ENGINEERING Energy 2026, 20(2): 10464
https://doi.org/10.1007/s11708-026-1046-4

RESEARCH ARTICLE

Parallel deep learning with attention-gated fusion for robust battery health monitoring under dynamic operating conditions

Xiaowen Sun[1], Yunfeng Jiang[4], Binhui Liu (✉)[3,5], Changying Liu[1], Changru Rong (✉)[4], Haiyan Lu (✉)[2]

1 College of Instrumentation and Electrical Engineering, Jilin University, Changchun 130012, China
2 College of Chemistry, Jilin University, Changchun 130012, China
3 China Electronic Product Reliability and Environmental Testing Research Institute, Guangzhou 510000, China
4 China FAW Group Corporation Limited Research and Development General Institute Battery System Development Department, Changchun 130011, China
5 Key Laboratory of MIIT for Intelligent Products Testing and Reliability, Guangzhou 510000, China

HIGHLIGHTS

- Parallel TCN-Transformer model captures multi-scale battery degradation patterns for robust SOH estimation.
- Four health features (K, b, $\sigma_{\Delta Q}$, $\sigma_{\delta\Delta Q}$) show > 0.95 correlation with actual SOH.
- Attention-gated fusion dynamically weights features, enhancing robustness and interpretability.
- Cross-dataset validation achieves RMSE < 1%, outperforming existing benchmarks significantly.

Keywords:
Lithium-ion battery
State-of-health (SOH)
Temporal convolutional network (TCN)
Transformer
Feature fusion
Dynamic operating conditions

ABSTRACT

With the rapid development of electric vehicles and energy storage systems (ESSs), accurate state-of-health (SOH) estimation for lithium-ion batteries has become crucial for ensuring safety and optimizing performance. However, SOH estimation under dynamic operating conditions remains challenging, as non-monotonic voltage profiles and irregular current patterns reduce the effectiveness of conventional measurement methods. This paper proposes a comprehensive approach that combines health feature extraction with a parallel deep learning architecture for robust SOH estimation. First, the method extracts four highly correlated health features (K, b, $\sigma_{\Delta Q}$, and $\sigma_{\delta\Delta Q}$) from dynamic measurement data collected by sensors, with correlation coefficients between these features and the actual SOH exceeding 0.95. These extracted features are then processed through a novel parallel Temporal Convolutional Networks (TCN)-Transformer hybrid architecture: the TCN captures multi-scale local temporal patterns, while the Transformer models global dependencies. An attention-gated fusion module dynamically integrates complementary feature representations from the two branches and adaptively weights different paths based on degradation features. Experimental validation on three standardized battery datasets (MIT, CALCE, Oxford) shows that the method achieves an estimation accuracy with a root mean square error (RMSE) below 1% under all operating conditions, representing an 8%–70% improvement over conventional methods. Attention weight analysis reveals correlations with aging mechanisms, providing interpretability for model decisions. The proposed method enables practical real-time battery health assessment in dynamic environments.

1 Introduction

With the acceleration of the global green energy transition, lithium-ion batteries have emerged as key carriers for clean energy storage and conversion due to their high energy density, high output power, long service life, and low self-discharge rate, playing a crucial role in electric transportation, portable electronic devices, and distributed energy storage systems (ESSs) [1]. However, as electrochemical ESSs, lithium-ion batteries inevitably undergo aging during service, manifesting as capacity loss, reduced power supply capability, and ultimately

✉ Corresponding authors. E-mails: liubinhui@ceprei.biz (B. Liu); rongchangru@139.com (C. Rong); luhy@jlu.edu.cn (H. Lu)

Article history: Received 9 September 2025, Revised 3 November 2025, Accepted 7 November 2025, Available online 1 January 2026

performance degradation and potential safety risks [2].

State-of-health (SOH) estimation represents one of the most challenging tasks in modern battery management systems (BMSs), requiring precise quantification of battery capacity degradation through sophisticated analysis of measured voltage, current, and temperature data [3]. The key challenge lies in extracting reliable health features from measured signals and developing robust estimation algorithms that can handle varying operational conditions [4]. This estimation problem has attracted significant attention due to its critical importance for electric vehicle safety and ESS reliability [5].

Traditional battery health estimation methods rely on controlled laboratory conditions with constant current discharge profiles, where voltage and capacity vary monotonically, enabling the use of conventional feature extraction techniques such as incremental capacity analysis (ICA) [6,7]. Specifically, ICA and differential voltage analysis (DVA) necessitate monotonically varying voltage profiles, typically achievable only under constant-current laboratory measurement conditions [8]. Due to the complexity and noise of actual measurement signals, these methods are ineffective in dynamic operating conditions [9,10].

Electrochemical impedance spectroscopy (EIS), as a crucial non-destructive analytical technique, provides key insights into battery health through systematic analysis of impedance spectra [11–13]. Nevertheless, its reliance on specialized measurement instruments, strict environmental control, and extended measurement durations severely limits its real-time application in dynamic operating scenarios. Equivalent circuit models (ECMs) and electrochemical models attempt to estimate SOH based on measured electrical responses; however, they face parameter uncertainty and high computational complexity under dynamic measurement conditions [14,15].

Data-driven methods have revolutionized SOH estimation [16,17]. By directly extracting health information from measured data and bypassing the need for detailed physicochemical models, these techniques have gained prominence [18–20]. Machine learning algorithms, including support vector regression [21], neural networks [22], and deep learning architectures [23], have been extensively explored. For example, Lin et al. [24] developed multimodal feature fusion techniques, while He et al. [25] improved Gaussian process regression, both enhancing feature-based SOH estimation.

In data-driven methods, health feature extraction is critical. Researchers have developed advanced frameworks for health feature extraction from battery measurements [26,27]. Cui et al. [28] employed long short-term memory (LSTM) networks to handle capacity regeneration features, while Xue et al. [29] proposed hybrid methods combining particle filters and LSTM learning, both highlighting the importance of advanced feature extraction in dynamic environments.

Recent studies have focused on adapting these techniques to dynamic scenarios. Feng et al. [30] introduced multi-feature strategies for high-capacity battery health prediction, Xia et al. [31] developed online filtering for anti-interference feature extraction from noisy data, and Marri et al. [32] compared machine learning strategies using extracted features. However, existing methods struggle in noisy, dynamic conditions, invalidating traditional assumptions and highlighting the need for new methodologies.

Deep learning has gained traction in battery health feature processing due to its superior pattern recognition capabilities [33]. Temporal convolutional networks (TCNs) capture multi-scale temporal patterns through dilated convolutions, while Transformer architectures model long-term dependencies via self-attention mechanisms [34,35]. Researchers have developed diverse deep learning approaches. Zhao et al. [36] integrated attention and transfer learning for battery lifespan prediction, and Akbar et al. [37] proposed robust SOH prediction models for electric vehicle batteries.

