of the PINN-based model: the learning rate, the number of collocation points, and the regularization loss weight. Finally, the optimized PINN is deployed to predict the SOH of the Center for Advanced Life Cycle Engineering (CALCE) battery dataset. Experimental results demonstrate that the proposed VMD-KPCA-PLO-PINN exhibits high prediction accuracy under both 7:3 and 5:5 training-to-testing data partitions. For example, under the 5:5 partition, the proposed model achieves an average R^2 of 0.983 and an average RMSE of 0.0085 on the tested CALCE cells.

Keywords: lithium-ion battery; state of health; variational mode decomposition; kernel principal component analysis; physics-informed neural network; polar lights optimizer

1. Introduction

Amid the global efforts to combat climate change and address the traditional energy crisis, electric vehicles (EVs) and new energy storage sectors have experienced rapid growth. Lithium-ion batteries (LIBs), known for their high energy density, long cycle life, low self-discharge rates, and environmental benefits, have become the main power source for EVs and various stationary energy storage systems [1,2]. However, as the cycle count of LIBs increases, battery performance inevitably degrades. Internally, a series of complex and irreversible physicochemical reactions occur, including electrolyte decomposition, thickening of the solid electrolyte interface layer, loss of active material, and structural damage to the negative electrode materials [3]. These processes result in diminished discharge capacity and increased internal resistance. Particularly under harsh operating conditions characterized by high power output and long service life, the battery's aging rate can be significantly accelerated. More critically, aging compromises the thermal stability of the battery, potentially leading to abnormal temperature rises that trigger internal short



Academic Editor: Michael Fowler

Received: 3 June 2026

Revised: 5 July 2026

Accepted: 10 July 2026

Published: 16 July 2026

Copyright: © 2026 by the authors.

Published by MDPI on behalf of the

World Electric Vehicle Association.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

circuits, thermal runaway, and even catastrophic events such as fires or explosions [4]. To mitigate the aforementioned failures and extend battery service life, the integration of a highly efficient Battery Management System (BMS) is of paramount importance [5]. Among the various monitoring indices within the BMS, State of Health (SOH) serves as the core parameter for quantifying the degree of battery aging [6]. Generally, SOH is defined as the ratio of the current battery capacity to its initial capacity under specific operating conditions, expressed as a percentage [7,8].

$$SOH_k = \frac{C_k}{C_0} \times 100\% \quad (1)$$

where C_k denotes the measured discharge capacity at the k -th cycle, and C_0 represents the initial measured discharge capacity of the corresponding battery. When the SOH drops below a predefined threshold, typically 80% of the initial capacity, the battery is generally considered to reach its end-of-life state [9]. Therefore, achieving SOH prediction not only facilitates the real-time identification of battery degradation levels to provide a reliable decision-making basis for BMS but also serves as a critical barrier ensuring the safe and stable operation of the battery system [10–12].

Currently, SOH estimation methods for LIBs are mainly divided into model-based and data-driven methods. Among them, model-based approaches primarily encompass empirical models, equivalent circuit models, and electrochemical models [13,14]. Empirical model-based methods develop estimation models by analyzing the number of cycles and the ohmic resistance accumulated during battery utilization [15]. However, the accuracy of these models is often limited due to the lack of comprehensive experimental data. Furthermore, they struggle to account for individual battery variations and the complex interactions of diverse aging mechanisms [16]. Equivalent circuit model-based methods construct state-space equations to characterize battery aging behavior using multiple passive circuit components. Due to their simplicity, stability, and interpretability, they are widely applied in industry [17,18]. However, their model parameters are susceptible to variations in operating conditions, and parameter identification typically requires electrochemical impedance spectroscopy techniques, which are difficult to implement in real vehicle applications [14]. Electrochemical model-based methods, based on complex internal electrochemical reaction equations, can accurately predict particle behavior inside the battery, thereby evaluating the battery status and performance under different operating conditions [19]. However, the parameters of these models can usually only be acquired in static laboratory environments. Furthermore, due to manufacturing processes, variations in parameters still exist among batteries of the same model, making the model difficult to generalize [19]. Concurrently, the model solution process is also accompanied by high computational complexity [14]. Data-driven methods are widely applied in the battery cell domain due to their independence from specific battery systems or models. By utilizing sufficient data, these methods can develop high-precision, application-specific models [20]. However, data-driven methods rely heavily on large datasets, and their performance degrades when the quality and quantity of measurement data are low [21]. Furthermore, they suffer from inherent interpretability deficiencies, occasionally yielding predictions that violate physical principles [22].

The Physics-Informed Neural Network (PINN) framework, which integrates data-driven methods with physical priors, is demonstrating broad application prospects in the field of lithium-ion battery SOH estimation [23–25]. Hofmann et al. [26] fused physicochemical mechanisms with neural networks, achieving root mean square error (RMSE) below 2% and 3% on synthetic and real-world datasets, respectively, thereby validating its superiority over purely data-driven models. Wang et al. [27] introduced empirical degradation models

and state-space equations, verifying their approach on over 310,000 samples. Notably, mean absolute percentage error (MAPE) of the SOH estimation was only 0.87%, and the model demonstrated outstanding generalization capability in small-sample and cross-condition migration experiments. Singh et al. [28] achieved high-accuracy and robust predictions of SOH under low-complexity and small-sample conditions by explicitly embedding the partial differential equations (PDEs) of Fick's law of diffusion from a single-particle model into the PINN training phase.

The health features (HFs) extracted from LIBs, similar to raw capacity data, exhibit non-linearity and capacity regeneration phenomena. When processing such complex, non-linear sequences, the traditional empirical mode decomposition (EMD) and its variants are still plagued by under-decomposition or over-decomposition issues [29,30]. As an advanced alternative, variational mode decomposition (VMD) effectively overcomes these inherent limitations of EMD, successfully decoupling the intricate time-series signals into a set of intrinsic mode functions (IMFs) in a highly efficient and well-isolated manner. Consequently, this technique has demonstrated significant advantages and promising prospects in the field of LIBs health prognosis [31].

As a robust tool for classification and dimensionality reduction, Kernel Principal Component Analysis (KPCA) excels in purging redundant information and suppressing noise interference, thereby preventing model overfitting while simultaneously minimizing architectural complexity and training durations [32]. To resolve the information redundancy inherent in extracted HFs and streamline computational demands, Zheng et al. [33] projected the selected HFs into a single-dimensional space via KPCA, employing the resultant feature as the primary input for SOH prognostics. Similarly, the framework developed by He et al. [34] leveraged KPCA to compress and reconstruct the stabilized components derived from the decomposition of raw lithium-ion battery SOH sequences, a strategy that successfully curtailed cumulative modeling discrepancies.

The prediction performance of PINN is heavily dependent on the configuration of their network hyperparameters. Traditional manual tuning strategies are notoriously time-consuming and highly susceptible to trapping the model in local optima. Nevertheless, intelligent optimization algorithms can efficiently search for the optimal combination of model hyperparameters, thereby augmenting the prediction accuracy and overall performance of the model [35]. The Polar Lights Optimizer (PLO) is a novel meta-heuristic swarm intelligence algorithm, fundamentally inspired by the dynamic motion behaviors of high-energy charged particles during the formation of natural auroras [36]. Notably, Wang et al. [37] applied the PLO algorithm to optimize the hyperparameters of a hybrid Transformer-LSTM model. Their comparative results suggested that the PLO-optimized variant achieved higher SOH estimation accuracy than the unoptimized baseline, indicating the potential of PLO in improving the performance of complex deep learning networks.

The overall research workflow is illustrated in Figure 1. First, raw cycling data from the Center for Advanced Life Cycle Engineering (CALCE) public dataset at the University of Maryland, specifically focusing on four batteries from the CS2 series (CS2-35, CS2-36, CS2-37, and CS2-38), undergo charge-discharge separation and outlier removal to extract five HFs. Second, VMD is applied to decompose each HF into six IMFs, resulting in a 30-dimensional IMF matrix. Subsequently, this high-dimensional matrix serves as the input for KPCA to reduce the dimensionality. Specifically, the first principal component, which achieves a contribution rate exceeding 95%, is selected as the Improved Health Feature (IHF). Then, the IHF is fed into PINN optimized by the PLO algorithm for SOH estimation. The PLO algorithm automatically searches for the optimal combination of three key hyperparameters—the learning rate, the number of collocation points, and the regularization loss weight—within a predefined search space. Finally, experiments are

conducted on four batteries of the CS2 series using a 7:3 training-to-testing split ratio to compare performance against three baseline models: VMD-KPCA-BP, VMD-KPCA-LSTM and VMD-KPCA-PINN. To further examine the prediction stability of the proposed method under a reduced training proportion, additional experiments are performed with a 5:5 split ratio. The results demonstrate that the proposed VMD-KPCA-PLO-PINN method exhibits significant advantages in four core evaluation metrics: mean absolute error (MAE), coefficient of determination (R^2), RMSE and MAPE.

