This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2026.3686077

Physics-Informed Deep Gated Recurrent Unit Network for Lithium-Ion Battery State of Health Estimation Under Field Conditions

Laijin Luo, Yu Wang, *Senior Member*, *IEEE*, Yang Li, *IEEE Senior Member*, and Qiushi Cui, *IEEE Member*

***Abstract*—Accurate estimation of lithium-ion battery state of health (SOH) in electric vehicles (EVs) under real-world conditions is much more challenging than using well-designed laboratory cycling data due to unreliable SOH labeling and segmented charging behavior, and the results often lack interpretability. To address these issues, this paper proposes a physics-informed deep gated recurrent unit (PIDGRU) architecture for robust and interpretable SOH estimation without requiring explicit physical modeling. First, a modified inverse ampere-hour integral method is combined with the Bayesian estimator of abrupt, seasonality, and trend (BEAST) algorithm to estimate battery capacity and characterize SOH uncertainty. A universal feature extraction and selection framework is then developed to handle segmented EV charging data, utilizing a hybrid linear-nonlinear redundancy analysis to ensure an optimal input feature set. The PIDGRU integrates empirical degradation modeling and nonlinear dynamic degradation learning through a deep gated recurrent unit network, utilizing a deep hidden physics model (DeepHPM). A Bayesian inference uncertainty-constrained (BIUC) strategy is introduced to enhance training reliability and uncertainty quantification. Extensive evaluations on both in-vehicle and cross-vehicle datasets demonstrate that the proposed method achieves high accuracy, robustness, and generalizability, with the mean absolute error and root mean squared error consistently below 1.10% and 1.31%, respectively.**

***Index Terms*—Lithium-ion battery, state of health, electric vehicles, physics-informed deep gated recurrent unit, BEAST algorithm, Bayesian inference uncertainty constrained strategy.**

I. Introduction

With the global energy landscape transitioning rapidly, development of green, low-carbon, and renewable energy systems has been recognized as a priority direction for national energy strategies [1], [2]. Lithium-ion batteries (LIBs), as a representative energy storage technology, have been widely deployed in electric vehicles (EVs), smart grids, and consumer electronics, owing to their high energy density, long cycle life, and flexible system integration [3]. However, under highly complex and uncertain real-world operating conditions, LIBs are prone to nonlinear capacity degradation, power fade, and safety-critical failures [4]. As such, achieving an accurate battery state of health (SOH) assessment is essential but challenging, as it ensures the safety, reliability, and efficiency of battery energy storage systems.

Existing battery SOH estimation methods can be broadly categorized as model-based and data-driven approaches [5]. Model-based SOH methods typically rely on dynamic modeling to describe degradation trajectory mechanisms, including empirical models [6], electrochemical models [7], and equivalent circuit models (ECMs) [8]. Empirical models fit predefined mathematical functions (e.g., polynomial or exponential form) to capacity degradation trajectories based on time, cycles, or operational variables, offering simple and efficient SOH prediction [9]. However, they often lack physical interpretability. Electrochemical models aim to capture underlying degradation mechanisms using partial differential equations (PDEs), accounting for factors such as lithium transference number, lithium plating, and solid electrolyte interphase (SEI) growth [10]. While offering high fidelity, the inherent high computational complexities and difficulty in parameter identification severely limit their applicability in practical systems. To improve practicality, ECMs simulate battery dynamics through low-order RC and fractional-order circuit elements [11]. However, their parameters are sensitive to unseen variations in temperature and aging conditions during model identification, often causing nonlinear drift that reduces estimation accuracy.

Data-driven methods, on the other hand, learn degradation patterns directly from historical data without relying on explicit dynamic modeling [12]. Due to their flexibility and powerful capabilities for function approximation, data-driven methods have yielded satisfactory SOH estimation results [13]. In fact, the performance of data-driven methods primarily depends on two critical aspects: advanced feature engineering and the development of estimation algorithms. In general, when degradation features are informative and well-selected, lightweight machine learning algorithms such as multilayer perceptron (MLP) [14], support vector machine (SVM) [15], and Gaussian process regression (GPR) [16] can achieve competitive performance. Recently, deep learning methods, including convolutional neural networks (CNNs) [17], long short-term memory (LSTM) [18], and gated recurrent unit (GRU) [19], have been employed to capture temporal degradation behavior. However, these models often suffer from limited interpretability and high dependency on large, high-quality datasets.

This work was supported by the National Key R&D Program of China (Sci-Tech Innovation 2030 Agenda) (Grant No. 2025ZD0805405). (Corresponding author: Yu Wang)

Laijin Luo, Yu Wang, and Qiushi Cui are with School of Electrical Engineering, Chongqing University, Chongqing, China (e-mail: 20241101048@stu.cqu.edu.cn; yu_wang@cqu.edu.cn; qcui@cqu.edu.cn).

Yang Li is with the School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China (e-mail: yangli@ieee.org)

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2026.3686077

To overcome the limitations of traditional purely data-driven and model-based approaches, physics-informed neural networks (PINNs) have emerged as a promising architecture [20]. Aykol et al. [21] categorized PINNs into two types: sequential serial integrated PINNs (Type-A) and hybrid integrated PINNs (Type-B). In essence, physical and data-driven models of Type-A are explicitly decoupled and can be directly constructed by serially concatenating existing physical models with advanced machine learning components [22]. Therefore, in the short term, Type-A architecture represents a practical and relatively straightforward battery SOH estimation strategy. For example, Li et al. [23] proposed a PINN framework for estimating electrode-level states in lithium-ion batteries. Hofmann et al. [24] further integrated simulation data generated from the pseudo-two-dimensional model and vehicle field data. As a more promising yet challenging architecture, Type-B PINNs effectively incorporate a series of nonlinear governing equations from physics-based dynamic models with data-driven models to simulate and learn the complex behaviors of battery degradation. For instance, Wang et al. [25] developed an interpretable PINN-based SOH estimation approach, leveraging MLPs to model degradation features and dynamics, and integrating physical knowledge into the loss function for improved interpretability and accuracy. Inspired by population growth modeling, Wen et al. [26] proposed a Verhulst model-based method to capture the degradation trend and estimate battery SOH. A recent study summarized the learning framework for embedding three physical knowledge, including input space guidance, network structure design, and loss function constraints [27]. In particular, electrochemistry-thermal coupling models offer higher fidelity mechanisms [28], but embedding them into computable learning objectives remains challenging for field data. To improve in-situ battery life prediction, Zhang et al. [29] extracted physically meaningful features from voltage relaxation responses under mixed operating conditions.

Although several PINN-based SOH estimation models have been developed for specific applications aiming at improving interpretability, we notice that relevant studies for battery pack SOH estimation under field conditions of EVs remain limited and face several challenges.

1) The complexity of high-fidelity physical models limits the flexibility of integrating degradation dynamics into PINN architectures [10], [23]. Moreover, most existing PINNs employ MLPs [25], which are inadequate for modeling the temporal evolution of degradation features governed by PDEs.

2) Feature engineering approaches are mostly developed and demonstrated using laboratory-tested aging data, which do not generalize well to EV field data [18], [23]. Real-world charging data are often segmented due to irregular user behavior, which makes it difficult to extract reliable degradation-relevant features. Furthermore, SOH labels are not directly available in raw data and must be inferred using external algorithms or costly experiments.

3) Conventional feature selection methods often assess SOH-feature relationships from a single perspective, such as Pearson correlation, and fail to account for linear or nonlinear redundancy among features [13], [18], [19].

To overcome the above challenges, a novel SOH estimation framework for EV field conditions is proposed in this study. The main contributions are summarized as follows.

1) A physics-informed deep gated recurrent unit (PIDGRU) architecture without explicit physical modeling is proposed, which integrates the information of potential nonlinear dynamic degradation and temporal dependencies, and automatic differentiation (AutoDiff) to achieve accurate SOH estimation. A Bayesian inference uncertainty constrained (BIUC) training strategy is incorporated to enhance estimation reliability by adaptively regulating uncertainty during training.

2) A robust SOH labeling strategy for EV field conditions is proposed, combining a modified inverse ampere-hour integral approach with the Bayesian estimator of abrupt, seasonality, and trend (BEAST) algorithm to reduce label noise and enhance label reliability.

3) A universal feature extraction and selection method is presented for handling segmented EV charging data. A dual-perspective selection approach, leveraging both Pearson and distance correlation, is employed to reduce linear and nonlinear redundancy.

II. Analysis of Field Data and SOH Label Uncertainty

A. Description of Battery Field Data

To investigate the performance evolution of EV batteries under real-world operating conditions, this study adopts 29 months of charging data collected from 20 commercial EVs [30]. Each EV is operated in typical urban driving scenarios, with its battery pack composed of 90 cells connected in series and equipped with 32 temperature sensors. More detailed battery pack specifications are summarized in Table I. As shown in Fig. 1, the charging process is presented as a time-sequenced concatenation, including current, total voltage, SOC, maximum and minimum cell voltages, and maximum and minimum temperatures recorded per 8-second intervals. The vehicle adopts a multi-stage constant current charging scheme, with each stage initiated at a random SOC.

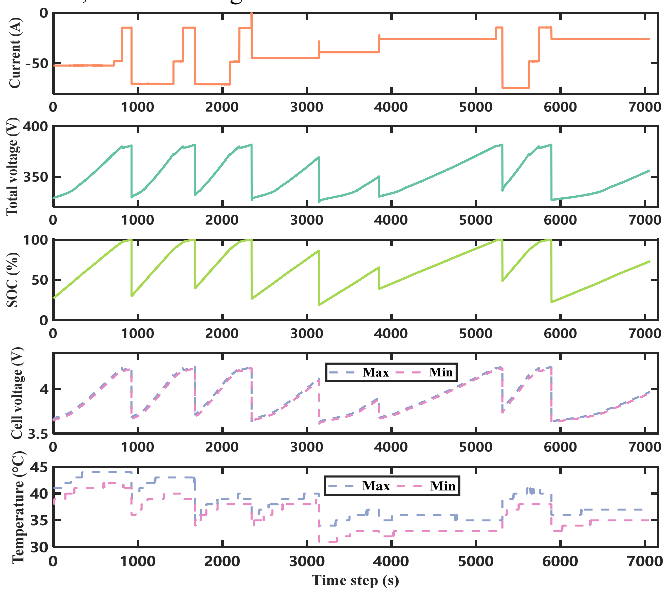


Fig. 1. Current, total voltage, SOC, cell voltage, and temperature profiles of a vehicle battery over multiple charge–discharge cycles.