Current deep learning research on SOH estimation for lithium-ion batteries has evolved along multiple technical paths, focusing on feature extraction accuracy, temporal modeling capability, and multi-task adaptability. For instance, in terms of the application of attention mechanisms, He et al. [38] proposed a multi-scale convolutional attention mechanism (MCA), extracting multi-scale information from charging data through convolutional kernels of different sizes and dynamically enhancing features strongly correlated with SOH degradation by combining channel and spatial attention modules, effectively solving the feature extraction challenge caused by the diversity of fast charging strategies. Regarding the multi-task learning framework, Zhang et al. [39] proposed a convolutional neural network-multi-gate mixture of gated recurrent units (CMMOG) model. This model uses a convolutional neural network (CNN) to extract high-dimensional health features, adopts GRU as the expert network for task mapping, combines a multi-gate network to allocate weights, and introduces a homoscedastic uncertainty loss function. These designs alleviate the problems of parameter redundancy and improving generalization of traditional single-task models. In terms of model-data fusion, Chen et al. [40] proposed a method combining a fractional-order ECM and an improved Vision Transformer (VIT), identifying model parameters through an optimization algorithm and selecting highly correlated parameters as health features to exploit complementary physical and data-driven

advantages. In terms of Transformer architecture innovation, Zhao et al. [41] proposed a multi-branch Vision Transformer (MB-ViT), designing independent branches for multi-source data such as voltage, current, and temperature to extract features and fuses global information, significantly improving early lifespan prediction accuracy. In terms of LSTM improvement, Bao et al. [42] developed a variant LSTM (VLSTM) with "peephole connections", coupled the forget gate and input gate to optimize information screening, and constructed a model by combining CNN and dimension attention mechanism, demonstrating excellent cross-dataset generalization.

However, single-network methods have inherent limitations: CNNs excel at local pattern recognition but struggle with long-range dependencies, while Transformers tend to overlook local details when modeling global patterns [43,44]. This limitation has driven the development of hybrid architectures that integrate complementary capabilities of different networks. Consequently, current deep learning methods for SOH estimation are moving toward "local-global feature collaborative modeling" and "adaptive multi-source information fusion," continuously improving estimation accuracy and robustness under complex operating conditions through architectural innovations and mechanism optimizations.

Real-world estimation presents three fundamental challenges:

Non-monotonic voltage and current variations: Dynamic operating conditions produce irregular voltage profiles due to current fluctuations, regenerative braking, and polarization effects, rendering conventional techniques unreliable.

Feature extraction from noisy signals: Dynamic current variations introduce measurement noise, degrading capacity calculation and health feature extraction accuracy.

Real-time feature extraction: Practical BMSs require continuous battery health assessment from limited measurement data without relying on complete charge-discharge cycles.

This paper addresses these challenges by proposing a comprehensive methodology, centered on a parallel TCN-Transformer hybrid architecture for integrating and analyzing health features from dynamic measurements. This TCN module efficiently extracts local changes of health features using causal and dilated convolutions, accurately capturing short-term battery state variations.

The Transformer module models long-term evolution of capacity degradation through multi-head attention, capturing overall SOH trends. An attention-gated fusion module dynamically integrates the outputs from both modules, enhancing global temporal correlations while preserving local details. Compared with existing single-architecture methods, this parallel structure reduces computational redundancy through feature reuse while ensuring high estimation accuracy, providing a novel approach for real-time SOH estimation under complex discharging conditions.

The proposed methodology includes four key technical contributions:

Dynamic condition feature extraction framework: A framework integrating established signal processing techniques (dual polarization equivalent circuit modeling and voltage segmentation) to derive four highly correlated health features (linear relationship parameters K, b and statistical features $\sigma_{\Delta Q}$, $\sigma_{\delta\Delta Q}$) from dynamic measurement signals, achieving correlation coefficients exceeding 0.95 with actual SOH across different operating scenarios.

Parallel TCN-Transformer hybrid architecture: A novel parallel deep learning framework synergistically combining TCNs for multi-scale local temporal pattern recognition with Transformer mechanisms for global dependency modeling, enabling collaborative processing of complementary temporal patterns within extracted health features while maintaining computational efficiency for real-time applications.

Attention-gated feature fusion mechanism: An adaptive fusion module that dynamically weights feature representations from both TCN and Transformer branches according to battery aging stages, significantly enhancing model robustness against measurement variations and providing interpretability for model decisions.

Comprehensive multi-dataset validation: Extensive testing on three representative measurement scenarios (constant-current, periodic pulse, complex dynamic) using MIT, CALCE, Oxford datasets, demonstrating superior estimation accuracy and practical applicability, including cross-battery-type transfer learning analysis and attention weight correlation studies that provide insights into the relationship between model decisions and battery aging mechanisms.

2 Overall methodology framework

Figure 1 presents the complete methodology framework, illustrating the systematic approach from dynamic measurement challenges to robust SOH estimation. The framework consists of three main stages: dynamic feature extraction that transforms irregular discharge data into reliable health indicators through OCV estimation and standardized voltage segmentation, parallel deep learning processing where the TCN captures multi-scale temporal patterns while the Transformer module models global dependencies, and attention-gated fusion that dynamic-