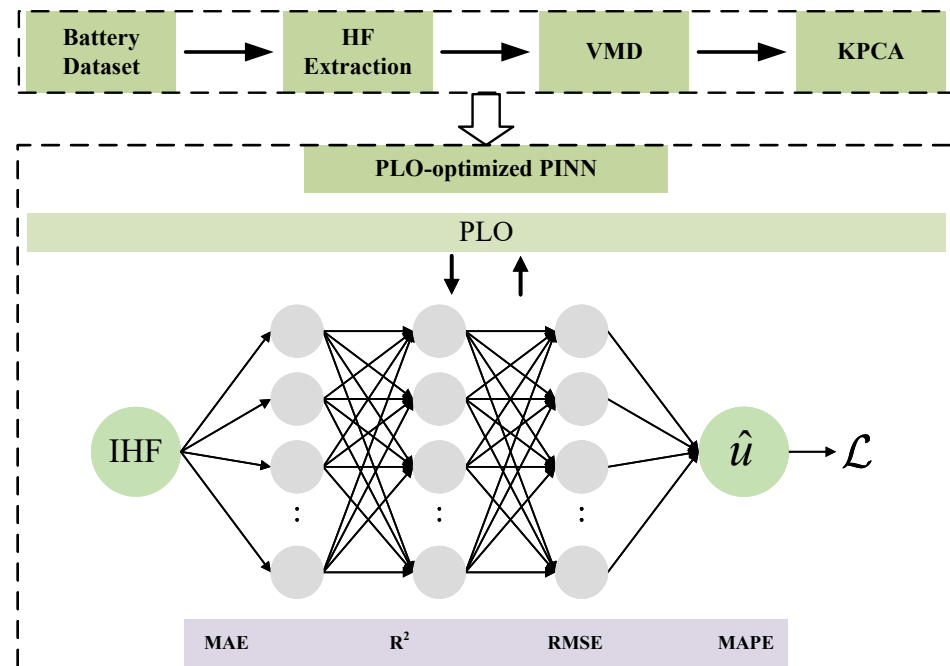


Figure 1. The overall research workflow.

To further clarify the contribution of each module, the proposed VMD-KPCA-PLO-PINN framework can be understood as a coordinated architecture consisting of four functional components. First, VMD acts as the multi-scale decomposition module. It decomposes the extracted health features into several intrinsic mode functions, allowing the global degradation trend and local fluctuation components to be separated more effectively. This process improves the quality of the input features and reduces the influence of nonlinear fluctuations and noise. Second, KPCA serves as the nonlinear feature reconstruction module. After VMD decomposition, the feature dimension increases significantly; therefore, KPCA is introduced to compress the decomposed IMF matrix into a compact improved health feature while preserving the dominant degradation information. Third, the PINN-based model is used as a degradation-informed prediction module. By incorporating an empirical degradation constraint into the loss function, the model introduces degradation-informed regularization into the SOH estimation process, which helps improve the consistency between the predicted SOH trajectory and the general battery degradation trend. Finally, PLO is adopted as the adaptive hyperparameter optimization module. It optimizes the learning rate, the number of collocation points, and the regularization loss weight, thereby reducing the dependence on manual parameter tuning and improving the adaptability of the PINN model. Therefore, the proposed framework is not intended as a simple stacking of existing methods, but as a coordinated SOH estimation strategy in which each module addresses a specific limitation in feature representation, degradation-trend consistency, or parameter selection.

2. Theoretical Framework and Methodology

2.1. VMD

VMD technique stands out in handling time-series data characterized by complex, nonlinear patterns. Employing an iterative search to refine a variational framework, the algorithm adaptively isolates various signal elements while pinpointing their respective center frequencies and bandwidths. This approach essentially breaks down non-stationary, nonlinear signals into distinct IMFs. In practice, the VMD algorithm is implemented by constructing and solving a constrained variational problem. The goal is to find K mode components with specific center frequencies while minimizing the total estimated bandwidth. For a given original time signal f , the constrained model is defined as follows:

$$\begin{aligned} \min_{\{u_k\}, \{\omega_k\}} & \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t.} & \sum_{k=1}^K u_k = f \end{aligned} \quad (2)$$

where u_k and ω_k represent the k -th mode component after decomposition and its corresponding center frequency, respectively; $\delta(t)$ is the Dirac delta function; $\otimes$ denotes the convolution operator; and j is the imaginary unit.

To transform the constrained variational problem in Equation (2) into an unconstrained one, a quadratic penalty factor α and a Lagrange multiplier λ are introduced. This leads to the construction of the augmented Lagrangian function:

$$\begin{aligned} \mathcal{L}(\{u_k\}, \{\omega_k\}, \lambda) = & \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ & + \left\| f(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \langle \lambda(t), f(t) - \sum_{k=1}^K u_k(t) \rangle \end{aligned} \quad (3)$$

where α denotes the penalty parameter that ensures reconstruction accuracy in the presence of Gaussian noise, and λ represents the Lagrange multiplier employed to enforce the strictness of the constraints.

To solve this function, the Alternating Direction Method of Multipliers (ADMM) is employed to iteratively update u_k , ω_k , and λ :

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (4)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \quad (5)$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left[\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right] \quad (6)$$

where $\hat{u}_k^{n+1}$, $\hat{f}(\omega)$, and $\hat{\lambda}(\omega)$ denote the Fourier transforms of u_k^{n+1} , $f(t)$, and $\lambda(t)$, respectively; n represents the iteration index; and τ is the update step size.

Given the discrimination accuracy ϵ , the algorithm ceases iteration once the stopping condition is met. The stopping criterion is expressed as:

$$\sum_{k=1}^K \frac{\|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \epsilon \quad (7)$$

where ϵ denotes the predefined convergence tolerance.

Finally, the original signal f is decomposed into K discrete components $\{u_1, u_2, \dots, u_K\}$. Each of these components is a bandwidth-limited modal signal with a specific center frequency, which effectively captures the local characteristics of the original signal across different frequency bands.

2.2. KPCA

KPCA maps data into a high-dimensional feature space using kernel functions, enabling nonlinear signals to be effectively processed within a linear framework. The m -dimensional IMFs and n battery samples are organized into an initial matrix $X_{nm} = [x_1, x_2, x_3, \dots, x_m]$ and normalized, providing a foundation for subsequent dimensionality reduction.

First, the centered kernel matrix K' is calculated. By utilizing the polynomial kernel function defined in Equation (8) as the nonlinear kernel, the initial kernel matrix K is obtained. Subsequently, K is centered to yield the centered kernel matrix K' , as expressed in Equation (9).

$$k_{poly}(u, v) = (u \cdot v + 1)^d \quad (8)$$

$$K = k_{poly}(x_k, x_g) \quad (9)$$

$$K' = K - \frac{1}{n}ee^TK - K\frac{1}{n}ee^T + \frac{1}{n^2}ee^TKee^T \quad (10)$$

where d is the polynomial degree, x_k and x_g are samples from the parameter matrix, and e is a column vector with all elements equal to 1.

Then, to map the data space into a high-dimensional space, a nonlinear function $\phi(x_k)$ is introduced to obtain the covariance matrix C , where its eigenvalues λ and eigenvectors V satisfy the conditions shown in Equation (12):

$$C = \frac{1}{m} \sum_{k=1}^m \phi(x_k) \cdot \phi(x_k)^T \quad (11)$$

$$\lambda \phi(x_k) V = \phi(x_k) C V \quad (12)$$

From the above equations, the eigenvector V of the covariance matrix C can be obtained. Subsequently, by introducing $K' = (\phi(x_k), \phi(x_g))$ and combining it with the previous expressions, the projection of the samples onto the nonlinear mapping $\phi(x)$ is yielded as follows:

$$V = \sum_{k=1}^m \alpha_k \phi(x_k) \quad (13)$$

$$V \cdot \phi(x) = \sum_{k=1}^m \alpha_k \phi(x_k) \phi(x) = \sum_{k=1}^m \alpha_k K(x_k, x) \quad (14)$$

where α is the eigenvector of K' .

Finally, the cumulative contribution rate D_Σ of the principal components is expressed as:

$$D_\Sigma = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^m \lambda_i}, \quad (k = 1, 2, \dots, m) \quad (15)$$

2.3. PINN

In this paper, a fully connected neural network $\mathcal{F}(x; \Phi)$ is constructed as the predictor, where Φ denotes the learnable weights and biases of the network. The network takes IHF x obtained after VMD-KPCA processing as its input, and yields the predicted SOH value $\hat{u}$

as the output. The network architecture comprises three hidden layers, each configured with 32 neurons. The hyperbolic tangent (tanh) function is employed as the activation function. The above architecture was selected considering the low-dimensional input of the proposed model. After VMD-KPCA processing, the input is reduced to a one-dimensional IHF, and the output is the SOH value. Therefore, a compact fully connected network with three hidden layers and 32 neurons per layer is sufficient to capture the nonlinear relationship between IHF and SOH while avoiding excessive model complexity. In addition, the tanh activation function is smooth and differentiable, which is suitable for calculating the degradation-informed regularization residual in the PINN-based framework. The mapping relationship can be mathematically expressed as:

$$\hat{u} = \mathcal{F}(x; \Phi) \quad (16)$$

The total loss function of the PINN-based framework ($\mathcal{L}$) is defined as a weighted combination of the data-driven loss ($\mathcal{L}_{\text{data}}$) and the degradation-informed regularization loss ($\mathcal{L}_{\text{phys}}$):

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \lambda_{\text{phys}} \mathcal{L}_{\text{phys}} \quad (17)$$

where λ_{phys} represents the regularization coefficient, which balances the trade-off between data fitting and compliance with the empirical degradation prior. $\mathcal{L}_{\text{data}}$ quantifies the mean squared error (MSE) between the network predictions and the experimental SOH labels, defined as:

$$\mathcal{L}_{\text{data}} = \frac{1}{N} \sum_{i=1}^N |u^i - \hat{u}^i|^2 \quad (18)$$

where N denotes the total number of training samples, and u^i represents the experimentally measured normalized capacity SOH label of the i -th sample. $\mathcal{L}_{\text{phys}}$ evaluates the degree of violation of the degradation-informed empirical constraint by the neural network output.