TABLE I
SPECIFICATIONS OF THE BATTERY PACK

Items	Value
Battery type	NCM
Rated capacity (cell)	1.61 Ah
Nominal voltage (cell)	3.6V
Number of cells	90
Rated capacity (pack)	145 Ah
Nominal voltage (pack)	324 V

B. Uncertainty Modeling of SOH Labels Under Field Conditions

In battery aging studies, the definition of SOH based on maximum available capacity (MAC) is widely adopted due to its intuitive interpretation and broad applicability [2]. The SOH is mathematically defined as:

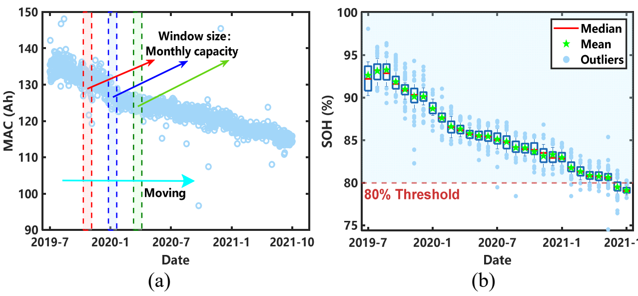
$$SOH_i = \frac{Q_i}{Q_{\text{rated}}} \tag{1}$$

where Q_i and Q_{rated} represent the MAC at the ith cycle and nominal capacity, respectively.

For real-world EVs, obtaining complete charge/discharge cycles is often impractical, which makes it difficult to directly measure the true MAC of the battery pack. Accordingly, based on field operational data, the inverse ampere-hour integral method is employed to estimate the MAC, calculated as:

$$Q_i = \frac{\int_{t_1}^{t_2} I(t)dt}{SOC_{t_2} - SOC_{t_1}} \tag{2}$$

where t_1 and t_2 denote the charging start and end times of a charging segment (t_1<t_2), SOC_{t_1} and SOC_{t_2} represent the corresponding battery SOC, and $I(t)$ is the battery charging current. As shown in Fig. 2(a), the derived MAC exhibits a clear overall declining trend, indicating progressive battery degradation over time. In addition, combined with (1), the monthly SOH statistics are illustrated in Fig. 2(b) using box plots, which highlight the non-stationarity of field data and the need for robust modeling approaches in SOH estimation.



(a) (b)

Fig. 2. Evolution of the derived MAC and SOH. (a) Illustration of the MAC monthly analysis with a moving window. (b) Monthly degradation trend of SOH shown by 29 boxplots, each representing the SOH distribution for a given month.

Therefore, the BEAST algorithm is employed to perform Bayesian inference on the MAC derived from the ampere-hour integral method [31]. It identifies potential seasonal or trend components, along with associated change points and their probabilities, to characterize the uncertainty in SOH degradation. The BEAST algorithm leverages the Bayesian model averaging (BMA) to combine multiple models into an average model, enabling joint inference of the MAC's probability distribution rather than a single deterministic value.

Specifically, the BEAST algorithm models the MAC time series using three components, namely, seasonal (S), trend (T), and noise (ε), and can be expressed as:

$$y(t) = S(t;\theta_S) + T(t;\theta_T) + \varepsilon(t) \tag{3}$$

where θ_S and θ_T are the parameter sets for seasonal change points (SCPs) and trend change points (TCPs), respectively, which are implicitly encoded within the seasonal and trend components; The noise term ε is assumed to follow a zero-mean Gaussian distribution with unknown variance σ^2.

By selecting P time nodes $\{\xi_p\}$ for p = 1, 2, …, P, the seasonal signal $S(t)$ can be approximated using a piecewise harmonic model comprising P + 1 segments. The pth segment is described by:

$$S(t) = \sum_{j=1}^{J_p}\left[a_{j,p}\cos\left(\frac{2\pi jt}{C}\right) + b_{j,p}\sin\left(\frac{2\pi jt}{C}\right)\right], \xi_{p-1} \le t < \xi_p \tag{4}$$

where J_p denotes the harmonic order of the pth segment, $a_{j,p}$ and $b_{j,p}$ are the coefficients of the corresponding cosine and sine terms, respectively. C represents the period of the seasonal component. Since the harmonic model is discontinuous across segments and C can be preset, it is necessary to estimate the following parameter set $\theta_S = \left\{\left(P, \xi_p, J_p, \{a_{j,p}, b_{j,p}\}_{j=1}^{J_p}\right) \mid p = 1, \cdots, P+1\right\}$ to fully characterize the piecewise seasonal harmonic curve.

Similar to the seasonal component, the trend $T(t)$ is modeled as a piecewise linear function with M time nodes $\{\tau_m\}$ for m = 1, 2, …, M, dividing $T(t)$ into M + 1 segments. Within the mth segment, $T(t)$ is a linear function of time, parameterized by a slope β_m and an intercept c_m, i.e.,

$$T(t) = \beta_m t + c_m, \quad \tau_{m-1} \le t < \tau_m \tag{5}$$

Therefore, the piecewise linear trend $T(t)$ is characterized by the parameter set $\theta_T = \left\{\left(M, \tau_m, \beta_m, c_m\right) \mid m = 1, \cdots, M+1\right\}$.

For clarity and structural interpretation, the two parameter sets θ_S and θ_T are reorganized into two groups, denoted by Ω and μ_Ω. The first group $\Omega = \{P, \xi_p, J_p, M, \tau_m\}$ defines the model structure, including the number of time nodes for the trend and seasonal components, as well as the harmonic orders of the seasonal terms. The second group determines the amplitude of the basis functions within a fixed model structure. In this way, (3) can be transformed into a general linear form:

$$y(t) = \mathcal{B}_\Omega(t)\mu_\Omega + \varepsilon(t) \tag{6}$$

where $\mathcal{B}_\Omega(t) = \{\, t, \cos(2\pi jt/C), \sin(2\pi jt/C)\}$ represents the basis function formed by the piecewise linear and harmonic functions in (4) and (5).

According to Bayes' theorem, the posterior probability is proportional to the product of the likelihood and the prior probabilities. Accordingly, (6) can be formulated as:

$$P(\mu_\Omega, \Omega, \sigma^2 \mid y) \propto P(y \mid \mu_\Omega, \Omega, \sigma^2) \cdot P(\mu_\Omega, \Omega, \sigma^2) \tag{7}$$

Due to the intractability of analytically solving the posterior probability, Markov Chain Monte Carlo (MCMC) sampling is

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2026.3686077

employed to generate sample chains $L\{\mu_{\Omega}^{(i)}, \Omega^{(i)}, \sigma^{2(i)}\}$ for posterior inference. The final estimate is obtained by averaging over these chains. The key parameters of the BEAST algorithm, including the seasonal period and time step, are selected based on both the characteristics of the field data and established practices in the literature [31]. Specifically, the vehicle dataset covers approximately 29 months, during which annual seasonal temperature variations influence battery degradation. Therefore, a period of one year is adopted to capture the annual cycle degradation trends. The time step is set to 1 month, which captures long-term degradation processes and seasonal trends while avoiding excessive computational overhead. Based on the above considerations and the analysis in Fig. 2, we employed 3 MCMC chains, each chain containing 8,000 samples with a sampling interval of 5. The detection results for the MAC are shown in Fig. 3. In addition, more demonstration examples for the BEAST algorithm toolbox in MATLAB and Python versions are available at https://github.com/zhaokg/Rbeast

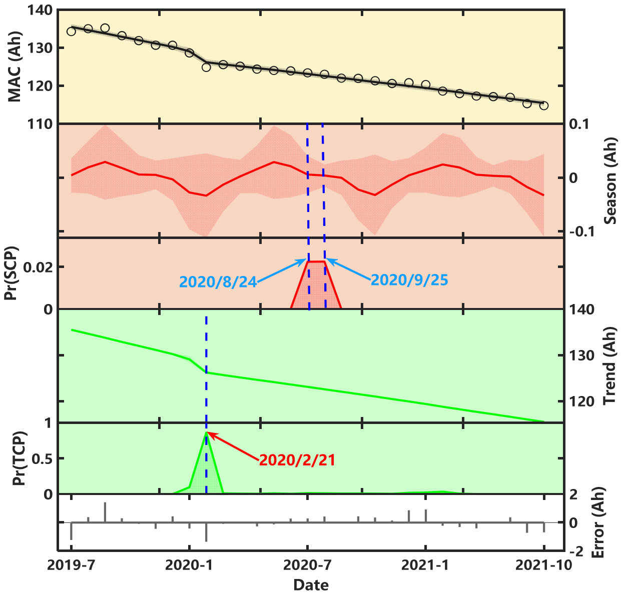


Fig. 3. Results of BEAST decomposition and change point detection, including original MAC data, the seasonal component and its uncertainty, the occurrence probability of seasonal change points "Pr(SCP)", the trend component, the occurrence probability of trend change points "Pr(TCP)", and the residual "Error". The envelopes around the fitted seasonal or trend signals represent 95% confidence intervals.

As shown in Fig. 3, the BEAST algorithm demonstrates strong capability in accurately identifying and separating the seasonal and trend degradation patterns of the battery MAC. In the seasonal component, two SCPs are detected on August 24, 2020, and September 25, 2020 (indicated by blue arrows). Although the associated occurrence probabilities are relatively low, the periodic fluctuations captured by the seasonal component closely match the multiple fluctuations observed in the MAC trajectory, highlighting the influence of seasonal variation on battery behavior. In the trend component, BEAST precisely detects an abrupt decline on February 21, 2020 (indicated by a red arrow) with a high probability of 0.86. This change point coincides with the main degradation phase of the MAC trajectory. Overall, these decomposition results reveal critical patterns in MAC evolution and provide a robust, interpretable foundation for subsequent feature engineering and SOH modeling. The seasonal component is used to select aging-related temperature features, while the slope of the degradation trend (β_k) and the trend change point (TCP) are incorporated into the regularization term of the model training objective.

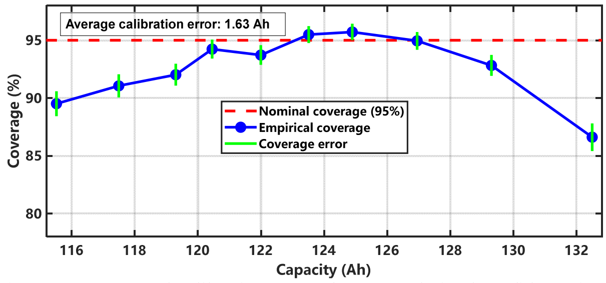


Fig. 4. Coverage and calibration error of BEAST-derived confidence intervals across different operating capacity ranges.