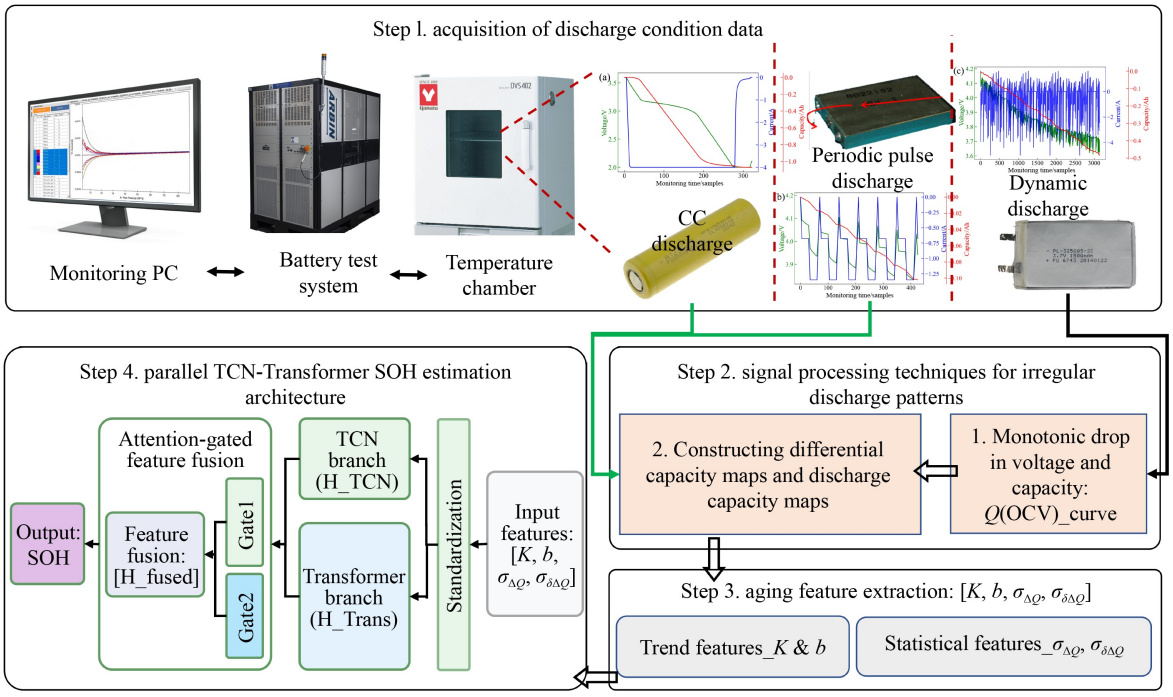

Fig. 1 Overall methodology framework for robust SOH estimation under dynamic operating conditions. (The framework integrates dynamic feature extraction with parallel deep learning architecture to address the challenges of non-monotonic voltage profiles and measurement noise in practical BMSs.)

ally integrates complementary information for accurate SOH prediction.

3 Data set description and dynamic feature extraction

3.1 Data set description

Three representative datasets covering different measurement scenarios are utilized: the MIT Dataset (controlled CC discharge, 124 LFP cells) [45], the CALCE Dataset (periodic pulse discharge, CX2-3 prismatic LCO cell) [10], and the Oxford Dataset (complex dynamic profiles, 6 LCO pouch cells) [46]. Detailed specifications and experimental configurations for each dataset are provided in Electronic Supplementary Material (SEM) S1.

3.2 Health feature extraction for dynamic conditions

3.2.1 Challenges in dynamic feature extraction

Feature extraction from dynamic battery measurements faces three critical challenges: non-monotonic voltage behavior that violates fundamental ICA/DVA assumptions, measurement noise that degrades signal-to-noise ratio of capacity calculations, and ensuring feature consistency across cycles with entirely different current profiles.

3.2.2 Open circuit voltage estimation for stable analysis

Under dynamic discharge conditions, significant terminal voltage fluctuations cause conventional $Q(V)$ curves to lose their smooth monotonicity. To address this, the present study constructs Q(OCV) curves using the open-circuit voltage, which is mainly determined by the state of charge and electrode capacity loss, and is strongly correlated with battery capacity degradation process.

To achieve this, the dual-polarization (DP) model within the ECM is utilized, as shown in Fig. 2(a). The DP model is widely recognized for accurately describing the external characteristics of a battery and is extensively applied in SOC estimation. The dual-polarization ECM provides accurate OCV estimation [47]:

$$U_{\mathrm{OC}} = U - U_1 - U_2 - IR_0, \tag{1}$$

where U is terminal voltage, U_1 and U_2 are polarization voltages, I is the discharge current, and R_0 is the ohmic internal resistance. Parameters are identified using forgetting factor recursive least squares (FFRLS) for real-time parameter identification.

Figure 2(b) illustrates a comparison between the

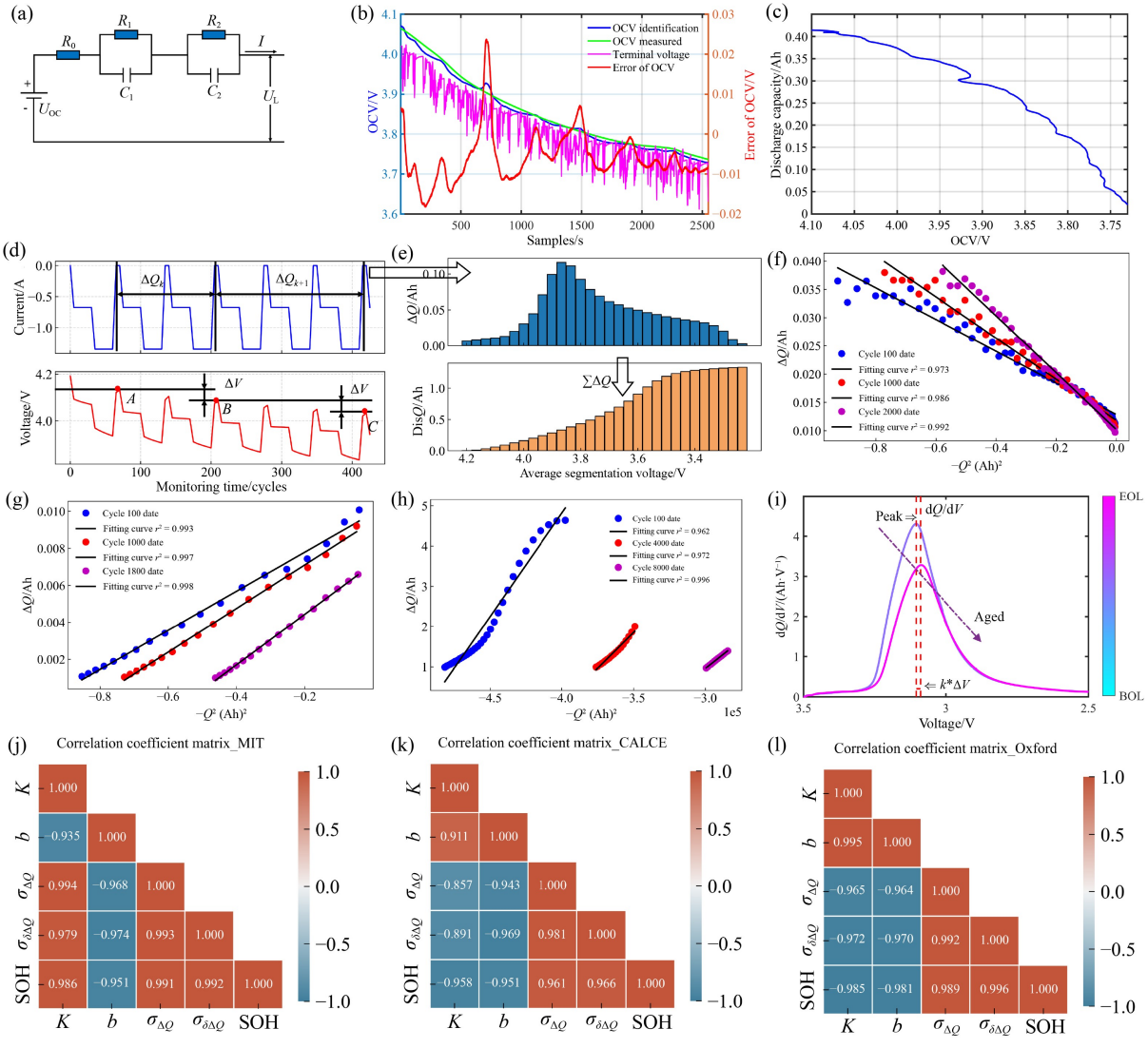

Fig. 2 Comprehensive dynamic feature extraction method.