It should be noted that the input variable $x \in \mathbb{R}$ in this study is the IHF obtained after VMD-KPCA processing, rather than a direct physical time, cycle number, or electrochemical state variable. Therefore, the constraint introduced in this work is not intended to represent a strict first-principles electrochemical governing equation. Instead, it is used as an empirical degradation-trend-informed regularization term. The motivation is that the IHF retains the dominant degradation information of the battery and shows a strong correlation with SOH. Based on the commonly observed smooth degradation tendency of lithium-ion batteries, the following exponential-type empirical constraint is introduced to regularize the network output:

$$\frac{du}{dx} + u = 0 \quad (19)$$

This constraint is used to encourage the predicted SOH to follow a smooth degradation trend in the transformed feature space, rather than to fully describe the internal electrochemical aging mechanism. Therefore, the term $\mathcal{L}_{\text{phys}}$ in this study should be interpreted as a degradation-informed regularization loss rather than a strict electrochemical constraint loss.

During the network training process, the partial derivative of the neural network output $\hat{u}$ with respect to the continuous input feature x is approximated using a first-order forward finite difference scheme:

$$\frac{\partial \hat{u}}{\partial x} \approx \frac{\hat{u}(x + \epsilon) - \hat{u}(x)}{\epsilon} \quad (20)$$

where $\epsilon = 0.1$ denotes the perturbation step size for the finite difference approximation. Consequently, the degradation-informed regularization loss $\mathcal{L}_{\text{phys}}$ is formulated as:

$$\mathcal{L}_{\text{phys}} = \frac{1}{N_{\text{phys}}} \sum_{i=1}^{N_{\text{phys}}} \left| \frac{\hat{u}(x_{\text{phys}}^i + \epsilon) - \hat{u}(x_{\text{phys}}^i)}{\epsilon} + \hat{u}(x_{\text{phys}}^i) \right|^2 \quad (21)$$

where N_{phys} is the number of collocation points, and x_{phys}^i denotes the i -th collocation point sampled within the domain boundaries of the training input space. These collocation points do not require true experimental SOH labels and are used only to impose the empirical degradation-informed regularization during network training.

2.4. PLO

PLO improves optimization efficiency and convergence performance by simulating the trajectory of charged particles in the Earth's magnetic field and effectively adjusting the search range and direction. PLO integrates three distinct stochastic strategies: rotational motion, auroral oval walking, and particle collision.

2.4.1. Population Initialization

The PLO algorithm utilizes a stochastic initialization strategy, where the iterative process begins with an initial population generated via pseudo-random numbers. Assuming the population consists of N candidate solutions within a D -dimensional search space, the entire population matrix $\mathbf{X}_{ND}$ can be defined as follows:

$$\mathbf{X}_{ND} = LB + R(UB - LB) = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,D} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & x_{N,2} & \cdots & x_{N,D} \end{bmatrix} \quad (22)$$

where LB and UB denote the lower and upper boundaries of the search space, respectively, and R is a random sequence uniformly distributed in the interval $[0, 1]$.

2.4.2. Rotational Motion

Particles spiraling along magnetic field lines under geomagnetic confinement map to the algorithm's local search, which meticulously fine-tunes candidate solutions to enhance their quality and accelerate convergence toward the global optimum. During this dynamic process, the velocity of the charged particles fluctuates continuously. The velocity $v(t)$ of a particle in the t -th generation is defined as:

$$v(t) = Ce^{\frac{qB-u}{m}t} \quad (23)$$

where C represents the integration constant; the particle charge q , magnetic induction strength B , and particle mass m are treated as constants; and u denotes the damping coefficient.

2.4.3. Auroral Oval Walking

This strategy involves rapid movement around candidate solutions to identify the global optimum or superior local solutions. It corresponds to the global exploration phase, where the search space is scanned with larger step sizes to prevent stagnation. The mutation value Ao for the auroral oval walking and the barycenter position X_{avg} of high-energy particles are respectively defined as:

$$Ao = \text{Levy}(d) \times (X_{\text{avg}} - X(i, j)) + LB + r_1 \cdot (UB - LB)/2 \quad (24)$$

$$X_{avg} = \frac{1}{N} \sum_{i=1}^N X_i \quad (25)$$

where $X(i, j)$ represents the current position of the high-energy particle; $Levy(d)$ denotes a Lévy flight with a step size d ; and r_1 is a random value uniformly distributed in $[0, 1]$.

By integrating the rotational motion and auroral oval walking strategies, the updated calculation model for the particle positions is obtained:

$$X_{new}(i, j) = X(i, j) + r_2(W_1 \times v(t) + W_2 \times A_o) \quad (26)$$

$$W_1 = \frac{1 - e^{-2(\frac{t}{T})^4}}{1 + e^{-2(\frac{t}{T})^4}} \quad (27)$$

$$W_2 = e^{-8(\frac{t}{T})^3} \quad (28)$$

where $X_{new}(i, j)$ denotes the updated position of the i -th particle in the j -th dimension; T represents the maximum number of iterations; t is the current iteration index; and W_1, W_2 are the adaptive weight coefficients that balance the global and local search strategies throughout the optimization process.

2.4.4. Particle Collision

High-speed charged particles from the solar wind travel towards Earth and collide with each other upon impacting the atmosphere. These chaotic collisions can occur at any unpredictable angle. This stochastic collision mechanism enables the PLO algorithm to effectively escape local optima. The updated position $X_{new}(i, j)$ after a particle collision is formulated as:

$$X_{new}(i, j) = X(i, j) + \sin(\pi r_3) \cdot (X(i, j) - X(a, j)) \quad (29)$$

$$r_4 < \sqrt{\frac{t}{T}} \quad (30)$$

$$r_5 < 0.05 \quad (31)$$

where $X(a, j)$ represents a randomly selected particle from the population; r_3, r_4 , and r_5 are random values uniformly distributed in $[0, 1]$; and T denotes the maximum number of iterations. In each iteration, two random values, r_4 and r_5 , are stochastically generated. The collision-based position update is executed only if both r_4 and r_5 concurrently satisfy the constraints defined in Equations (30) and (31).

In this study, PLO is adopted to optimize the key hyperparameters of the PINN-based model, including the learning rate, the number of collocation points, and the regularization loss weight. The main reason for selecting PLO is that its rotational motion, auroral oval walking, and particle collision mechanisms provide a flexible search strategy for balancing global exploration and local exploitation. This characteristic is helpful for improving the adaptability of PINN hyperparameter selection. However, this study does not aim to prove the universal superiority of PLO over other optimization algorithms.

2.5. Model Evaluation Metrics

To quantitatively evaluate the estimation accuracy of the models, MAE, R^2 , RMSE, and MAPE are adopted as the core evaluation metrics in this study.

MAE represents the average of the absolute differences between the predicted and true values, where all individual differences are weighted equally. A smaller MAE value indicates higher prediction accuracy, and it is expressed as:

$$MAE = \frac{1}{n} \sum_{k=1}^n |SOH_{est}^k - SOH_{true}^k| \quad (32)$$

R^2 coefficient represents the model's ability to explain the variance in the data. A value closer to 1 indicates a better fitting performance of the model, and it is defined as:

$$R^2 = 1 - \frac{\sum_{k=1}^n (SOH_{est}^k - SOH_{true}^k)^2}{\sum_{k=1}^n (SOH_{mean} - SOH_{true}^k)^2} \quad (33)$$

RMSE intuitively reflects the degree of deviation between the predicted and true values, where a smaller value indicates superior performance. It is expressed as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (SOH_{est}^k - SOH_{true}^k)^2} \quad (34)$$

MAPE measures the relative error between the predicted and true values as a percentage. It eliminates the influence of scale, facilitating comparisons between data of different magnitudes. A smaller value indicates higher accuracy, and it is defined as:

$$MAPE = \frac{1}{n} \sum_{k=1}^n \left| \frac{SOH_{est}^k - SOH_{true}^k}{SOH_{true}^k} \right| \times 100\% \quad (35)$$

where n denotes the total number of samples; SOH_{est}^k represents the predicted SOH value for the k -th cycle in the sample; SOH_{true}^k is the corresponding ground truth for that cycle; and SOH_{mean} refers to the average of all true SOH values across the dataset.

3. Dataset and Feature Extraction

3.1. Computational Environment

All experiments were conducted on a desktop workstation equipped with an Intel® Core™ i5-10200H CPU (@ 2.40 GHz, 4 cores, 8 logical processors), 16 GB of RAM, and an NVIDIA GeForce GTX 1650 GPU with 4 GB of dedicated video memory. The system operated on Windows 10 (version 10.0.19045). All simulations and analyses in this and subsequent sections were implemented using MATLAB 2025b. The main MATLAB toolboxes used in this study included the Deep Learning Toolbox and Statistics and Machine Learning Toolbox. Other procedures, including VMD, PLO optimization, and evaluation metrics, were implemented using custom MATLAB scripts.

3.2. CALCE Dataset

In this study, battery aging datasets provided by the CALCE at the University of Maryland were employed for model validation. Specifically, four battery cells labeled CS2-35, CS2-36, CS2-37, and CS2-38 were selected as experimental subjects. These cells feature a lithium cobalt oxide cathode with a nominal capacity of 1.1 Ah. The degradation experiments were conducted following a standard Constant Current-Constant Voltage charging protocol. During the charging phase, the cells were initially charged at a constant current of 0.5 C until the terminal voltage reached a cutoff of 4.2 V. Subsequently, the system maintained a constant voltage of 4.2 V until the charging current decreased to below 0.05 A. For the discharge phase, a constant current of 1 A was applied until the discharge termination voltage dropped to 2.7 V. This cyclic aging process was performed repeatedly until the batteries reached their end-of-life criteria.