In addition, to evaluate the coverage and calibration of the SOH tags, we conducted empirical coverage and calibration error analyses, which used the confidence intervals derived from BEAST as uncertainty references. Fig. 4 shows the coverage of confidence intervals based on BEAST across different capacity ranges. It can be observed that beyond 80% of the capacity aging range (118 Ah-130 Ah), the empirical coverage exceeds 90%, with an average calibration error of 1.6 Ah, indicating that the uncertainty intervals are well calibrated throughout most of the aging process.

III. Methodology

A. Feature Engineering

Differing from laboratory settings, battery discharge behavior in real-world applications is typically uncontrollable and strongly influenced by user-specific driving habits [18]. In contrast, the charging process usually occurs under more stable and regular conditions, facilitating the reliable monitoring and transmission of current, voltage, and temperature data. Therefore, the charging curves are utilized for extracting aging features in this study.

1) Aging Feature Pool Construction

As illustrated in Fig. 5, feature extraction was performed on approximately 1,656 charging segments collected from a vehicle over a period of 29 months. It is observed that the starting voltages of these segmented charging curves vary randomly within the range of 310–370 V, and the voltage profile exhibits substantial variation across segments. These factors present significant challenges for consistently and systematically extracting aging features.

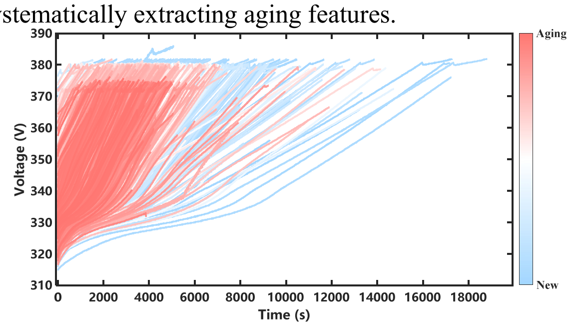


Fig. 5. Voltage−time profiles of battery charging data segments collected over a period of 29 months. Each curve represents a charging event, with color indicating the aging status, ranging from new (blue) to aged (red).

Based on this observation, further analysis reveals that the seasonal component of MAC identified in Section II-B is highly similar to the temperature feature curve, with a correlation coefficient of 0.61. This indicates that monthly temperature variations are a primary driver of the underlying seasonal pattern in MAC, as further supported by Fig. 6. To isolate seasonal effects and retain the long-term degradation trend, the voltage-related time series is decomposed using the BEAST algorithm. The extracted voltage trend components are then incorporated into the feature pool to enhance the model's ability to capture long-term SOH evolution. The MAC/SOH labels are defined and acquired at the cycle level, while voltage measurements form time-series data within each charge cycle, with all feature extraction procedures performed independently for each cycle.

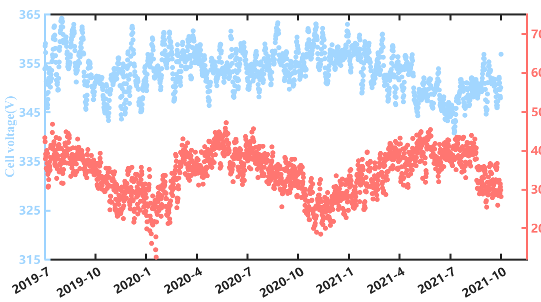


Fig. 6. Seasonal fluctuations in voltage and temperature.

To construct a universal feature pool, we extracted 18 features from key field signals. These features are derived from battery pack current and voltage, SOC, maximum cell voltage, minimum cell voltage, temperature, and cell-to-cell voltage differences. They include both direct statistical descriptors, such as mean, entropy, and standard deviation, and trend-based voltage descriptors used to capture long-term degradation-related evolution patterns. In particular, the voltage-related trend components are emphasized in this study as degradation-sensitive representations for segmented charging data under field conditions. A complete list of feature symbols and descriptions is provided in Table S1 of the Supplementary Material.

2) Feature Correlation Analysis

Traditional methods typically employ Pearson correlation analysis to identify aging features that exhibit strong linear relationships with SOH. However, Pearson correlation is limited to capturing the linear dependencies and fails to detect nonlinear dependencies, thereby constraining the comprehensiveness and effectiveness of feature selection. To address this limitation, distance correlation is introduced to evaluate both linear and nonlinear relationships between aging features and SOH. Distance correlation is defined based on distance covariance (dCov) and distance variance (dVar). Given a feature pool X and the corresponding battery pack SOH values Y, the distance correlation coefficient dCorr(X, Y) is calculated as:

$$\mathrm{dCorr}(X,Y)=\frac{\mathrm{dCov}(X,Y)}{\sqrt{\mathrm{dVar}(X)\mathrm{dVar}(Y)}} \tag{8}$$

Higher values of dCorr indicate stronger dependencies, particularly in capturing nonlinear relationships. Figs. 7(a) and (b) present the visual comparison of Pearson correlation (PCorr) and distance correlation, respectively.

From the first row of Figs. 7(a) and (b), it is observed that the identified aging features with highly correlated SOH are largely consistent between the two approaches, particularly the correlation coefficients of F4, F5, F13, and F14, all of which exceed 0.95. Conversely, several features, such as F7, F8, F16, and F17, that show weak linear correlation with SOH under the Pearson metric, yet are distinctly captured by distance correlation, which indicates the presence of the neglected nonlinear dependencies between these features and SOH.

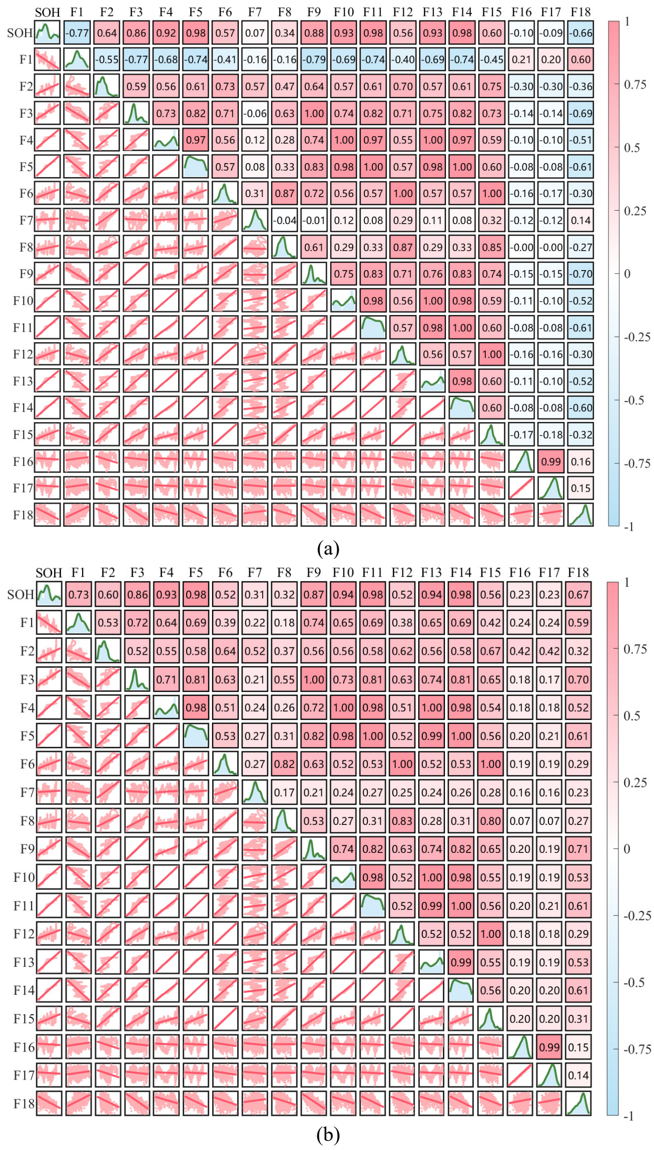


Fig. 7. Visualization of correlations between features and SOH. (a) Pearson correlation heatmap. (b) Distance correlation heatmap. In the lower triangular matrix, scatter plots illustrate the distributions and potential nonlinear relationships between pairs of variables. The upper triangular matrix displays numerically labeled, color-coded correlation coefficients, capturing both linear and nonlinear dependencies. Diagonal elements present kernel density plots that characterize the marginal distributions of each variable.

It is worth noting that the upper triangular regions of the heatmaps reveal strong mutual correlations among many features (indicated by darker colors). For instance, the feature

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2026.3686077

pairs F4 and F5, as well as those from F9 to F15, display correlation coefficients close to 1, reflecting a high degree of linear or nonlinear redundancy, and either can be removed. In addition, the scatter plots provide an intuitive visualization of the overlapping information between these features. These results emphasize the importance of feature selection in reducing redundancy.

3) *Feature Selection*

To address linear and nonlinear redundancy, a feature selection procedure is further designed by integrating Pearson correlation and distance correlation analysis to identify optimal feature subsets. The specific procedure is summarized in Table II, where Pearson correlations greater than 0.8 or distance correlations greater than 0.6 are generally considered to indicate a strong association with SOH. In contrast, correlation coefficients greater than 0.98 between features are highly redundant.

These thresholds were established based on references [20], [30], and our correlation analysis results, which are not strictly theoretical boundaries but serve as practical selection guidelines. Furthermore, to rigorously evaluate the statistical significance of correlations between SOH and features, we conducted t-tests, with results presented in Supplementary Fig. R1. These results further confirm that the selected features exhibit a significant correlation with SOH ($p < 0.001$).

TABLE II
FEATURE SELECTION BASED ON PEARSON AND DISTANCE CORRELATION

Feature selection procedure

Input: Feature matrix $X = [f_1, f_2, ..., f_m]$,
Target variable Y=SOH,
Pearson correlation threshold α_p=0.8,
Distance correlation threshold α_d=0.6,
Redundant threshold α_R=0.98.
Output: Optimal feature subsets F.
Steps:
1: **For** each feature f_i in X:
2: Compute the Pearson correlation coefficient PCorr$_i$ between f_i and Y,
Compute the distance correlation coefficient dCorr$_i$ between f_i and Y,
4: **End For**
5: Select features satisfying $|\text{PCorr}_i| > \alpha_p$ or dCorr$_i > \alpha_d$, and sort them in descending order.
6: Construct the initial feature set F with the selected features.
7: **For** i = 1 to m **do**
8: **For** j = i+1 to m **do**
9: Compute cross-Pearson correlation PCorr$_{ij}$ between f_i and f_j,
Compute cross-distance correlation dCorr$_{ij}$ between f_i and f_j,
10: **If** $|\text{PCorr}_{ij}| > \alpha_R$ or dCorr$_{ij} > \alpha_R$
11: Remove f_j from F.
12: **End If**
13: **End For**
14: **End For**

TABLE III
DIFFERENT FEATURE SUBSETS

Feature Subset	Included Feature	Subset Description
X_1	F1-F18	Full feature pool as baseline
X_2	F1, F2, F5, F15, F16, F18	Selected features with selection mechanisms
X_3	F1, F2, F5, F15, F18	Same as X_2 but without temperature features
X_4	F1-F5, F9-F11, F13-F14, F18	Highly SOH-correlated features
X_5	F1-F18	Same as X_1 but without BEAST decomposed inputs

Following the predefined feature selection process, six key features (F1, F2, F5, F15, F16, and F18) are identified that exhibit a strong correlation with SOH while maintaining low redundancy. Based on these, four representative and mutually distinct feature subsets (X_1–X_5) are constructed for subsequent performance validation of the SOH estimation model. Detailed descriptions of these subsets are provided in Table III.