(a) DP ECM of lithium-ion battery; (b) OCV identification results with error distribution; (c) monotonically decreasing Q(OCV) curve; Standardized voltage segmentation strategy for consistent feature extraction: (d) differential capacity corresponding to the average segmented voltage point and average segmented voltage point selection; (e) sequential distribution of differential capacity and sequential distribution of discharge capacity; (f) linear model fitting results for CX2-3; (g) linear model fitting results for MIT; (h) linear model fitting results for Oxford; (i) variation of dQ/dV for different cycling processes. Correlation coefficient matrices between four features and SOH for different datasets: (j) MIT dataset; (k) CX2-3 dataset; (l) Oxford dataset.

identified U_{OC} obtained using FFRLS and the measured OCV under dynamic operating conditions. The results validate the accuracy and robustness of the FFRLS method in estimating U_{OC}. Subsequently, the identified U_{OC} is used to construct the Q(OCV) curve, shown in Fig. 2(c). Unlike the fluctuating $Q(V)$ curve under dynamic conditions, the Q(OCV) curve is monotonic, providing a smooth and reliable basis for feature extraction. This demonstrates the effectiveness of the proposed voltage-smoothing strategy in addressing challenges posed by dynamic operating conditions.

3.2.3 Standardized voltage segmentation for consistent feature extraction

Figures 2(d) and 2(e) illustrates a periodic pulse discharge and standstill process. After each pulse, the voltage experiences a short rise during the standstill due to relaxation. However, this voltage rise remains below the voltage of the previous stage (e.g., points A, B, and C), maintaining a principle of monotonic decline. The repetition period of discharge and rest stages is denoted as N_m, but voltage drops across adjacent intervals are

unequal. For example, assuming the voltage drop in equal intervals is ΔV, segment AB spans two periods, whereas BC spans three. To ensure consistency and comparability of the characteristics between different discharge cycles, a standardized voltage interval approach is applied which uniformly divides the analysis voltage range:

$$\Delta V = \frac{V_{\text{high}} - V_{\text{low}}}{N_{\text{p}}}, N_{\text{p}} < N_{\text{m}}, \tag{2}$$

where V_{high} and V_{low} are the start and end discharge voltages, respectively, N_{p} represents the number of voltage intervals. For each voltage interval $[V_i, V_{i+1}]$, the corresponding discharge capacity is calculated by ampere-time integration:

$$\Delta Q(V_i) = \int_{t(V_i)}^{t(V_{i+1})} I(t)\mathrm{d}t, \tag{3}$$

where $t(V_i)$ indicates the time when the voltage reaches V_i, and $I(t)$ is the discharge current.

Based on these calculations, two important distribution curves are constructed: the differential capacity distribution $\Delta Q(V)$, reflecting capacity change rate over equal voltage intervals, and the cumulative discharge capacity $Q(V)$, indicating total capacity from discharge start:

$$Q(V_i) = \sum_{j+1}^{i} Q(V_j). \tag{4}$$

This segmentation approach provides robust performance under irregular discharge patterns by maintaining fixed voltage references regardless of current profile variations.

3.2.4 Multi-scale health feature extraction

From the processed voltage and capacity data, four health features highly correlated with battery SOH are extracted:

Trend feature extraction: A linear relationship exists between differential capacity distribution $\Delta Q(V)$ and the square of discharge capacity distribution $Q(V)^2$ [48]:

$$\Delta Q(V_i) = -K[(Q(V_i)]^2 + b + \varepsilon, \tag{5}$$

where K and b are slope and intercept parameters, and ε is Gaussian white noise. The voltage interval near the $\Delta Q(V)$ peak to low-voltage region is selected for fitting, characterized by relatively slow capacity changes and high data resolution. The model was applied to the CX2-3 dataset at the 100th, 1000th, and 2000th cycles. Parameter values (K and b) were computed for each cycle, and the corresponding fitting results are illustrated in Fig. 2(f), demonstrating that as the battery ages, the slope K and intercept b change systematically, reflecting the degradation of the battery.

Statistical feature extraction: To capture subtle variations in capacity distribution, the standard deviation of the differential capacity for a single cycle ($\sigma_{\Delta Q}$) is calculated:

$$\sigma_{\Delta Q} = \sqrt{\frac{1}{N_{\text{p}}}\sum_{i=1}^{N_{\text{p}}}(\Delta Q(V_i) - \overline{\Delta Q})^2}. \tag{6}$$

As shown in Fig. 2(i), considering that the $\Delta Q(V)$ peak shifts toward low-voltage regions during battery aging, a hysteresis parameter k is introduced to align peak positions across different cycles:

$$k = \left\lceil (\Delta Q^{\text{cur}}(V)_{\text{peak}} - \Delta Q^{\text{ref}}(V)_{\text{peak}}/\Delta V \right\rceil, \tag{7}$$

where $\lceil \cdots \rceil$ denotes upward rounding. $\Delta Q^{\text{cur}}(V)_{\text{peak}}$ and $\Delta Q^{\text{ref}}(V)_{\text{peak}}$ denote the differential capacities of the current cycle and the reference cycle, respectively.

The interleaved differential capacity sequence accounts for peak shifts:

$$\delta\Delta Q(V_i) = \Delta Q^{\text{cur}}(V_{i+k}) - \Delta Q^{\text{ref}}(V_i). \tag{8}$$

The standard deviation of this staggered differential capacity sequence ($\sigma_{\delta\Delta Q}$) is calculated as the second statistical feature:

$$\sigma_{\delta\Delta Q} = \sqrt{\frac{1}{N-k}\sum_{i=1}^{N-k}(\delta\Delta Q(V_i) - \overline{\delta\Delta Q})^2}. \tag{9}$$

3.3 Health feature correlation analysis

Quantitative correlation assessment: Pearson correlation analysis demonstrates strong relationships between the extracted features and actual SOH measurements. The segmentation parameter N_{p} = 30 was determined based on the actual operational characteristics of the CX2-3 dataset, where periodic pulse discharge and rest conditions produce 45–80 effective cycles across the battery's lifespan from fresh to aged states. This parameter selection ensures adequate capture of voltage-capacity relationship variations across different aging stages while avoiding information loss or excessive noise introduction. Excessive segmentation ($N_{\text{p}} >> 30$) introduces unnecessary measurement noise from current fluctuations, while insufficient segmentation ($N_{\text{p}} << 30$) risks losing critical electrochemical signatures, such as $\mathrm{d}Q/\mathrm{d}V$ peak shifts indicative of capacity fade mechanisms. A comprehensive parameter sensitivity analysis (ESM S2) using the MIT dataset as a representative example, validates that N_{p} = 30 achieves optimal feature-SOH correlation (> 0.95).