3.2.1. HF Extraction

Multiple HFs were extracted from the charging and discharging curves of four batteries from the CALCE dataset. Due to the inconsistent variation trends among different HFs, it is challenging to directly determine their correlation with battery capacity. To further

quantify the relationship and select the most effective HFs, this study employs the Pearson and Spearman correlation coefficients. Their calculation formulas are as follows:

$$Pearson = \frac{E(\mathbf{XY}) - E(\mathbf{X})E(\mathbf{Y})}{\sqrt{E(\mathbf{X}^2) - E^2(\mathbf{X})}\sqrt{E(\mathbf{Y}^2) - E^2(\mathbf{Y})}} \quad (36)$$

$$Spearman = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} \quad (37)$$

where $\mathbf{X}$ and $\mathbf{Y}$ represent the sample populations of health features and capacity, respectively. A higher absolute value of the Pearson or Spearman coefficient indicates a stronger correlation. An absolute value of 1 signifies that the two variables are perfectly correlated. The Pearson coefficient is most appropriate when variables consist of continuous data, follow a normal distribution, and exhibit a linear relationship. However, if any of these conditions are not met, the Spearman coefficient is more efficient. Therefore, both methods are utilized here for a comprehensive screening of health features.

Before health feature extraction, the raw cycling data of the CALCE batteries were first preprocessed. According to the voltage, current, and time information, the raw data were divided into charging and discharging datasets, and invalid cycles with missing records or incomplete charge/discharge processes were removed. Since the health features used in this study are mainly extracted from the charging process, the cleaned charging data were used for subsequent feature extraction. Finally, five HFs were selected for analysis: HF1, the accumulated charging quantity over the whole charging process; HF2, the charging time interval from 3.8 V to 4.0 V; HF3, the time from the beginning of charging to 4.2 V; HF4, the accumulated charging quantity from the beginning of charging to 4.2 V; and HF5, the maximum IC peak extracted from the dQ/dV curve.

For each charging cycle, the sampled relative time, terminal voltage, and charging current were used to determine the characteristic voltage points. For a given voltage threshold, such as 3.8 V, 4.0 V, or 4.2 V, the corresponding sampling point was identified as the first point at which the terminal voltage reached or exceeded the specified threshold. Based on the cleaned charging data, HF1 was obtained by calculating the accumulated charging quantity over the entire charging process using trapezoidal integration of the current-time curve. Its unit is A·s, and it can be converted to Ah by dividing by 3600. HF2 was calculated as the time difference between the points at which the terminal voltage reached 3.8 V and 4.0 V. HF3 was defined as the elapsed time from the beginning of charging to the moment when the terminal voltage reached 4.2 V. HF4 was obtained by calculating the accumulated charging quantity from the beginning of charging to the 4.2 V point, also using trapezoidal integration. HF5 was extracted from the incremental capacity (IC) curve. First, the charging data before 4.2 V were selected, and the relative time was converted from seconds to hours. Then, the accumulated charging capacity was calculated from the charging current, and the voltage sequence was resampled with a voltage interval of 0.01 V. The IC curve was subsequently obtained using the finite-difference method with a difference step of one. To reduce fluctuations caused by numerical differentiation, a Kalman filter was applied to smooth the calculated IC curve. The process noise covariance, measurement noise covariance, and initial estimation error covariance were set to 0.01, 0.05, and 0.01, respectively. Finally, the maximum value of the smoothed IC curve was selected as HF5, with the unit of Ah/V. During feature extraction, cycles without the required voltage threshold points or with non-finite IC values were regarded as invalid and removed. After the five health feature sequences were obtained, they were used for correlation analysis and subsequent VMD-KPCA-based feature construction for SOH estimation.

The evolutionary trends of the five HFs and the SOH are illustrated in Figure 2, while Table 1 quantifies this correlation using Pearson and Spearman correlation coefficients. Among them, to facilitate a direct comparison of their evolutionary trends, all HFs shown in Figure 2 are normalized using the following normalization formula:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (38)$$

where x^* denotes the normalized value, while $x_{\max}$ and $x_{\min}$ represent the maximum and minimum values of the original data, respectively.

As shown in Figure 2, all HFs show clear degradation trends throughout their lifetime. Specifically, the trajectories of HF1–HF4 exhibit good consistency and smoothness, and can effectively track the macro-degradation trend of battery capacity. In contrast, although HF5 follows a similar downward trend, it shows noticeable localized fluctuations and abrupt spikes. This indicates that HF1–HF4 are relatively stable in reflecting global aging, whereas HF5 is more susceptible to localized noise, which motivates the subsequent dimensionality reduction for effective feature aggregation.

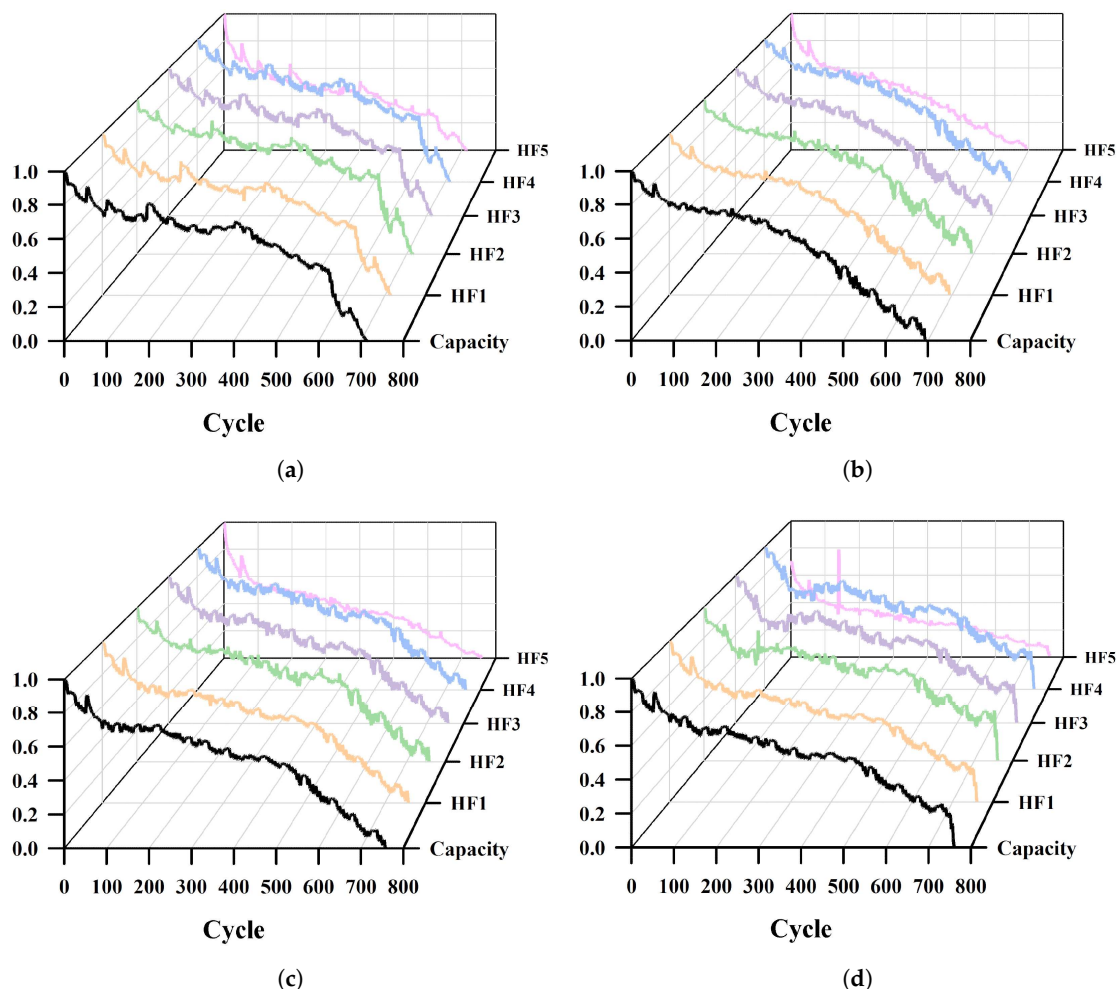


Figure 2. Normalized trends of capacity and five selected health features (HFs): (a) Battery CS2-35. (b) Battery CS2-36. (c) Battery CS2-37. (d) Battery CS2-38.

Table 1. Correlation analysis of the extracted HF5.

Battery	Coefficient	HF1	HF2	HF3	HF4	HF5
CS2-35	Pearson	0.9983	0.9945	0.9967	0.9967	0.9658
	Spearman	0.9960	0.9917	0.9941	0.9941	0.9884
CS2-36	Pearson	0.9988	0.9959	0.9976	0.9976	0.9803
	Spearman	0.9981	0.9949	0.9972	0.9971	0.9957
CS2-37	Pearson	0.9982	0.9950	0.9968	0.9968	0.9772
	Spearman	0.9968	0.9914	0.9953	0.9952	0.9924
CS2-38	Pearson	0.9975	0.9805	0.9792	0.9792	0.9614
	Spearman	0.9969	0.9734	0.9707	0.9708	0.9911

Table 1 further quantifies the correlation metrics. Both Pearson and Spearman coefficients for HF1–HF4 exceed 0.97, indicating their strong correlation with battery capacity degradation. In contrast, HF5 exhibits slightly lower Pearson coefficients due to localized fluctuations, but its Spearman coefficients remain above 0.98. Overall, all five HFs show strong correlation with SOH and are therefore used as inputs for the subsequent SOH prediction framework.