B. *Physics-Informed Deep Gated Recurrent Unit Architecture*

The general framework of the physics-informed deep gated recurrent unit (PIDGRU) architecture for battery SOH estimation under field conditions is illustrated in Fig. 8. Given N training samples $\mathcal{D} = \{t_i, \mathbf{x}_i, u_i\}_{i=1}^{N} \in \mathbb{R}^{N\times 8}$, where $\mathbf{x} \in \mathbb{R}^{N\times 6}$ denotes the selected optimal feature subsets, $t \in \mathbb{R}^{N\times 1}$ and $u \in \mathbb{R}^{N\times 1}$ represent the corresponding cycle times and target SOH labels, respectively, the estimation process proceeds as follows. First, a surrogate neural network (NN) parameterized by Φ is employed to model the empirical degradation features. These are then combined with a deep hidden physics model (DeepHPM) to infer latent nonlinear degradation dynamics that are not directly observable from the field data. To facilitate this inference, automatic differentiation (AutoDiff) is used to compute the required gradient information. In addition, Bayesian inference is incorporated as a constraint to capture the uncertainty in SOH degradation, thereby enabling robust and reliable estimation under real-world conditions.

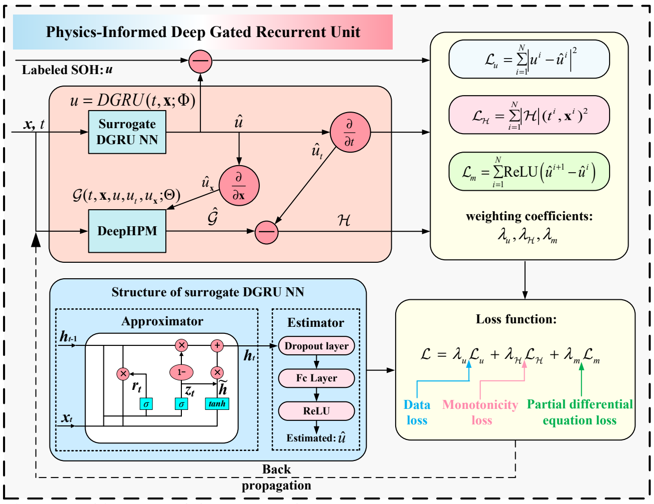


Fig. 8. Architecture of the proposed PIDGRU.

1) *Surrogate NN Model*

As illustrated in Fig. 8, the surrogate NN designed to capture temporal dependencies consists of two primary modules: an approximator and an estimator. Considering field condition data of electric vehicle battery packs are typically noisy, heterogeneous, and limited in effective degradation trajectories, the approximator employs a stacked two-layer GRU architecture referred to as the Deep GRU (DGRU). Compared with LSTM, GRU retains sequence dependency modeling capabilities while employing a more compact gating structure and fewer trainable parameters, which enhances training efficiency and robustness under field conditions [18]. Although attention-based Transformer models demonstrate strong potential in sequence modeling, their advantages are typically more prominent when large-scale datasets are available and long-range labeled interactions can be fully leveraged [14]. Therefore, a two-layer GRU is adopted as a lightweight and effective temporal encoder. The estimator comprises a dropout layer, followed by a fully connected (FC) layer and a ReLU activation function. The overall SOH degradation modeling process can be expressed as follows:

$$h_t = \mathrm{DGRU}(t, \mathbf{x}; \Phi) \tag{9}$$

$$\tilde{h}_t = \mathrm{Dropout}(h_t) \tag{10}$$

$$o_t = w_o * \tilde{h}_t + b_o \tag{11}$$

$$u_t = \mathrm{ReLU}(o_t) \tag{12}$$

Here, the hidden state h_t is computed based on the internal update and reset gates of the DGRU, where t denotes the corresponding time step. A dropout operation is applied to h_t to enhance generalization, yielding the intermediate output $\tilde{h}_t$. This is then passed through an FC layer parameterized by weights w_o and bias b_o to obtain o_t. Finally, the ReLU activation is applied to o_t to generate the final output u_t.

2) Data-Driven Dynamic SOH Degradation Model

Without loss of generality, the dynamic degradation of SOH can be described by its decline rate as:

$$\frac{\partial u}{\partial t} = g(t, \mathbf{x}, u; \Theta) \tag{13}$$

where the function $g(\cdot)$ represents a nonlinear mapping of time t, features $\mathbf{x}$, and model outputs u. In principle, this mapping can be explicitly described by a PDE parameterized by Θ.

However, in real-world battery systems, $g(\cdot)$ typically involves complex, coupled degradation mechanisms across multiple physical scales, such as electrode-level material aging and the growth of the SEI film [10]. Consequently, it is extremely difficult to formulate a unified, high-fidelity PDE that accurately captures these degradation mechanisms while remaining applicable to segmented and noisy field data.

Motivated by this challenge, we define a flexible and generalizable deep hidden physics model (DeepHPM), parameterized by a set Θ, and equipped with AutoDiff to approximate the nonlinear mapping $g(\cdot)$ [32]. The DeepHPM model emphasizes reducing reliance on explicitly parameterized physical models, particularly in field applications where battery internal degradation states are difficult to observe and acquire. The physically consistent electrical response patterns provide a more operationally feasible and applicable approach for integrating domain knowledge into the learning framework. Driven by field data and their partial derivatives, DeepHPM enables the characterization of implicit degradation mechanisms that govern the temporal evolution of SOH. Accordingly, (13) can be reformulated as

$$u_t = \frac{\partial u}{\partial t} \approx \mathcal{G}(t, \mathbf{x}, u, u_t, u_{\mathbf{x}}, u_{\mathbf{xx}}, \cdots; \Theta) \tag{14}$$

where $\mathcal{G}$ represents the DeepHPM network, as constructed by another DGRU, $u_{\mathbf{x}} = \left[\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \cdots, \frac{\partial u}{\partial x_n}\right]^{\mathrm{T}}$ denotes the first-order derivative of u with respect to $\mathbf{x}$, and $u_{\mathbf{xx}}$ represents the second-order derivative, and so on. In this study, a first-order time derivative is considered to reflect the accumulative and irreversible battery aging, while the ReLU activation is adopted in the DeepHPM network to enforce non-negative aging rates. Therefore, the input dimension of DeepHPM follows a structured relationship of twice plus one, which enables it to learn underlying degradation mechanisms beyond the surrogate model.

3) Model Fusion

Based on the previous formulation, the DGRU-based surrogate NN solves the approximation process $u_t \approx \mathrm{DGRU}(t, \mathbf{x}; \Phi)$, thereby forming a parameterized prior for the unknown SOH. In parallel, the DeepHPM implicitly approximates the PDEs $\partial u / \partial t \approx \mathcal{G}(t, \mathbf{x}, u, u_t, u_{\mathbf{x}}, u_{\mathbf{xx}}, \cdots; \Theta)$ that describe the dynamic SOH degradation process. Given the sparsity of field observations and the complexity of the underlying physical mechanisms, this solution process involves a high-dimensional inverse problem aimed at discovering nonlinear degradation dynamics. Based on the above analysis, the mathematical formulation of PIDGRU is defined by (13) and (14), as presented below:

$$\mathcal{H}(t, \mathbf{x}; \Phi, \Theta) := \frac{\partial}{\partial t}\mathrm{DGRU}(t, \mathbf{x}; \Phi) - \mathcal{G}(t, \mathbf{x}, u, u_t, u_{\mathbf{x}}; \Theta) \tag{15}$$

where $\partial \mathrm{DGRU}(t, \mathbf{x}; \Phi) / \partial t$ denotes the partial derivative of the output of the DGRU with respect to time t.

C. Bayesian Inference Uncertainty Constrained (BIUC) Training Strategy

To achieve physics-consistent parameter learning within the PIDGRU framework, we design three loss functions: data fitting loss, PDE loss, and monotonicity loss.

$$\mathcal{L}_u(\Phi) = \sum_{i=1}^{N} \left| u(t_i, \mathbf{x}_i; \Phi) - u_i \right|^2 \tag{16}$$

$$\mathcal{L}_{\mathcal{H}}(\Phi, \Theta) = \sum_{i=1}^{N} \left| \mathcal{H}(t_i, \mathbf{x}_i; \Phi, \Theta) \right|^2 \tag{17}$$

$$\mathcal{L}_m(\Phi) = \sum_{i=1}^{N} \mathrm{ReLU}\left(u(t_{i+1}, \mathbf{x}_{i+1}; \Phi) - u(t_i, \mathbf{x}_i; \Phi) \right) \tag{18}$$

where $\mathcal{L}_u$ is the data loss, provided by the DGRU surrogate network to quantify the deviation between estimated and observed SOH values; $\mathcal{L}_{\mathcal{H}}$ denotes the physics-informed loss, which implicitly approximates the underlying degradation dynamics governing the evolution of SOH via the DeepHPM and partial derivative of the DGRU module; $\mathcal{L}_m$ represents the monotonicity loss, which enforces the physical prior that SOH degradation should not exhibit unphysical recovery. However, in complex real-world scenarios, relying solely on loss functions is insufficient to address the uncertainty inherent in the SOH degradation process.