To ensure methodological consistency and enable fair performance comparisons, the same N_{p} = 30 configuration was uniformly applied across all three datasets (MIT, CALCE, Oxford), eliminating potential systematic biases

from data set-specific parameter variations. Figures 2(g) and 2(h) demonstrate linear model fitting results for the MIT and Oxford datasets at selected cycles. As the battery ages, the K and b values change systematically, reflecting the declining trend of battery capacity.

Figures 2(j)–2(l) show feature correlation results across the three datasets. All extracted features achieve correlation coefficients exceeding 0.95 with actual SOH, demonstrating that the proposed feature extraction methodology effectively preserves the physical degradation characteristics of battery degradation while maintaining statistical robustness.

4 Methodology

4.1 Overall architecture design

As shown in Fig. 3, the proposed estimation system integrates advanced health feature extraction with parallel deep learning architectures to achieve robust SOH estimation under dynamic operating conditions. The parallel architecture synergistically leverages the complementary strengths of two distinct deep learning paradigms:

$$\begin{cases} \mathbf{H}_{\text{TCN}} = \mathcal{F}_{\text{TCN}}(\mathbf{X};\theta_{\text{TCN}}) \\ \mathbf{H}_{\text{Trans}} = \mathcal{F}_{\text{Trans}}(\mathbf{X};\theta_{\text{Trans}}) \\ \mathbf{H}_{\text{fused}} = \mathcal{F}_{\text{fusion}}(\mathbf{H}_{\text{TCN}},\mathbf{H}_{\text{Trans}};\theta_{\text{fusion}}) \\ \hat{y} = \mathcal{F}_{\text{output}}(\mathbf{H}_{\text{fused}};\theta_{\text{output}}) \end{cases}, \tag{10}$$

where $\mathbf{X}$ are the input health features, $\mathbf{H}_{\text{TCN}}$ and $\mathbf{H}_{\text{Tran}}$ are the output features from the TCN and Transformer branches respectively, $\mathbf{H}_{\text{fused}}$ is the fused feature representation, $\hat{y}$ is the predicted SOH values and θ denotes the trainable parameters of each module.

4.2 TCN branch for multi-scale local pattern recognition

The TCN branch employs a cascaded structure of residual temporal convolutional blocks, designed to capture battery degradation patterns across multiple temporal scales. Each TCN block utilizes dilated convolutions with exponentially increasing dilation rates. Specifically, the dilation rate for block i is defined as $d_i = 2^{i-1}$, enabling a large receptive field while maintaining linear parameter growth in the number of parameters.

4.3 Transformer branch for global dependencies

The Transformer branch captures long-term dependencies and global degradation trends through multi-head self-attention mechanisms. Input features are first projected to higher-dimensional space and augmented with positional encodings. The multi-head self-attention mechanism then computes relationships between all feature positions simultaneously:

$$\text{Attention}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \tag{11}$$

Input features are augmented with positional encoding to preserve temporal order information.

4.4 Attention-gated feature fusion

The attention-gated fusion module represents a key innovation for integrating complementary information

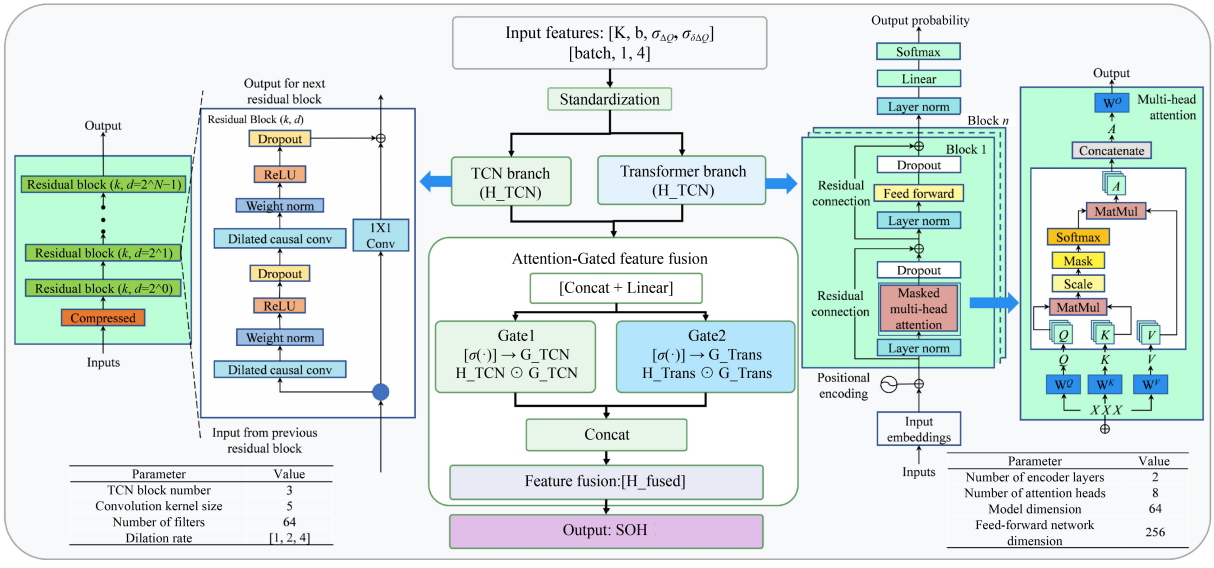

Fig. 3 Architecture diagram of parallel TCN-Transformer hybrid network with attention-gated fusion.

from both branches. This mechanism addresses the critical challenge of optimally combining local and global feature representations while preserving essential degradation information.

The fusion module computes adaptive importance weights for each branch based on current input features:

$$\begin{cases} G_{\text{TCN}} = \sigma(\text{Conv1D}(\text{ReLU}(\text{Conv1D}(\mathbf{H}_{\text{TCN}} \oplus \mathbf{H}_{\text{Trans}})))) \\ G_{\text{Trans}} = \sigma(\text{Conv1D}(\text{ReLU}(\text{Conv1D}(\mathbf{H}_{\text{Trans}} \oplus \mathbf{H}_{\text{TCN}})))) \end{cases}, \tag{12}$$

where $\oplus$ denotes concatenation, σ is sigmoid activation. The gating mechanism allows dynamic emphasis of different pathways based on degradation patterns.