3.2.2. VMD Decomposition

Since the HFs extracted in the previous step still contain random noise, VMD is introduced to process the five HFs. Taking HF1 of battery CS2-35 as an example, the resulting time-domain waveforms and frequency distributions are shown in Figure 3 and Figure 4, respectively.

The VMD parameters were selected based on the decomposition quality and frequency separation of the IMFs. The decomposition number K determines the number of extracted modes. A small K may lead to under-decomposition, whereas an excessively large K may introduce redundant modes. In this study, $K = 6$ was selected because it can separate the main degradation trend and multi-scale fluctuation components of the HFs. The penalty factor α controls the bandwidth of each mode. Through repeated testing, $\alpha = 2235$ was adopted because it produced stable IMF components with clear frequency separation, as shown in Figures 3 and 4. Additionally, the noise tolerance τ is set to 0 and the convergence threshold ϵ is 10^{-7} to ensure decomposition accuracy.

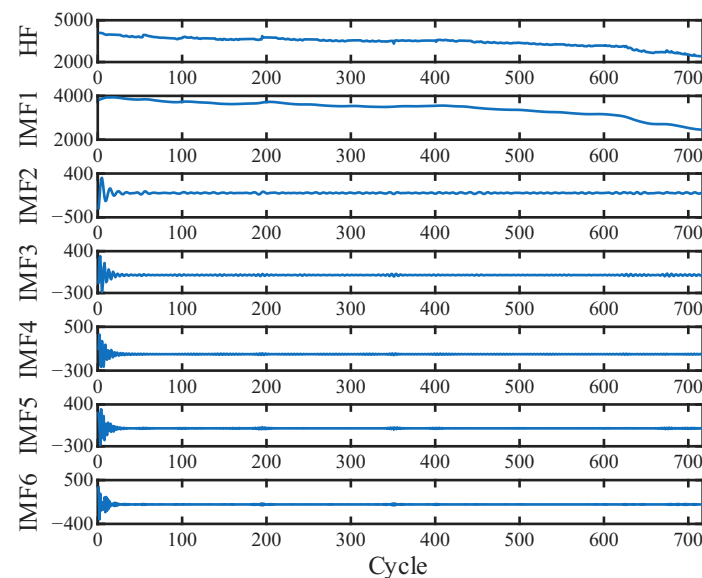
**Figure 3.** VMD decomposition results of HF1 for battery CS2-35.

Figure 3 displays the six IMF components obtained from the VMD decomposition of HF1 for battery CS2-35. It can be observed that IMF1 captures the dominant degradation trend of HF1, with its amplitude generally decreasing as the cycle number increases. This reflects the overall degradation behavior of the battery. In contrast, IMF2 to IMF6 represent fluctuation components at different scales. These components mainly describe local variations and noise-related fluctuations in the HF sequence, indicating that VMD can decompose the original HF into a dominant trend component and several multi-scale fluctuation components.

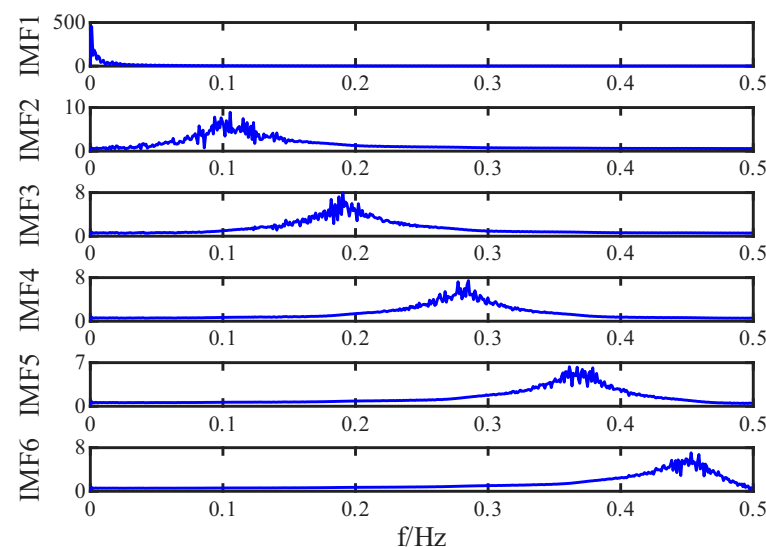


Figure 4. Frequency spectra of the IMF components obtained from the VMD decomposition of HF1 for battery CS2-35.

Figure 4 further illustrates the frequency characteristics of each IMF. The energy of IMF1 is mainly concentrated in the extremely low-frequency band near 0 Hz, corresponding to the slow degradation trend of the battery. The dominant frequency bands of IMF2 to IMF6 gradually shift toward higher frequencies. The relatively separated frequency distributions indicate that VMD provides effective frequency separation and helps alleviate the mode mixing problem.

The above analysis demonstrates that VMD can separate the dominant degradation trend from the multi-scale fluctuation components within the battery HFs. This provides more structured decomposed features for subsequent dimensionality reduction using KPCA.

3.2.3. KPCA Dimensionality Reduction

To reduce the redundancy of the 30-dimensional features obtained after VMD and to capture the nonlinear correlations among the decomposed components, KPCA with a polynomial kernel function (kernel parameter $d = 10$) was employed for nonlinear mapping and dimensionality reduction. Taking battery CS2-35 as an example, as shown in Figure 5, the VMD-derived feature matrix was transformed by KPCA to obtain several principal components. The cumulative contribution rate of the first principal component (PC1) exceeds 95%, indicating that PC1 retains most of the dominant information contained in the original VMD-based feature space. Therefore, PC1 was defined as the improved health feature (IHF) and used as the input of the subsequent PLO-PINN model. As summarized in Table 2, the cumulative contribution rates of the top three principal components for the CALCE battery datasets, including CS2-35, CS2-36, CS2-37, and CS2-38, further support the selection of PC1 as the IHF for SOH estimation.

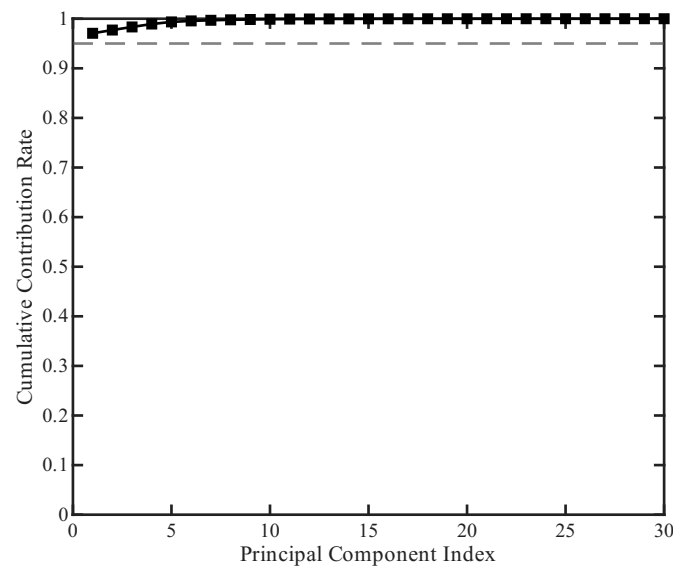


Figure 5. cumulative contribution rates of the principal components for battery CS2-35. The gray dotted line indicates the 95% cumulative contribution rate threshold.

Table 2. The contribution rates of the first three principal components.

Principal Component	CS2-35	CS2-36	CS2-37	CS2-38
PC1	0.9706	0.9987	0.9980	0.9658
PC2	0.0065	0.0008	0.0012	0.0077
PC3	0.0062	0.0002	0.0003	0.0072
Cumulative (PC1)	97.06%	99.87%	99.80%	96.58%

4. Model Experiments and Discussion of Results

4.1. Test Setting

In this section, the SOH estimation performance of the proposed VMD-KPCA-PLO-PINN model is evaluated using the University of Maryland CALCE battery dataset (CS2 series). Three key hyperparameters, namely the learning rate, the number of degradation-informed collocation points, and the degradation-informed regularization loss weight, are automatically optimized by the PLO algorithm. The search ranges of these parameters are set to $[1 \times 10^{-5}, 1 \times 10^{-2}]$, $[1000, 5000]$, and $[0.1, 0.6]$, respectively. For the PLO algorithm, the population size is set to 6, and the maximum number of iterations is set to 5. All models in this study are implemented in MATLAB R2025b. The Adam optimizer is employed to update the learnable parameters, with the maximum number of training epochs set to 2000. Finally, four evaluation metrics, including MAE, R^2 , RMSE, and MAPE, are used to quantitatively assess the SOH estimation performance of each model.

4.2. 70% Training Proportion

To demonstrate the superiority of the proposed method, three comparative models are established: VMD-KPCA-BP, VMD-KPCA-LSTM, and VMD-KPCA-PINN. For fairness, all four models (including the proposed one) adopt the same VMD-KPCA feature processing workflow. Specifically, the 1-dimensional principal component obtained after KPCA dimensionality reduction is used as the input feature to ensure the objectivity of the comparison. The training and testing sets are divided at a 7:3 ratio, meaning that the first 70% of the cycle data is used for training, while the remaining 30% is used to evaluate the prediction performance of the models on the testing interval. The quantitative error metrics for the

four battery samples (CS2-35, CS2-36, CS2-37, and CS2-38) are summarized in Table 3, and the SOH prediction comparison curves are shown in Figure 6a–d.