To address this challenge, we propose a Bayesian inference uncertainty constrained (BIUC) training strategy, which enables the model to minimize loss functions (16)-(19) while quantifying and constraining uncertainty in SOH estimation. The proposed BIUC strategy constrains the training process by augmenting the standard regression loss with physically informed uncertainty regularization terms derived from BEAST analysis. By embedding prior physical information from BEAST (i.e., uncertainty intervals for seasonal degradation components, significant drop points TCP, and their corresponding slopes as shown in Fig 3) into the loss function, the proposed PIDGRU model constrains the estimation within physically meaningful confidence regions. During training, these additional loss terms directly participate in gradient-based optimization and influence parameter updates by penalizing deviations from the identified uncertainty priors.

$$\begin{aligned}\min_{\Phi,\Theta}\ & \mathcal{L}_{All} = \lambda_u \mathcal{L}_u(\Phi) + \lambda_{\mathcal{H}} \mathcal{L}_{\mathcal{H}}(\Phi,\Theta) + \lambda_m \mathcal{L}_m(\Phi) \\ \text{s.t.}\ & u(t_i, \mathbf{x}_i; \Phi) \in [0.9\hat{u}_{BIUC}, 1.1\hat{u}_{BIUC}],\ \hat{u}_{BIUC} = P(\mu_\Omega, \Omega, \sigma^2 |\ u_i) \\ & 0 < \frac{\partial u}{\partial t} < \beta_k \\ & 0 < t < \text{TCP}\end{aligned} \tag{19}$$

where λ_u, $\lambda_{\mathcal{H}}$, and λ_m are the loss weights corresponding to the loss terms $\mathcal{L}_u$, $\mathcal{L}_{\mathcal{H}}$, and $\mathcal{L}_m$, respectively; TCP is the trend change points, β_k is the corresponding slope, and P is the Bayesian posterior result. These weights are used to balance the contributions of different loss terms during the training process. The parameter optimization of PIDGRU under the BIUC strategy is detailed in Algorithm 1.

Algorithm 1: Training the proposed PIDGRU model

Input: Training data $\mathcal{D} = \{t, \mathbf{x}, u\}$, model hyperparameters.
Output: Network $\mathcal{H}(t, \mathbf{x}; \Phi, \Theta)$.
Steps:
1: **For** epoch = 1, 2, $\cdots$ **do**
2: Compute surrogate NN solution value: $\hat{u} \leftarrow u(t_i, \mathbf{x}_i; \Phi)$
3: Compute time derivative: $\hat{u}_t \leftarrow \text{AutoDiff}(t, \hat{u})$
4: Compute first-order derivative: $\hat{u}_{\mathbf{x}} \leftarrow \text{AutoDiff}(t, \mathbf{x})$
5: $\cdots$ (Compute higher-order derivatives as required)
6: Evaluate dynamic model: $\hat{\mathcal{G}} \leftarrow \mathcal{G}(t, \mathbf{x}, u, u_t, u_{\mathbf{x}}, u_{\mathbf{xx}}, \cdots; \Theta)$
7: Compute PIDGRU: $\hat{\mathcal{H}} \leftarrow \hat{u}_t - \hat{\mathcal{G}}$
8: Compute data loss $\hat{\mathcal{L}}_u(\Phi)$, PDE loss $\hat{\mathcal{L}}_{\mathcal{H}}(\Phi,\Theta)$, and monotonicity loss $\hat{\mathcal{L}}_m(\Phi)$ by (16)-(18)
9: Combine total loss: $\hat{\mathcal{L}}_{All}$ by (19) and back propagation
10: Update network parameter Φ and Θ:
$\nabla_\Phi \mathcal{L}_{\text{All}} = \lambda_u \nabla_\Phi \mathcal{L}_u(\Phi) + \lambda_H \nabla_\Phi \mathcal{L}_H(\Phi,\Theta) + \lambda_m \nabla_\Phi \mathcal{L}_m(\Phi)$
$\nabla_\Theta \mathcal{L}_{\text{All}} = \lambda_H \nabla_\Theta \mathcal{L}_H(\Phi,\Theta)$
11: **End For**

IV. Results and Discussion

A. Experiment Setup

To evaluate the effectiveness of the proposed method, all experiments are conducted on a system equipped with an NVIDIA GeForce RTX 4060 GPU. The proposed PIDGRU model is implemented in Python 3.9 using PyTorch 1.11.0 as the deep learning framework. The hyperparameter settings and detailed configurations of the PIDGRU model are summarized in Table IV, with all hyperparameters determined based on domain expertise and systematically tuned through iterative experiments to maximize model performance. During training, a cosine annealing schedule is employed, gradually decaying the learning rate from an initial value of 2×10^{-3} to a final value of 2×10^{-4} to achieve stable convergence.

TABLE IV
Hyperparameter Configuration of the Proposed PIDGRU Model

Description	Configuration
Approximator structure (DGRU)	hidden layer number=2; hidden size=150; activation= ReLU
Estimator structure	hidden size=32; dropout=0.2 activation= ReLU
Dynamic degradation structure (DeepHPM)	hidden layer number=2; hidden size=150;
Loss weight	$\lambda_u = 1; \lambda_{\mathcal{H}} = 0.7; \lambda_m = 0.3$
Epoch	500
Batch size	128
Optimizer	Adam
Initial learning rate (Global)	0.002
End learning rate	0.0002

Three evaluation metrics, including mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and Kling–Gupta efficiency (KGE) [33], are employed to assess the model's performance.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{25}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|^2} \tag{26}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \tag{27}$$

$$\text{KGE} = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{\hat{y}}}{\sigma_y} - 1\right)^2 + \left(\frac{\mu_{\hat{y}}}{\mu_y} - 1\right)^2} \tag{28}$$

where N denotes the total number of data entries considered in the evaluation; y_i, $\hat{y}_i$, and $\bar{y}_i$ represent the SOH label, estimated SOH, and average of the SOH label, respectively; r represents linear correlation; $\sigma_{\hat{y}}$ and σ_y denote the standard deviation of estimated SOH and the standard deviation of SOH label, respectively; $\mu_{\hat{y}}$ and μ_y indicate the estimated SOH mean and the SOH label mean, respectively.

B. Ablation Study

In this section, two ablation studies are conducted to systematically evaluate the impact of different input feature subsets (X_1–X_5, as defined in Table III) and various model components on the performance of the proposed PIDGRU model. To ensure the generalizability of the results, two representative EVs are selected for SOH estimation experiments, corresponding to high-frequency charging vehicle (EV-HF) and low-frequency charging vehicle (EV-LF) datasets. For each EV, early-life data are used to estimate the late-life SOH trajectory by splitting the dataset into a 1:1 ratio.

1) Ablation of Different Feature Subset Inputs

Figs. 9(a) and (b) present the SOH estimation results for the EV-HF and EV-LF datasets, respectively. The estimated SOH degradation trajectories using the optimal feature subsets X_2 show close agreement with the ground truth. It can be seen that the estimated SOH curves remain strictly within the 95% confidence interval (CI) throughout the entire cycle range,

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2026.3686077

with the maximum error limited to within 2%, as confirmed by the error box plots in Figs. 9(c) and (d).

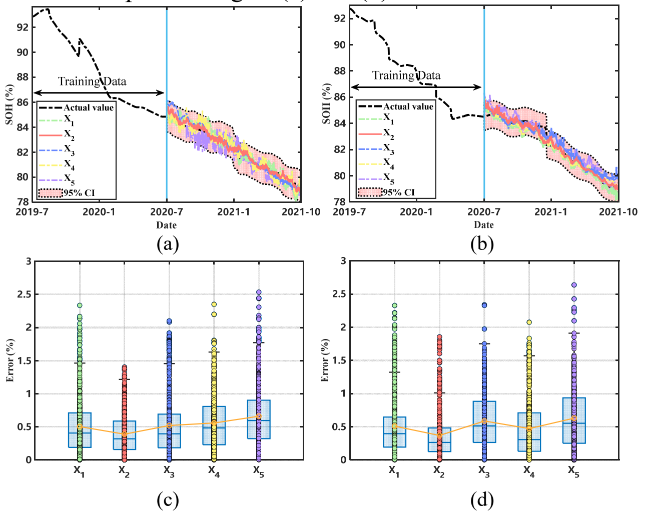


Fig. 9. Results of different feature subsets. (a) EV-HF. (b) EV-LF. (c) The error box plots of EV-HF. (d) The error box plots of EV-LF.

Table V further quantifies the impact of feature selection and temperature-related features through four evaluation metrics. In particular, the optimal feature subset X_2 achieves the lowest MAE and RMSE, along with the highest R^2 and KGE, across both datasets. Specifically, for the EV-HF dataset, X_2 yields a minimum MAE of 0.3912%, an RMSE of 0.4892%, an R^2 of 96.15%, and a KGE of 94.02%, significantly outperforming the baseline full feature pool X_1. A similar advantage is observed in the EV-LF dataset. Crucially, when the temperature-related features are excluded (feature subset X_3), estimation accuracy decreases significantly, with both MAE and RMSE increasing by over 30%, and R^2 dropping below 94%. These results demonstrate the significant contribution of temperature features to the estimation of SOH. In addition, although X_4 consists of features highly correlated with SOH, its performance is worse than that of X_2, which indicates that feature selection strategies that consider both relevance and redundancy are more effective than correlation-based methods alone. Meanwhile, X_5 achieved satisfactory results even without BEAST-derived input features, which confirms that performance improvements are not solely driven by trend extraction.

TABLE V
COMPARISON OF SOH ESTIMATION PERFORMANCE ON DIFFERENT INPUT FEATURE SETS

Dataset	Feature subset	MAE (%)	RMSE (%)	R^2 (%)	KGE (%)
EV-HF	X_1	0.5055	0.6533	93.1276	88.7221
	X_2	**0.3912**	**0.4892**	**96.1469**	**94.0232**
	X_3	0.5213	0.6962	92.1955	81.7256
	X_4	0.5605	0.6939	92.2457	91.5562
	X_5	0.6619	0.8024	89.6315	89.5589
EV-LF	X_1	0.5099	0.6782	94.1074	97.6168
	X_2	**0.3716**	**0.5195**	**96.5428**	**98.1868**
	X_3	0.5876	0.7122	93.5018	90.1433
	X_4	0.4739	0.6416	94.7264	92.2529
	X_5	0.6311	0.7784	92.2376	93.1488

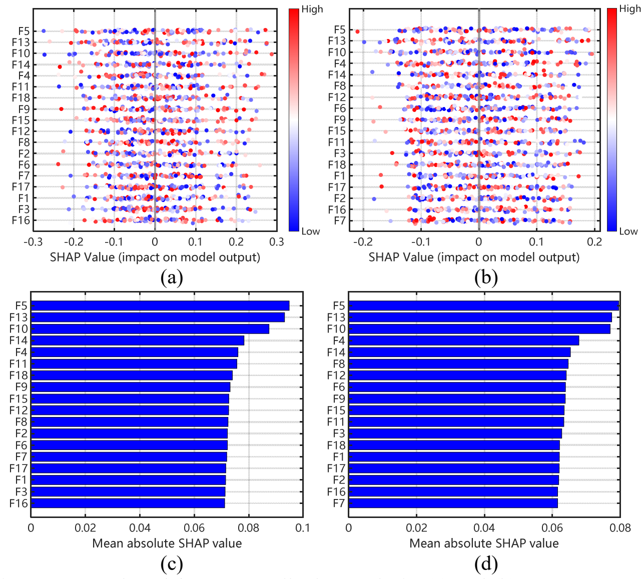


Fig. 10. SHAP-based feature contribution analysis. (a) and (b) are Beeswarm plots for F1-F18 in EV-HF and EV-LF. (c) and (d) are the importance rankings of F1-F18 importance in EV-HF and EV-LF.