To enhance information exchange between branches, bidirectional cross-attention is applied:

$$\begin{cases} F = \text{Attention}(\mathbf{H}^{\text{g}}_{\text{TCN}}, \mathbf{H}^{\text{g}}_{\text{Trans}}, \mathbf{H}^{\text{g}}_{\text{Trans}}) \\ F = \text{Attention}(\mathbf{H}^{\text{g}}_{\text{Trans}}, \mathbf{H}^{\text{g}}_{\text{TCN}}, \mathbf{H}^{\text{g}}_{\text{TCN}}) \end{cases}, \tag{13}$$

Model training uses mean squared error loss with L1/L2 regularization:

$$L_{\text{Reg}} = \frac{1}{B}\sum_{i=1}^{B}(y_i - \hat{y}_i)^2 + \lambda_1 \sum_i |\theta_i| + \lambda_2 \sum_i \theta_i^2. \tag{14}$$

Detailed network architectures, mathematical formulations, and hyperparameter configurations are provided in ESM S3.

5 Experimental validation

5.1 Validation experimental setup

These three datasets, varying in sizes, each requires distinct division strategies. The MIT dataset comprises five cells (Cell_91, Cell_94, Cell_100, Cell_119, and Cell_124) with different fast-charging protocols. Health features and capacity data were extracted separately for each cell. Specifically, Cell_91, Cell_94, Cell_100, and Cell_119 (6483 samples in total) were designated as the training set, while Cell_124 (1800 samples) served as the test set, yielding a 78.2%/21.8% training-test split. For the cyclic dynamic discharge and stationary process, data from a single cell (CX2-3, 1264 samples) was used, with 50% randomly selected for training and the remaining 50% for testing. The Oxford dataset, under dynamic discharge conditions, includes six cells (Cell_1/2/3/6/7/8), totaling 360 samples.

To validate the effectiveness of the proposed method, comparative experiments were conducted using classical deep learning methods, including CNN-LSTM [42], ViT [41], TCN [35], and Transformer [17], as benchmarks. Additionally, a tandem fusion model combining TCN and Transformer was introduced to illustrate the advantages of the parallel framework. To ensure fairness, all benchmark methods utilized the same feature inputs and the same dataset partitioning strategies:

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}, \\ \text{MAPE} &= \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y_i}\right|/y_i \times 100\%, \\ R^2 &= 1 - \sum_{i=1}^{n}(y_i - \widehat{y_i})^2 / \sum_{i=1}^{n}(y_i - \bar{y})^2, \end{aligned} \tag{15}$$

where y_i represents the ground truth SOH of the ith sample.

5.2 Performance comparison results

Quantitative results are presented in Table 1, and qualitative performance (SOH prediction curves and error distributions) is visualized in Fig. 4. The proposed parallel TCN-Transformer hybrid architecture consistently outperforms all benchmark methods, including ViT, across diverse dynamic operating conditions.

Table 1 Performance comparison across datasets

Dataset	Method	MAE/%	RMSE/%	R^2
MIT	CNN-LSTM	0.82	0.88	0.954
	ViT	0.69	0.75	0.966
	TCN	0.62	0.67	0.973
	Transformer	0.63	0.70	0.971
	TCN-Trans (Serial)	0.79	1.03	0.937
	Proposed	0.47	0.55	0.982
CALCE	CNN-LSTM	1.48	1.70	0.969
	ViT	0.87	1.17	0.985
	TCN	0.53	0.75	0.994
	Transformer	0.66	0.94	0.990
	TCN-Trans (Serial)	0.64	0.83	0.992
	Proposed	0.44	0.68	0.995
Oxford	CNN-LSTM	0.95	1.27	0.958
	ViT	0.79	0.93	0.977
	TCN	1.05	1.29	0.957
	Transformer	0.53	0.66	0.988
	TCN-Trans (Serial)	0.52	0.66	0.988
	Proposed	0.48	0.61	0.991

5.2.1 CC discharge (MIT Dataset)

The proposed method precisely tracks capacity degradation trends, with prediction errors concentrated