Table 3. SOH estimation performance comparison of different models for different battery samples.

Battery	Model	MAE	R^2	RMSE	MAPE (%)
CS2-35	VMD-KPCA-BP	0.0332	0.6664	0.0416	4.84
	VMD-KPCA-LSTM	0.0175	0.8894	0.0240	2.59
	VMD-KPCA-PINN	0.0145	0.9254	0.0197	2.14
	VMD-KPCA-PLO-PINN	0.0073	0.9807	0.0100	1.07
CS2-36	VMD-KPCA-BP	0.0231	0.6216	0.0260	3.20
	VMD-KPCA-LSTM	0.0118	0.8892	0.0141	1.63
	VMD-KPCA-PINN	0.0082	0.9419	0.0102	1.13
	VMD-KPCA-PLO-PINN	0.0055	0.9669	0.0077	0.74
CS2-37	VMD-KPCA-BP	0.0263	0.5578	0.0303	3.61
	VMD-KPCA-LSTM	0.0110	0.9167	0.0131	1.51
	VMD-KPCA-PINN	0.0111	0.9236	0.0126	1.48
	VMD-KPCA-PLO-PINN	0.0040	0.9833	0.0059	0.53
CS2-38	VMD-KPCA-BP	0.0105	0.8599	0.0131	1.37
	VMD-KPCA-LSTM	0.0071	0.9277	0.0094	0.93
	VMD-KPCA-PINN	0.0066	0.9354	0.0089	0.87
	VMD-KPCA-PLO-PINN	0.0043	0.9705	0.0060	0.55

According to the quantitative metrics in Table 3, the proposed VMD-KPCA-PLO-PINN method achieves the best performance across all four batteries. Taking CS2-37 as an example, the R^2 of the proposed method reaches 0.9833, which represents an improvement of 76.3%, 7.3%, and 6.5% compared to VMD-KPCA-BP ($R^2 = 0.5578$), VMD-KPCA-LSTM ($R^2 = 0.9167$), and VMD-KPCA-PINN ($R^2 = 0.9236$), respectively. Furthermore, the RMSE is reduced to 0.0059, which is 80.5%, 55.0%, and 53.2% lower than those of the three comparative models. Across the four batteries, the proposed method achieves an average R^2 of 0.9754, an average RMSE of only 0.0074, an average MAE of 0.0053, and an average MAPE as low as 0.72%, demonstrating significant superiority over the three contrastive methods.

As intuitively illustrated in Figure 6, the prediction curves of the VMD-KPCA-PLO-PINN model show the tightest fit with the actual SOH curves across the testing sets of all four batteries. The model effectively tracks the real degradation trajectory throughout the entire process, maintaining high precision even during the deep degradation stage at the end of the testing set. The VMD-KPCA-PINN model achieves the second-best performance, exhibiting minor deviations in regions marked by localized capacity regeneration. The remaining two models show significantly poorer predictive performance.

The above results indicate that the introduction of degradation-informed regularization enables the PINN-based model to better incorporate the empirical degradation trend into the SOH estimation process compared with pure data-driven methods such as BP and LSTM, thereby improving prediction accuracy. In addition, compared with the PINN without hyperparameter optimization, the PLO algorithm adaptively searches for a suitable parameter combination in the hyperparameter space, which further enhances the fitting ability and prediction performance of the PINN-based model. Moreover, the consistent performance observed across the four CALCE CS2 battery cells suggests that the proposed method exhibits good adaptability among cells from the same dataset.

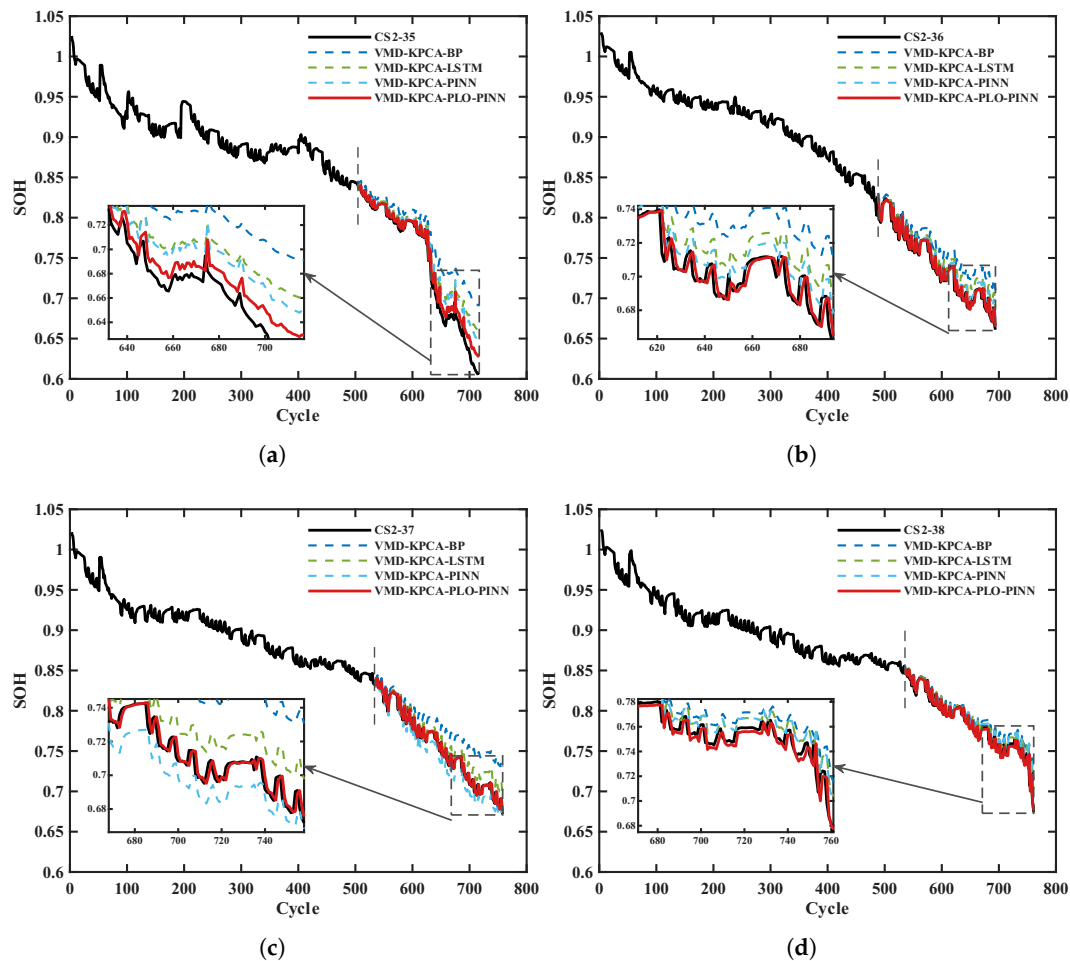


Figure 6. State of Health (SOH) prediction results: (a) Battery CS2-35. (b) Battery CS2-36. (c) Battery CS2-37. (d) Battery CS2-38. The gray dotted line indicates the boundary between the training and testing sets.

4.3. 50% Training Proportion

In practical BMS applications, the accumulation of battery degradation data often requires long operational cycles. Therefore, scenarios with limited available training data are common in engineering applications. To further examine the prediction performance of the proposed VMD-KPCA-PLO-PINN method under a reduced training proportion, the training-to-testing data ratio is adjusted from 7:3 to 5:5. The experiments in this section were conducted on the CS2-35, CS2-36, CS2-37, and CS2-38 batteries. The quantitative metrics are summarized in Table 4, and the results are illustrated in Figure 7a–d.

Table 4. Performance comparison of the proposed model under different training proportions.

Battery	Training:Test	MAE	R^2	RMSE	MAPE (%)
CS2-35	7:3	0.0073	0.9807	0.0100	1.07
	5:5	0.0065	0.9841	0.0105	0.93
CS2-36	7:3	0.0055	0.9669	0.0077	0.74
	5:5	0.0085	0.9753	0.0109	1.11
CS2-37	7:3	0.0040	0.9833	0.0059	0.53
	5:5	0.0056	0.9856	0.0073	0.72
CS2-38	7:3	0.0043	0.9705	0.0060	0.55
	5:5	0.0037	0.9870	0.0053	0.46

According to the quantitative metrics in Table 4, the R^2 values for CS2-35, CS2-36, CS2-37, and CS2-38 under the 5:5 split are 0.9841, 0.9753, 0.9856, and 0.9870, respectively. The corresponding MAE values are 0.0065, 0.0085, 0.0056, and 0.0037. In addition, all RMSE values remain below 0.011, and the MAPE values are controlled within 1.2%. As shown in Figure 7, the predicted SOH curves generally follow the actual degradation trajectories of the four tested batteries. These results indicate that the proposed VMD-KPCA-PLO-PINN model can achieve stable prediction accuracy on the tested CALCE cells under the 5:5 data partition.