Additionally, to enhance the physical interpretability of the proposed model, SHAP analysis is employed to quantify the contribution of each feature in SOH estimation [34]. As shown in Fig. 10, features with high correlations in Fig. 7 (e.g., F4, F5, F10, F11, F13, and F14) all present significant SHAP values, indicating their critical roles in the model's state decision-making. It is observed that F5 (trend component of the entropy of total voltage) is the most influential feature, indicating that irregular charging processes play an important role in battery degradation.

From a physical perspective, an increase in the entropy of total voltage reflects more irregular charging behavior, which is commonly associated with internal resistance growth or lithium plating, thereby accelerating battery degradation. Meanwhile, features F1, F2, F15, and F18 from the optimal subset X_2 exhibit moderate contributions to SOH estimation, which is consistent with the designed feature selection and redundancy analysis procedure.

2) Component Ablation of the PIDGRU Model

Using the optimal feature subset X_2 as input, an ablation study is conducted to evaluate the contributions of key components within the proposed PIDGRU model. The comparison includes variants such as PIDGRU without the BIUC training strategy, a PINN-MLP baseline, and the standalone DGRU model. The SOH estimation results for each configuration are systematically compared in Fig. 11.

As illustrated in Figs. 11(a) and (b), the DGRU and PINN-MLP exhibit notable deviations from the actual values, in particular during periods of rapid SOH degradation and at later cycles, which can also be observed in the error boxplots of Figs. 11(c) and (d), where the maximum error reaches nearly 4%. Additionally, the partial SOH estimation results of PIDGRU without BIUC fall outside the CIs, and its error distribution is visibly more dispersed than that of the complete PIDGRU, further highlighting the importance of the BIUC training strategy.

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2026.3686077

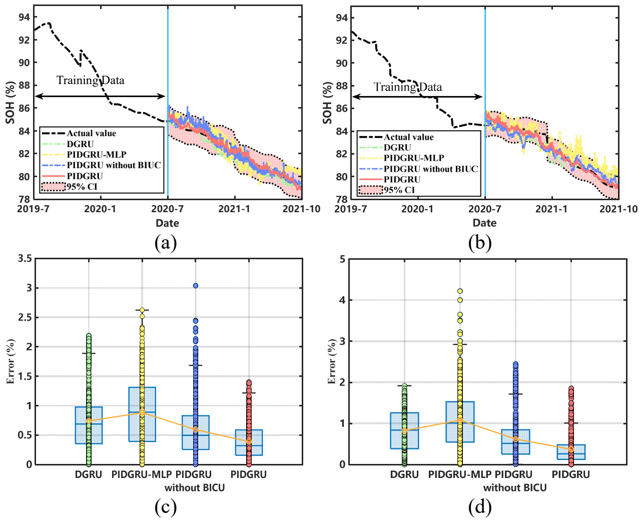


Fig. 11. Results of different components. (a) EV-HF. (b) EV-LF. (c) The error box plots of EV-HF. (d) The error box plots of EV-LF.

TABLE VI
COMPARISON OF PERFORMANCE OF SOH ESTIMATION ON DIFFERENT COMPONENTS

Dataset	Method	MAE (%)	RMSE (%)	R^2 (%)	KGE (%)
EV-HF	DGRU	0.7418	0.9108	86.6406	90.0516
	PINN-MLP	0.8831	1.0443	82.4373	89.4683
	PIDGRU w/o BIUC	0.5949	0.7459	91.0395	84.7947
	PIDGRU	**0.3912**	**0.4892**	**96.1469**	**94.0232**
EV-LF	DGRU	0.8279	0.9581	88.2798	85.4983
	PINN-MLP	1.0828	1.2898	78.6861	80.8887
	PIDGRU w/o BIUC	0.6223	0.7914	91.9749	85.0928
	PIDGRU	**0.3716**	**0.5195**	**96.5428**	**98.1868**

The SOH estimation performance of different components is summarized in Table VI, which is further quantified through MAE, RMSE, R^2, and KGE metrics for ablation studies. The complete PIDGRU consistently achieves R^2 values above 96% on both datasets, demonstrating strong agreement with the ground truth and robust explanatory capability. By effectively integrating DeepHPM and the BIUC training strategy, the model can adapt to various latent SOH degradation patterns, achieving both high accuracy and robustness. When the BICU strategy is removed, as in the PIDGRU without the BICU, the MAE increases by 52.1% and 67.5% in the two datasets, respectively; while the R^2 and KGE drops by approximately 5% and 10%, respectively. Figs. 11(c) and (d) also show error boxplots with higher median errors and wider error distributions. These results indicate that the absence of BIUC constraints significantly weakens the model's ability to learn degradation evolution in a reliable manner, especially for long-term dependencies and abrupt SOH transitions. Compared to PIDGRU, the DGRU model can leverage sequential modeling to capture temporal dependencies more effectively. However, its lack of physical constraints and uncertainty quantification makes it susceptible to error accumulation and instability. In contrast, the PINN-MLP model enforces physical consistency but is inherently limited in sequential modeling due to the feedforward nature of MLPs. As a result, both DGRU and PINN-MLP are outperformed by PIDGRU in terms of accuracy, reliability, and generalization.

C. Comparison With Existing Methods at Different Points

In this section, we further compare the performance of five state-of-the-art methods for SOH estimation, including CNN, SVM, deep kernel extreme learning machine (DKELM), deep LSTM (DLSTM), and the proposed PIDGRU. These models were evaluated on the EV-HF and EV-LF datasets, with comparative experiments conducted under 50% and 25% data availability scenarios, to assess performance with limited data.

To ensure a fair comparison among different models, all baseline models were fine-tuned under the same training and testing splits to avoid data leakage or evaluation bias. The CNN, DLSTM, and PINN-MLP were implemented with a unified two-layer architecture and hyperparameter initialization strategies consistent with prior studies [15]-[18], including batch size = 128, training epochs = 500, convolution kernel size = 3, number of filters = 32, and number of hidden layer units = 150. Regarding learning rate strategies, the cosine annealing learning rate schedule was adopted for all deep learning models, gradually decaying the learning rate from 2×10^{-3} to 2×10^{-4} to ensure a stable and smooth convergence process. For kernel-based SVM and DKELM models, the sparrow search algorithm (SSA) [35] was employed to optimize kernel parameters, ensuring competitive model performance. The corresponding parameter settings and learning rate adjustment schemes are summarized in Table S3 of the Supplementary Materials.

Figs. 12 and 13 illustrate the SOH estimation results for the EV-HF and EV-LF datasets. Regardless of whether 50% or 25% of the input data length is used, the PIDGRU consistently outperforms the other methods, with both MAE and RMSE maintained below 1%. In particular, even with only 25% of EV-HF data, the R^2 remains above 94%, highlighting the proposed PIDGRU model's strong regression capability under limited training data conditions. Furthermore, the 95% CIs of SOH estimations are also shown in Figs. 12 and 13, indicating that PIDGRU's estimated SOH values almost entirely cover the ground truth observed SOH trajectories, which confirms the superior accuracy and robustness of the proposed model.

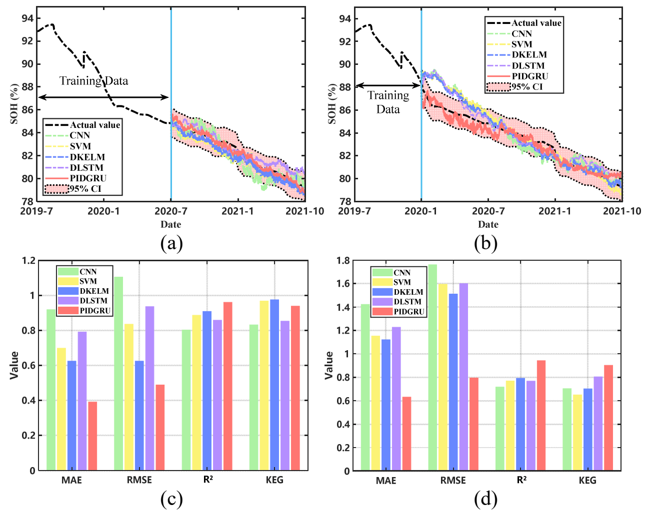


Fig. 12. SOH estimation results of different methods on the EV-HF dataset. (a) Comparison results for the first 50% of the data training. (b) Comparison results for the first 25% of the data training. (c) Performance metrics under 50% training data. (d) Performance metrics under 25% training data.

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2026.3686077

In addition, the error histograms presented in Figs. 12(c)–(d) and 13(c)–(d) offer additional insights into model performance. Among all methods, CNN yields the largest estimation errors, followed by DLSTM. In contrast, both SVM and DKELM demonstrate comparatively better and closely matched performance.

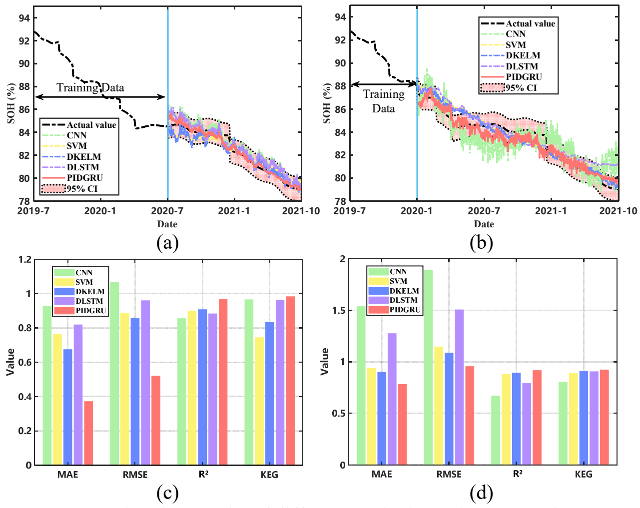


Fig. 13. SOH estimation results of different methods on the EV-LF dataset. (a) Comparison results for the first 50% of the data training. (b) Comparison results for the first 25% of the data training. (c) Performance metrics under 50% training data. (d) Performance metrics under 25% training data.