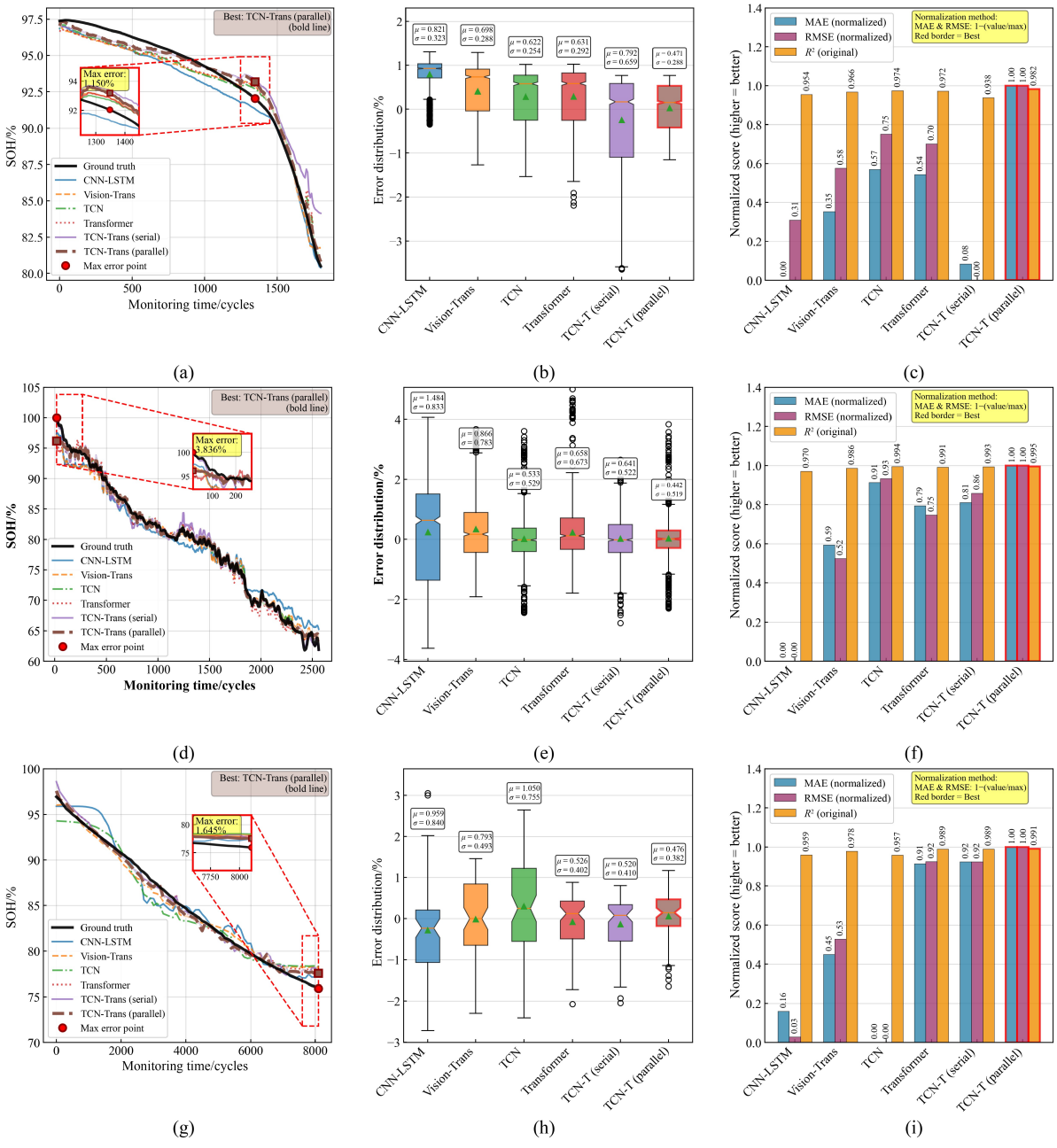

Fig. 4 Comparison of SOH prediction performance of different models on three data sets.

(a)–(c) MIT dataset; (d)–(f) CX2-3 dataset; (g)–(i) Oxford dataset (MAE/RMSE (with percentage dimension) are normalized to a dimensionless 0–1 scale (higher score = better performance) to resolve dimension inconsistency and clarify differences between closely clustered values; R^2 (dimensionless) is displayed as the original value.)

within ±1% (Figs. 4(a)–4(c)). Compared with the best single-model TCN, it reduces MAE from 0.62% to 0.47% (a reduction of 24.2%) and RMSE from 0.67% to 0.55% (a reduction of 17.9%), while R^2 increases from 0.973 to 0.982, indicating higher model fitting accuracy. The TCN-Transformer serial model performs worst, with a MAE of 0.79% and an RMSE of 1.03%, confirming that sequential feature fusion fails to fully leverage TCN's local pattern capture and Transformer's global dependency modeling. ViT and Transformer achieve intermediate performance (ViT: MAE = 0.69%, RMSE = 0.75%; Transformer: MAE = 0.63%, RMSE = 0.70%), but both lack the synergistic multi-scale feature modeling capability of the proposed parallel architecture, as reflected in their wider error distributions in Fig. 4(b).

5.2.2 Periodic measurement scenarios (CALCE Dataset)

Under complex cyclic discharge patterns, the proposed method maintains high stability, with prediction curves closely matching the ground truth (Figs. 4(d)–4(f)). Compared to TCN, MAE decreases from 0.53% to 0.44% (a reduction of 9.4%), RMSE from 0.75% to 0.68%, and R^2 increases to 0.995, approaching perfect fitting. Transformer performs slightly worse than TCN (MAE = 0.66%) but outperforms CNN-LSTM (MAE = 1.48%), due to its ability to capture long-term periodic dependencies. The TCN-Transformer serial model performs poorly (MAE = 0.64%, RMSE = 0.83%) from insufficient sequential feature interaction. ViT improves over CNN-LSTM (MAE = 0.87%, RMSE = 1.17%) but still lags behind the proposed method due to its limited adaptation to irregular feature fluctuations, resulting in a wider error distribution than the proposed method (Fig. 4(e)).

5.2.3 Complex dynamic measurement (Oxford Dataset)

The proposed method demonstrates strong anti-interference capability, producing the narrowest error distribution, as is evident in Figs. 4(g)–4(i). Relative to the Transformer baseline, MAE decreases from 0.53% to 0.48% (a reduction of 9.4%), RMSE from 0.66% to 0.61% (a reduction of 7.5%), and R^2 increases to 0.991. TCN performs poorly on this small-sample, complex dynamic dataset (MAE = 1.05%, RMSE = 1.29%), because dilated convolutions fail to capture long-range dependencies with limited data. The serial TCN-Transformer model shows no fusion advantage (MAE = 0.52%), confirming sequential fusion inefficiency. ViT outperforms CNN-LSTM (MAE: 0.79% versus 0.95%) and TCN but remains inferior to the proposed method and Transformer due to its reliance on global visual feature mapping, which limits its ability to capture local temporal fluctuations in dynamic discharge, leading to higher RMSE (0.93%).

In summary, across all three data sets, the proposed parallel TCN-Transformer architecture achieves optimal performance, with average MAE ranging from 0.47% to 0.52% and RMSE constantly below 1%. It significantly outperforms serial fusion models and single-network models, including ViT. Its core advantage lies in the parallel structure that synergizes TCN's multi-scale local feature capture capability with Transformer's global dependency modeling. The attention-gated fusion module further enables adaptive feature weighting, resulting in a highly precise and robust estimation method for battery SOH under dynamic operating conditions.

5.3 Ablation study

5.3.1 Feature combination impact

Table 2 systematically presents the effects of different feature combinations on SOH prediction performance, with all evaluations conducted using the complete parallel TCN-Transformer hybrid network. Corresponding prediction results and performance metrics are visualized in Fig. 5 (CX2-3 and Oxford Datasets).

The experimental results show that the full feature combination achieves the best performance across all evaluation metrics, delivering the lowest MAE and RMSE and the highest coefficient of determination (R^2). This confirms the critical role of using the complete feature set in improving prediction accuracy.