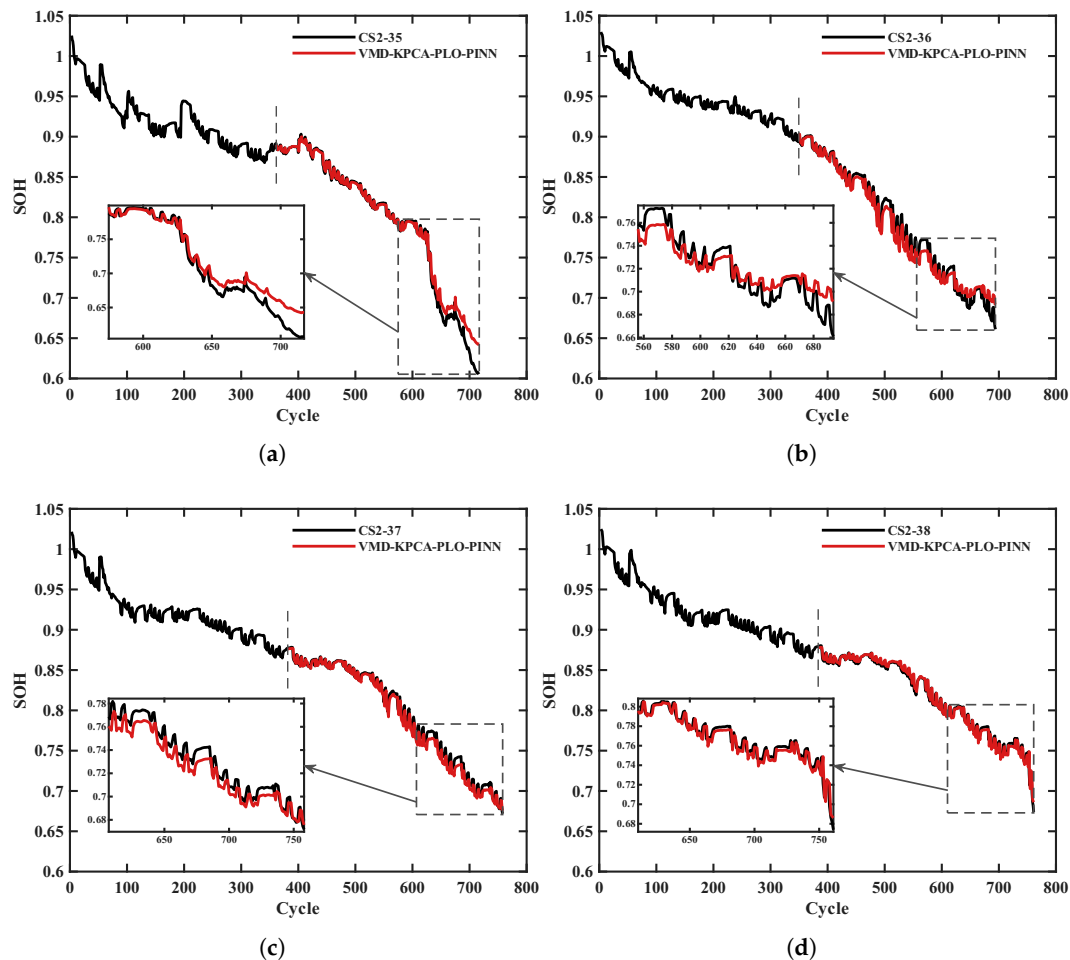


Figure 7. SOH prediction results: (a) Battery CS2-35. (b) Battery CS2-36. (c) Battery CS2-37. (d) Battery CS2-38. The gray dotted line indicates the boundary between the training and testing sets.

It should also be noted that the 5:5 partition should not be simply interpreted as a stricter proof of broad generalization ability. Compared with the 7:3 split, the testing interval in the 5:5 split starts earlier and covers a broader part of the degradation trajectory. In this interval, the capacity degradation trend is relatively smoother and more monotonic than that in the terminal aging stage. By contrast, the 7:3 split concentrates the testing samples more heavily in the later degradation stage, where nonlinear capacity fade and local fluctuations may increase the prediction difficulty.

Therefore, the comparable results between the 7:3 and 5:5 partitions indicate that the proposed framework maintains stable prediction performance under the tested CALCE data partitions. However, since both experiments are conducted on the same CS2-series dataset under similar operating conditions, these results should be regarded as preliminary evidence of prediction stability within the tested dataset, rather than sufficient evidence of universal generalization capability. Further validation using additional public datasets, different battery chemistries, temperatures, C-rates, and operating conditions is still required.

5. Conclusions

Based on the systematic experiments conducted on the four CALCE battery cells, the following core conclusions can be drawn.

First, under the 7:3 standard training-to-testing partition, the proposed VMD-KPCA-PLO-PINN framework achieves superior prediction performance compared with the three benchmark models, namely VMD-KPCA-BP, VMD-KPCA-LSTM, and VMD-KPCA-PINN, across all evaluation metrics. This demonstrates the effectiveness of integrating VMD-KPCA feature processing, PLO-based hyperparameter optimization, and PINN-based degradation-informed regularization.

Second, under the 5:5 training-to-testing partition, the proposed framework still maintains high estimation accuracy, with the average R^2 reaching 0.983 and the average RMSE limited to 0.0085. This result indicates that the proposed method can maintain relatively stable prediction performance when the amount of training data is reduced, showing its potential under limited-data conditions.

Third, the consistent evaluation results across the four CALCE cells indicate that the proposed framework can effectively track SOH degradation trends for different cells within the same battery series. This suggests that the proposed method has preliminary potential for battery SOH estimation under controlled experimental conditions.

Overall, this study provides a degradation-informed regularized framework for lithium-ion battery SOH estimation. The results demonstrate that the coordinated integration of VMD-KPCA feature processing, PLO-based hyperparameter optimization, and PINN-based regularized prediction can improve SOH estimation performance on the tested CALCE cells.

6. Limitations and Future Work

Several limitations should be acknowledged. First, the degradation constraint used in this study is an empirical trend-informed regularization term rather than a strict first-principles electrochemical governing equation. Since the input variable is the IHF obtained after VMD-KPCA transformation, its physical meaning differs from direct time, cycle number, or electrochemical state variables. Future work will investigate more rigorous constraints based on measurable physical variables and electrochemical aging mechanisms.

Second, the validation scope of this study is still limited. All experiments were conducted on four CS2-series batteries from the CALCE dataset under similar operating protocols. Therefore, the current results mainly demonstrate the effectiveness of the proposed framework on the tested CALCE cells, while its generalization ability under different datasets, battery chemistries, temperatures, C-rates, and operating conditions still requires further validation.

In addition, the extracted HFs rely on relatively complete charging curves under the CC-CV protocol. In practical EV or BESS applications, partial charging, variable current profiles, temperature variation, and measurement noise may affect the availability and reliability of some HFs. Future work will further explore adaptive feature extraction, cross-dataset testing, repeated experiments, statistical analysis, and comparisons with other optimization algorithms such as PSO, GA, WOA, and GWO.

Author Contributions: Conceptualization, R.D. and G.L.; methodology, Z.Y., G.L. and Q.D.; software, Z.Y.; validation, Z.Y.; formal analysis, Z.Y.; investigation, Z.Y. and Q.D.; resources, R.D.; data curation, Z.Y.; writing—original draft preparation, Z.Y.; writing—review and editing, R.D., G.L., Z.Y. and Q.D.; visualization, Z.Y.; funding acquisition, G.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (NSFC) under grant No. 52207233. The APC was funded by NSFC.

Data Availability Statement: The battery dataset analyzed in this study was obtained from the Center for Advanced Life Cycle Engineering (CALCE) at the University of Maryland and is publicly available at <https://calce.umd.edu/battery-data> (accessed on 13 April 2026). The processed health features and model prediction codes generated during this study are available from the corresponding author upon reasonable request.

Acknowledgments: The author would like to thank the reviewers for their valuable comments and guidance. Concurrently, the author deeply appreciates the support provided by the Hubei Key Laboratory for High-efficiency Utilization of Solar Energy and Operation Control of Energy Storage System.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BMS	battery management system
CALCE	Center for Advanced Life Cycle Engineering
HF	health features
IHF	improved health feature
IMFs	intrinsic mode functions
KPCA	kernel principal component analysis
LIBs	lithium-ion batteries
MAE	mean absolute error
MAPE	mean absolute percentage error
PINN	physics-informed neural network
PLO	polar lights optimizer
RMSE	root mean square error
SOH	state of health
VMD	variational mode decomposition