However, when the training data is reduced from 50% to 25%, the error metrics for CNN, SVM, and DKELM increase sharply, with MAE and RMSE rising by over 60%, and the R^2 values decreasing by more than 10% in most cases. It can be seen that although DLSTM also shows performance degradation, the relative increase in error is moderate and more stable. Nevertheless, the proposed PIDGRU demonstrates the smallest performance drop among all models, highlighting its superior generalization capability and robustness. In addition, we conducted Wilcoxon signed-rank tests for pairwise comparisons of models under the 25% training data condition, with corresponding results summarized in Table S2 of the Supplementary Materials. These paired test results further confirm that the observed performance improvement is statistically significant ($p < 0.05$).

TABLE VII
COMPARISON OF COMPUTATIONAL COST AND INFERENCE EFFICIENCY OF DIFFERENT MODELS

Model	Trainable parameters	Memory (MB)	Inference time (ms)
CNN	2.8137E04	2.75	3.65
SVM	–	0.12	0.35
DKELM	–	0.16	0.18
DLSTM	20.7601E04	0.63	1.57
PINN-MLP	3.9402E04	1.95	0.26
PIDGRU	21.3371E04	3.06	1.20

To evaluate the potential onboard feasibility, we compared the computational cost and inference efficiency of different SOH models in a uniform CPU environment with a batch size of 128. As shown in Table VII, kernel-based SVM and DKELM achieve the lowest latency and memory usage but exhibit limited representation capability. In contrast, CNN and DLSTM suffer from relatively large parameter scales and higher inference costs. The proposed PIDGRU model maintains a favorable balance between accuracy and efficiency, with a low inference time (1.20 ms) and moderate memory usage (3.06 MB). Although this evaluation is conducted on a generically available CPU rather than directly on hardware platforms representing embedded vehicle controllers, it still provides a preliminary alternative assessment of deployment efficiency and suggests the potential suitability of PIDGRU for real-time onboard SOH estimation without GPU acceleration.

D. Validation of Cross-Vehicle SOH Estimation

To evaluate the generalization capability of our proposed SOH estimation method across different vehicles, we conducted a 5-fold cross-validation experiment using data from 20 vehicles. The data were divided into five mutually exclusive subsets, with each subset containing data from four vehicles. In each fold, four subsets were used for training, and the remaining subset was used for testing.

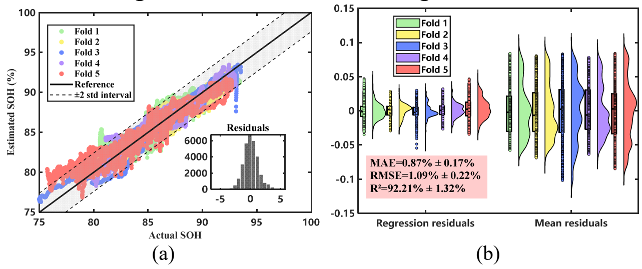


Fig. 14. Cross-vehicle SOH validation results on the 20 EV dataset. (a) SOH regression results for each fold. (b) The violin plots of the residual distributions.

Fig. 14(a) illustrates the regression results for each fold. The regression points align closely with the reference line, with only minor deviations, indicating strong predictive accuracy. Moreover, the residuals are tightly clustered around zero, demonstrating the model's estimation precision. Fig. 14(b) presents violin plots of the residual distributions across the five folds. The MAE, RMSE, and R^2 of the model are 0.87%, 1.09%, and 92.21%, respectively, reflecting an excellent overall performance and strong robustness across all validation sets. Importantly, the residuals for the estimated SOH are consistently smaller and exhibit less fluctuation compared to the residuals of the true labels, which further validates the superiority of our proposed method to reliably perform cross-vehicle SOH estimation.

In addition, to further assess statistical significance, non-parametric bootstrapping was performed across vehicles. Each vehicle's four SOH metrics were first obtained through 5-fold cross-validation, where each vehicle appears exactly once in the test set, ensuring strictly out-of-sample evaluation. Bootstrap resampling (2000 iterations) was then applied to the errors for each vehicle to estimate 95% CI of the four SOH evaluation metrics.

The average performance metrics and corresponding 95% CIs are summarized in Table VIII. As shown, the average MAE and RMSE of the proposed model are 0.8705% and 1.0962%, respectively, with relatively narrow confidence intervals, indicating both high estimation accuracy and stable generalization performance across different vehicles. Similarly, consistently high R^2 and KGE values further confirm the robustness of the proposed method for cross-vehicle SOH estimation under field conditions.

TABLE VIII
CROSS-VEHICLE BOOTSTRAPPED PERFORMANCE WITH 95% CIs

Metric	Mean	95% CI (Bootstrapping)
MAE (%)	0.8705	[0.7548, 1.0001]
RMSE (%)	1.0962	[0.9221,1.2235]
R^2 (%)	92.2177	[89.6909, 93.7157]
KGE (%)	87.3169	[83.6302,90.5929]

E. Sensitivity Analysis for Different Loss Weights

To further explore the sensitivity effects of different monotonicity and PDE loss weights, we conducted a statistical analysis of the capacity regeneration distribution across 20 EVs. As shown in Fig. 15, the probability of capacity regeneration events (defined as a 0–5% increase between consecutive cycles) ranged from 10% to 40% across different vehicles, which indicates that battery aging patterns can be significantly influenced by various factors such as measurement noise, load conditions, and temperature fluctuations. Therefore, it is essential to select an appropriate combination of monotonicity and PDE constraints to ensure the model maintains a reasonable tolerance for capacity regeneration events while consistently capturing long-term degradation trends.

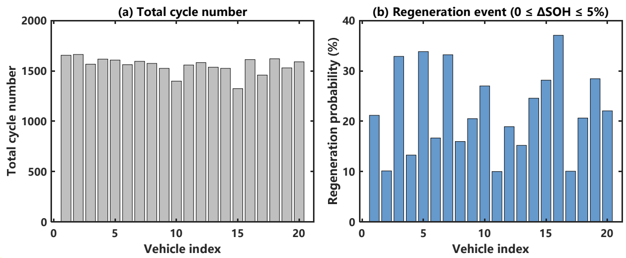


Fig. 15. Probability distribution of SOH regeneration for the 20 EVs dataset.

Based on this observation, we fixed the data fitting loss weight to 1; varied the monotonicity loss weight $\lambda_m \in [0,1]$, while defining the PDE loss weight as $\lambda_{\mathcal{H}} = 1 - \lambda_m$; and evaluated model performance under different weight combinations with the RMSE results of 5-fold cross-vehicle SOH. It can be observed from Fig. 16 that when the monotonicity loss is completely removed ($\lambda_m = 0$), the model exhibits performance degradation and larger cross-fold variance, which indicates that even with PDE constraints, it is insufficient to capture the capacity regeneration phenomenon.

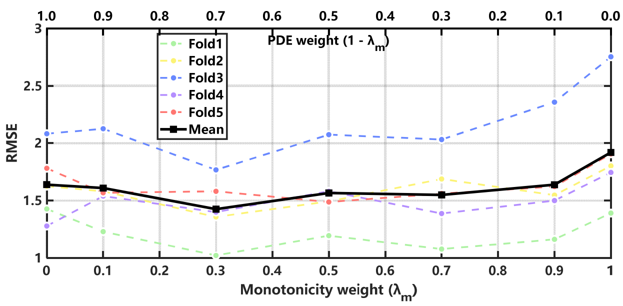


Fig. 16. Sensitivity analysis of the loss weight balance between the monotonicity and PDE.

In contrast, when the monotonicity loss dominates (λ_m=1), the performance deteriorates more significantly, which suggests that overly enforcing monotonicity may conflict with actual battery behavior and reduce the model's ability to adapt to noisy operational data. Importantly, in the intermediate region ($\lambda_m \approx 0.3$), the mean RMSE and inter-fold variance are lowest, achieving a good balance between PDE and monotonicity constraints. Accordingly, the selection of loss weights is determined through quantitative sensitivity analysis based on cross-vehicle RMSE and tolerance for SOH regeneration behavior in field data. This result provides a quantitative basis for balancing different loss terms and highlights the importance of incorporating domain knowledge into the physics-informed training objective.

V. CONCLUSION

In this study, a robust and generalizable battery SOH estimation framework, namely physics-informed deep gated recurrent unit (PIDGRU), is proposed for electric vehicles under real-world operating conditions. By integrating empirical degradation modeling via a deep gated recurrent unit network with nonlinear dynamic learning through DeepHPM, the proposed method effectively captures both observable and latent degradation behaviors. The introduction of the BEAST algorithm enables the modeling of uncertainty in SOH degradation, while a general feature extraction and linear and nonlinear redundant feature selection scheme ensures that an informative and representative input feature set is constructed from segmented charging data. Furthermore, the BIUC strategy enhances the training stability and reliability by addressing the uncertainty of SOH degradation. Extensive experiments conducted under both intra-vehicle and cross-vehicle scenarios validate the effectiveness of the proposed method. These results demonstrate the potential of PIDGRU as a practical and generalizable solution for deployment in real-world battery health management systems.

Although the proposed framework is not explicitly parameterized by chemistry-dependent electrochemical variables, this does not mean that all of its components are chemistry- independent. In particular, the segmented charging data and the BEAST-derived degradation features are extracted from observable electrical responses. The sensitivity and effectiveness of these features may vary across different battery chemistries. For example, the flatter voltage platform and different degradation profiles of LPF batteries may weaken the transferability of learned signal-level prior knowledge. However, with the rapid advancement of large language models (LLMs) as advanced inference and knowledge integration components, it offers promising directions for large-scale and multi-scenario deployment of battery health assessment.

REFERENCES

[1] Z. Wei, K. Liu, X. Liu, et al., "Multi-level data-driven battery management: From internal sensing to big data utilization," *IEEE Trans. Transport. Electrific.*, vol. 9, no. 4, pp. 4805-4823, Dec. 2023.
[2] K. Liu, Y. Liu, S. Zhao, X. Li, and Q. Peng, "An ultrasonic wave-based method for efficient state-of-health estimation of Li-ion batteries," *IEEE Trans. Ind. Electron.*, vol. 72, no. 4, pp. 3792-3802, Apr. 2025.

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2026.3686077

[3] X. Han et al., “A review on the key issues of the lithium ion battery degradation among the whole life cycle,” *eTransp*., vol. 1, Jul. 2019, Art. no. 100005.
[4] Y. Zhang, T. Wik, J. Bergström, and C. Zou, “State of health estimation for lithium-ion batteries under arbitrary usage using data-driven multimodel fusion,” *IEEE Trans. Transport. Electrific*., vol. 10, no. 1, pp. 1494-1507, Mar. 2024.
[5] Z. Wei, K. Liu, X. Liu, et al., “Multi-level data-driven battery management: From internal sensing to big data utilization,” *IEEE Trans. Transport. Electrific*., vol. 9, no. 4, pp. 4805-4823, 2023.
[6] D. Bodnár, G. Ram Chandra Mouli, F. Ďurovský, et al., “Semi-empirical model of nickel manganese cobalt (NMC) lithium-ion batteries including capacity regeneration phenomenon,” *IEEE Trans. Transport. Electrific*., vol. 11, no. 1, pp. 797-813, Feb. 2025.
[7] S. Zhang, Z. Liu, Y. Xu, G. Chen, and H. Su, “An electrochemical aging-informed data-driven approach for health estimation of lithium-ion batteries with parameter inconsistency,” *IEEE Trans. Power Electron*, vol. 40, no. 5, pp. 7354-7369, May 2025
[8] A. I. Rodriguez-Cea, D. Morinigo-Sotelo, and F. V. Tinaut, “A procedure for evaluating the SOH of Li-ion batteries from data during the constant voltage charge phase and the use of an ECM with internal resistance,” *J. Energy Storage*, vol. 108, Art. no. 115074, Feb. 2025.
[9] Kong J, Yang F, Zhang X, et al., “Voltage-temperature health feature extraction to improve prognostics and health management of lithium-ion batteries,” *Energy*, vol. 223, Art. no. 120114, May 2021.
[10] Y. Sun, R. Xiong, X. Meng, X. Deng, H. Li, and F. Sun, “Battery degradation evaluation based on impedance spectra using a limited number of voltage-capacity curves,” *eTransp.*, vol. 22, 2024, Art. no. 100347, Dec. 2024.
[11] S. Dong, X. Zhang, R. Huang, et al., “A simplified fractional-order model adapted to temperature and aging for fast estimation of state of power of lithium-titanate batteries,” *IEEE Trans. Transport. Electrific.*, vol. 11, no. 2, pp. 5484-5496, Apr. 2025.
[12] J. Tang, Y. Li, S. Wang, et al., “Data-driven state of health estimation method of lithium-ion batteries for partial charging curves,” *IEEE Trans. Energy Convers*., vol. 39, no. 4, pp. 2230-2243, Dec. 2024.
[13] Z. Wei, H. Ruan, Y. Li, et al., “Multistage state of health estimation of lithium-ion battery with high tolerance to heavily-partial charging,” *IEEE Trans. Power Electron*., vol. 37, no. 6, pp. 7432-7442, Jun. 2022.
[14] H. Xiao, L. Fu, C. Shang, et al., “A lithium-ion battery state-of-health prediction model combining convolutional neural network and masked multi-head attention mechanism,” *IEEE Trans. Energy Convers.*, vol. 40, no. 1, pp. 218-232, Mar. 2025.
[15] X. Hu, J. Jiang, D. Cao, et al., “Battery health prognosis for electric vehicles using sample entropy and sparse Bayesian predictive modeling,” *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2645-2656, Apr. 2016.
[16] C. Lin, J. Xu, J. Hou, et al., “Ensemble method with heterogeneous models for battery state-of-health estimation,” *IEEE Trans. Ind. Informat*., vol. 19, no. 10, pp. 10160-10169, 2023.
[17] L. Luo, Y. Wang, Y. Li, and Q. Cui, “Degradation capacity spatial-temporal embedding RUL prediction framework for lithium-ion batteries,” *IEEE J. Emerg. Sel. Topics Ind. Electron.*, 2025.
[18] X. Xiong, Y. Wang, C. Jiang, et al., “End-to-end deep learning powered battery state of health estimation considering multi-neighboring incomplete charging data,” *Energy*, vol. 292, Art. no. 130495, 2024.
[19] Y. Fan, F. Xiao, C. Li, et al., “A novel deep learning framework for state of health estimation of lithium-ion battery,” *J. Energy Storage*, vol. 32, Art. no. 101741, 2020.
[20] Y. Liu, H. Chen, L. Yao, et al., “A physics-guided approach for accurate battery SOH estimation using RCMHCRE and BatteryPINN,” *Adv. Eng. Inform*, vol. 65, 2025, Art. no. 103211.
[21] M. Aykol, C. B. Gopal, A. Anapolsky, et al., “Perspective—combining physics and machine learning to predict battery lifetime,” *J. Electrochem. Soc.*, vol. 168, no. 3, Art. no. 030525, 2021.
[22] H. Meng and Y.-F. Li, ‘‘A review on prognostics and health management (PHM) methods of lithium-ion batteries,’’ *Renew. Sustain. Energy Rev*., vol. 116, Dec. 2019, Art. no. 109405.
[23] W. Li, J. Zhang, F. Ringbeck, et al., “Physics-informed neural networks for electrode-level state estimation in lithium-ion batteries,” *J.Power Sources*, vol. 506, Art. no. 230034, 2021.
[24] T. Hofmann, J. Hamar, M. Rogge, et al., “Physics-informed neural networks for state of health estimation in lithium-ion batteries,” *J. Electrochem. Soc.*, vol. 170, no. 9, Art. no. 090524, 2023.
[25] F. Wang, Z. Zhai, Z. Zhao, et al., “Physics-informed neural network for lithium-ion battery degradation stable modeling and prognosis,” *Nat. Commun.*, vol. 15, no. 1, Art. no. 4332, 2024.
[26] P. Wen, Z.-S. Ye, Y. Li, et al., “Physics-informed neural networks for prognostics and health management of lithium-ion batteries,” *IEEE Trans. Intell. Veh.*, vol. 9, no. 1, pp. 2276–2289, Jan. 2024.
[27] L. Yang, M. He, Y. Ren, et al., “Physics-informed neural network for co-estimation of state of health, remaining useful life, and short-term degradation path in lithium-ion batteries,” *Appl. Energy*, vol. 398, Art. no. 126427, 2025.
[28] L. Yang, M. He, Y. Ren, et al., “Electrochemical and thermal analysis of square lithium-ion battery based on a multidimensional electrochemical-thermal coupled model,” *J. Energy Storage*, vol. 110, Art. no. 115257, 2025.
[29] Y. Zhang, X. Feng, M. Zhao, et al., “In-situ battery life prognostics amid mixed operation conditions using physics-driven machine learning,” *J. Power Sources*, vol. 577, Art. no. 233246, 2023.
[30] Z. Deng, L. Xu, H. Liu, et al., “Prognostics of battery capacity based on charging data and data-driven methods for on-road vehicles,” *Appl. Energy*, vol. 339, Art. no. 120954, 2023.
[31] K. Zhao, M. A. Wulder, T. Hu, et al., “Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm,” *Remote Sens. Environ.*, vol. 232, Art. no. 111181, 2019.
[32] M. Raissi, “Deep hidden physics models: deep learning of nonlinear partial differential equations,” *J. Mach. Learn. Res*., vol. 19, no. 1, pp. 932–955, 2018.
[33] W. J. M. Knoben, J. E. Freer, and R. A. Woods, “Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores,” *Hydrol. Earth Syst. Sci*., vol. 23, no. 10, pp. 4323–4331, 2019.
[34] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” *Proc. Adv. Neural Inf. Process. Syst*., vol. 30, 2017, pp. 1–10.
[35] J. Xue and B. Shen, ‘‘A novel swarm intelligence optimization approach: Sparrow search algorithm,’’ *Syst. Sci. Control Eng*., vol. 8, no. 1, pp. 22–34, Jan. 2020.

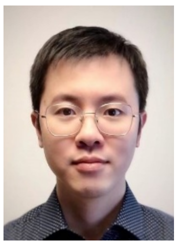
Laijin Luo received the M.S. degree in information and communication Engineering from Anqing Normal University, Anhui, China, in 2024.

He is currently pursuing the Ph.D. degree with the School of Electrical Engineering, Chongqing University, Chongqing, China. His research interests include battery modeling, state estimation and deep learning.

Yu Wang (Senior Member, IEEE) received B.Eng. degree in Electrical Engineering and Automation from Wuhan University, China in 2011, M.Sc. and Ph.D. degree in Power Engineering from Nanyang Technological University (NTU), Singapore in 2012 and 2017, respectively.

Currently, he is a professor at the School of Electrical Engineering, Chongqing University, China. He was a Marie Skłodowska-Curie Individual Fellow at Control and Power Group, Imperial College London.

Dr. Wang serves as the Associated Editor of *IEEE Transactions on Smart Grid, Journal of Modern Power Systems and Clean Energy, IET Generation, Transmission and Distribution*. His research interests include microgrid operation and control, power system dynamics and control, and cyber-physical power systems.



Yang Li (Senior Member, IEEE) received the B.E. degree in electrical engineering from Wuhan University, Wuhan, China, in 2007, and the M.Sc. and Ph.D. degrees in power engineering from Nanyang Technological University (NTU), Singapore, in 2008 and 2015, respectively.

He was with the Energy Research Institute, NTU, the School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, QLD, Australia, and the School of Automation, Wuhan University of Technology, Wuhan, the Department of

Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden, from 2016 to 2025. Since 2025, he has been a Professor with the School of Electrical Engineering and Automation, Wuhan University. His research interests include modeling, control, and management of energy storage systems in the power and transport sectors.

Dr. Li serves as an Associate Editor for several journals, such as IEEE Transactions on Industrial Electronics, IEEE Transactions on Transportation Electrification, and IEEE Transactions on Energy Conversion.

Qiushi Cui (Member, IEEE) earned his M.Sc. degree from Illinois Institute of Technology, and the Ph.D. degree from McGill University, both in Electric Engineering.

Currently, he is an associate professor from School of Electrical Engineering at Chongqing University. He was working as a postdoctoral researcher at of Arizona State University (ASU) and served as the Associate Director of Machine Learning Laboratory for Power Systems in the Ira A. Fulton Schools of Engineering of the same university. Prior to joining ASU, he was a Research Engineer at OPAL-RT Technologies Inc. from Nov. 2015 to Nov. 2017. His research interests are in the areas of machine learning and big data applications in power systems, power system protection, smart cities, microgrid, EV integration, and realtime simulation in power engineering. Dr. Cui has won three Best Paper Awards from UK, China, and the USA, all ranking the first. He was the winner of the Chunhui Cup Innovation and Entrepreneurship Competition for Overseas Chinese Scholars in the Energy Sector in 2018. Dr. Cui received the Postdoctoral Research Scholarship from both Natural Sciences and Engineering Research Council of Canada (NSERC) and Quebec Research Fund-Nature and Technology (FRQNT) and held the MITACS Accelerate Research Program Fellowship from Canada.