References

- Li, Q.; Xue, W. A review of feature extraction toward health state estimation of lithium-ion batteries. *J. Energy Storage* **2025**, *112*, 115453. [\[CrossRef\]](#)
- Song, Y.; Liu, D.; Liao, H.; Peng, Y. A hybrid statistical data-driven method for on-line joint state estimation of lithium-ion batteries. *Appl. Energy* **2020**, *261*, 114408. [\[CrossRef\]](#)
- Man, Q.; Shen, H.; Wei, C.; Xi, B.; Xiong, S.; Feng, J. Rational design of F, N-rich artificial interphase via chemical prelithiation initiation strategy enabling high coulombic efficiency and stable micro-sized SiO anodes. *J. Energy Chem.* **2024**, *92*, 224–232. [\[CrossRef\]](#)
- Elmahallawy, M.; Elfouly, T.; Alouani, A.; Massoud, A.M. A Comprehensive Review of Lithium-Ion Batteries Modeling, and State of Health and Remaining Useful Lifetime Prediction. *IEEE Access* **2022**, *10*, 119040–119070. [\[CrossRef\]](#)
- Gabbar, H.A.; Othman, A.M.; Abdussami, M.R. Review of Battery Management Systems (BMS) Development and Industrial Standards. *Technologies* **2021**, *9*, 28. [\[CrossRef\]](#)
- Weddington, J.; Niu, G.; Chen, R.; Yan, W.; Zhang, B. Lithium-ion battery diagnostics and prognostics enhanced with Dempster-Shafer decision fusion. *Neurocomputing* **2021**, *458*, 440–453. [\[CrossRef\]](#)
- Li, C.; Yang, L.; Li, Q.; Zhang, Q.; Zhou, Z.; Meng, Y.; Zhao, X.; Wang, L.; Zhang, S.; Li, Y.; et al. SOH estimation method for lithium-ion batteries based on an improved equivalent circuit model via electrochemical impedance spectroscopy. *J. Energy Storage* **2024**, *86*, 111167. [\[CrossRef\]](#)
- Wu, Y.; Rao, B.; Tian, J.; Du, J.; Jiang, J. SSA-BiLSTM Model-Based SOH Estimation for Lithium-Ion Batteries. *Energies* **2026**, *19*, 1499. [\[CrossRef\]](#)
- Hu, X.; Xu, L.; Lin, X.; Pecht, M. Battery lifetime prognostics. *Joule* **2020**, *4*, 310–346. [\[CrossRef\]](#)
- Lipu, M.H.; Ansari, S.; Miah, M.S.; Meraj, S.T.; Hasan, K.; Shihavuddin, A.; Hannan, M.; Muttaqi, K.M.; Hussain, A. Deep learning enabled state of charge, state of health and remaining useful life estimation for smart battery management system: Methods, implementations, issues and prospects. *J. Energy Storage* **2022**, *55*, 105752. [\[CrossRef\]](#)

11. Alsuwian, T.; Ansari, S.; Ammirul Atiqi Mohd Zainuri, M.; Ayob, A.; Hussain, A.; Hossain Lipu, M.; Alhawari, A.R.; Almawgani, A.; Almasabi, S.; Taher Hindi, A. A review of expert hybrid and co-estimation techniques for SOH and RUL estimation in battery management system with electric vehicle application. *Expert Syst. Appl.* **2024**, *246*, 123123. [\[CrossRef\]](#)
12. Kurucan, M.; Özbaltan, M.; Yetgin, Z.; Alkaya, A. Applications of artificial neural network based battery management systems: A literature review. *Renew. Sustain. Energy Rev.* **2024**, *192*, 114262. [\[CrossRef\]](#)
13. Yi, Y.; He, X.; Zhu, Y.; Wang, J.; Chen, Z.; Li, H. Hybrid Harris hawks and particle swarm optimized kernel extreme learning machine for battery SOH estimation. *J. Power Sources* **2026**, *680*, 240212. [\[CrossRef\]](#)
14. Liu, M.; Lu, Z. A Data-Driven Framework for Lithium-Ion Battery SOH Estimation Using VMD-GRU Hybrid Approach with Multi-Scale Feature Analysis. *Energy Eng.* **2026**, *123*, 12. [\[CrossRef\]](#)
15. Lei, X.; Xie, F.; Wang, J.; Zhang, C. A review of lithium-ion battery state of health and remaining useful life estimation methods based on bibliometric analysis. *J. Traffic Transp. Eng. (Engl. Ed.)* **2024**, *11*, 1420–1446. [\[CrossRef\]](#)
16. Sulzer, V.; Mohtat, P.; Aitio, A.; Lee, S.; Yeh, Y.T.; Steinbacher, F.; Khan, M.U.; Lee, J.W.; Siegel, J.B.; Stefanopoulou, A.G.; et al. The challenge and opportunity of battery lifetime prediction from field data. *Joule* **2021**, *5*, 1934–1955. [\[CrossRef\]](#)
17. Guo, Z.; Li, Y.; Zhou, Z.; Zhang, C.; Yan, Z.; Chow, M.Y. A virtual-component-embedded equivalent circuit model for lithium-ion battery state estimation. *Int. J. Electr. Power Energy Syst.* **2025**, *170*, 110896. [\[CrossRef\]](#)
18. Wang, L.; Yang, T.; Hu, B. A Battery State-of-Health Estimation Method for Real-World Electric Vehicles Based on Physics-Informed Neural Networks. *IEEE Sens. J.* **2025**, *25*, 15577–15587. [\[CrossRef\]](#)
19. Park, J.; Jin, Y.; Kam, W.; Han, S. A practical semi-empirical model for predicting the SoH of lithium-ion battery: A novel perspective on short-term rest. *J. Energy Storage* **2024**, *96*, 112659. [\[CrossRef\]](#)
20. Li, Y.; Liu, K.; Foley, A.M.; Zülke, A.; Berecibar, M.; Nanini-Maury, E.; Van Mierlo, J.; Hoster, H.E. Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review. *Renew. Sustain. Energy Rev.* **2019**, *113*, 109254. [\[CrossRef\]](#)
21. Sun, X.; Wei, Z.; Li, Y.; Lu, H.; Liu, C. New Insights Into Estimating Lithium-Ion Battery Cell Health Leveraging Physics-Informed Machine Learning. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 1–15. [\[CrossRef\]](#)
22. Ma, Y.; Liu, Z.; Ye, S.; An, D.; Zhang, G. Generative-Enhanced Physics-Informed Neural Network for Lithium-Ion Battery State of Health Estimation. *IEEE Sens. J.* **2025**, *25*, 31162–31174. [\[CrossRef\]](#)
23. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-informed machine learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440. [\[CrossRef\]](#)
24. Raissi, M.; Perdikaris, P.; Karniadakis, G. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [\[CrossRef\]](#)
25. Jiang, C.; Zhong, T.; Choi, H.; Youn, B.D. Physics-informed Gaussian process probabilistic modeling with multi-source data for prognostics of degradation processes. *Reliab. Eng. Syst. Saf.* **2025**, *258*, 110893. [\[CrossRef\]](#)
26. Hofmann, T.; Hamar, J.; Rogge, M.; Zoerr, C.; Erhard, S.; Philipp Schmidt, J. Physics-Informed Neural Networks for State of Health Estimation in Lithium-Ion Batteries. *J. Electrochem. Soc.* **2023**, *170*, 090524. [\[CrossRef\]](#)
27. Wang, F.; Zhai, Z.; Zhao, Z.; Di, Y.; Chen, X. Physics-informed neural network for lithium-ion battery degradation stable modeling and prognosis. *Nat. Commun.* **2024**, *15*, 4332. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Singh, S.; Ebongue, Y.E.; Rezaei, S.; Birke, K.P. Hybrid Modeling of Lithium-Ion Battery: Physics-Informed Neural Network for Battery State Estimation. *Batteries* **2023**, *9*, 301. [\[CrossRef\]](#)
29. Chen, Y.; Huang, X.; He, Y.; Zhang, S.; Cai, Y. Edge-cloud collaborative estimation lithium-ion battery SOH based on MEWOA-VMD and Transformer. *J. Energy Storage* **2024**, *99*, 113388. [\[CrossRef\]](#)
30. Zhu, T.; Wang, S.; Fan, Y.; Hai, N.; Huang, Q.; Fernandez, C. An improved dung beetle optimizer- hybrid kernel least square support vector regression algorithm for state of health estimation of lithium-ion batteries based on variational model decomposition. *Energy* **2024**, *306*, 132464. [\[CrossRef\]](#)
31. Yuan, Z.; Tian, T.; Hao, F.; Li, G.; Tang, R.; Liu, X. A hybrid neural network based on variational mode decomposition denoising for predicting state-of-health of lithium-ion batteries. *J. Power Sources* **2024**, *609*, 234697. [\[CrossRef\]](#)
32. He, X.; Wang, H.; Jiang, Y.; Kang, J.; Zhu, J.; Wang, L. State of Health Estimation for Lithium-Ion Batteries Based on Global Feature Extraction and KPCA-ISSA-LSSVM Model. *J. Electrochem. Soc.* **2025**, *172*, 110525. [\[CrossRef\]](#)
33. Zheng, D.; Man, S.; Guo, X.; Ning, Y. Joint prediction of state of health and remaining useful life for lithium-ion batteries based on health features optimization and multi-model fusion. *Ionics* **2024**, *30*, 6239–6252. [\[CrossRef\]](#)
34. He, R.; Peng, T.; Zhang, X.; Chen, Z.; Yao, J.; Nazir, M.S.; Zhang, C. A novel hybrid model for state of health prediction in lithium batteries based on non-stationary transformers optimized by tree-structured Parzen estimator considering health factors. *Appl. Energy* **2026**, *402*, 127030. [\[CrossRef\]](#)
35. Ge, Y.; Zhang, C.; Wang, Y.; Chen, J.; Wang, Z.; Shahzad Nazir, M.; Peng, T. A novel hybrid model based on multiple influencing factors and temporal convolutional network coupling ReOSELN for wind power prediction. *Energy Convers. Manag.* **2024**, *313*, 118632. [\[CrossRef\]](#)

36. Yuan, C.; Zhao, D.; Heidari, A.A.; Liu, L.; Chen, Y.; Chen, H. Polar lights optimizer: Algorithm and applications in image segmentation and feature selection. *Neurocomputing* **2024**, *607*, 128427. [[CrossRef](#)]
37. Wang, H.; Chen, K.; Chen, Z.; Li, S.; Wang, J.; Wang, P.; Yang, L. State-of-health prediction of lithium-ion batteries: A novel PLO-transformer-LSTM framework with DRT feature engineering. *Comput. Chem. Eng.* **2026**, *205*, 109473. [[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.