When comparing individual features, combining trend features (K + b) with statistical features ($\sigma_{\Delta Q}$ + $\sigma_{\delta\Delta Q}$) significantly improves prediction performance, with the trend feature pair demonstrating the strongest individual contribution. This highlights a complementary relationship among features with similar characteristics. Moreover, integrating three feature types together further

Table 2 Impact of different feature combinations on prediction performance

Feature combination	MIT			CX2-3			Oxford		
	MAE/%	RMSE/%	R^2	MAE/%	RMSE/%	R^2	MAE/%	RMSE/%	R^2
K	1.79	2.05	0.7535	1.35	2.3	0.9451	0.94	1.12	0.9681
b	1.61	2.35	0.6758	1.86	2.45	0.938	0.89	1.06	0.9711
$\sigma_{\Delta Q}$	0.83	1.03	0.938	1.41	1.79	0.967	1.14	1.39	0.9508
$\sigma_{\delta\Delta Q}$	0.68	0.88	0.9541	2.18	2.57	0.9316	0.95	1.11	0.9687
K+b	0.61	0.66	0.9741	0.97	1.82	0.9659	0.78	0.99	0.9749
$\sigma_{\Delta Q}$ +$\sigma_{\delta\Delta Q}$	0.63	0.87	0.9554	1.11	1.48	0.9772	1.21	1.56	0.9381
$K + b + \sigma_{\Delta Q}$	0.56	0.66	0.9744	0.79	0.96	0.9905	0.82	0.96	0.9764
$K + b + \sigma_{\delta\Delta Q}$	0.58	0.67	0.9739	1.15	1.7	0.9703	0.79	0.93	0.978
All features	0.47	0.55	0.982	0.44	0.7	0.995	0.48	0.61	0.991

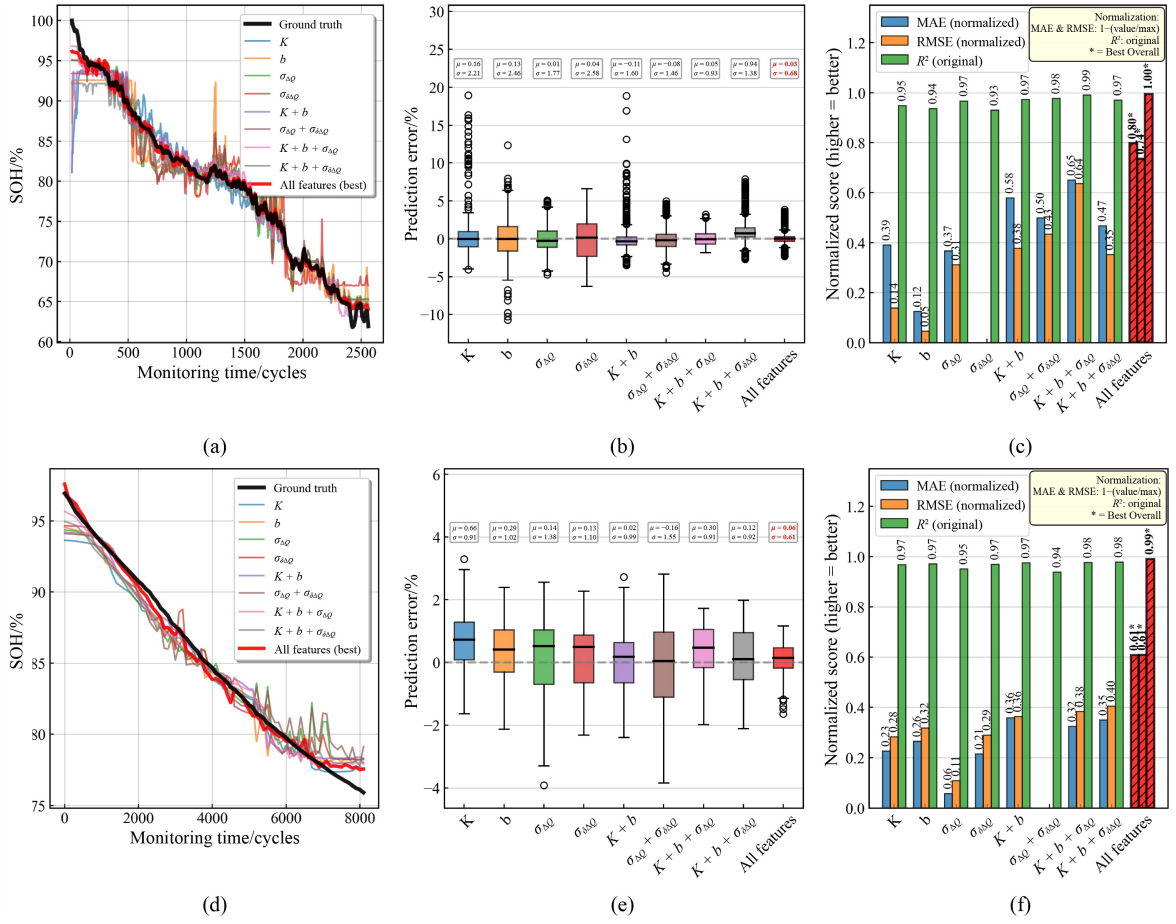

Fig. 5 Impact of different feature combinations on prediction performance.

(a)–(c) CX2-3 dataset; (d)–(f) Oxford dataset (MAE/RMSE (with percentage dimension) are normalized to a dimensionless 0–1 scale (higher score = better performance) to resolve dimension inconsistency and clarify differences between closely clustered values; R^2 (dimensionless) is displayed as the original value.)

enhances model performance, approaching the results achieved by the full feature set. This indicates that this feature subset could serve as an effective option for designing lightweight models without substantial accuracy loss.

Overall, these findings not only validate the effectiveness of the extracted features but also underscore the importance of multi-feature fusion strategies for improving SOH prediction accuracy.

5.3.2 Fusion strategy analysis

The experimental results across the three datasets demonstrate the superiority of dynamic feature weighting strategies over simple concatenation for SOH prediction. As shown in Table 3, the Attention-Gated fusion consistently outperforms Cross-Attention and Simple Concatenation methods.

On the MIT dataset, the Attention-Gated method achieves reductions of 42% and 38% in MAE and RMSE respectively, while improving R^2 from 0.953 to 0.982. For the CALCE dataset, MAE and RMSE decrease by 51% and 44%, respectively, while R^2 reaching 0.995, indicating near-perfect model fit. On the Oxford dataset, Attention-Gated reduces MAE and RMSE by 55% and 52% respectively, and increases R^2 by 3.3% compared to Simple Concatenation. Corresponding prediction curves error distributions are illustrated in Fig. 6 (CX2-3 and Oxford Datasets).

The superior performance stems of the Attention-Gated mechanism arise from its dynamically allocate weights, emphasizing critical features while suppressing redundant or noisy signals. In contrast, Cross-Attention exhibits comparable R^2 values in some cases (e.g., CALCE: 0.993 versus 0.995; Oxford: 0.96 versus 0.991) but shows higher prediction errors, suggesting less effective feature selection. Simple Concatenation performs worst due to its fixed weighting, which fail to exploit feature

Table 3 Impact of feature weight allocation methods prediction performance

Feature weight allocation	MIT			CX2-3			Oxford		
	MAE/%	RMSE/%	R^2	MAE/%	RMSE/%	R^2	MAE/%	RMSE/%	R^2
Simple concatenation	0.81	0.89	0.953	0.89	1.21	0.984	1.06	1.26	0.959
Cross-attention	0.59	0.69	0.972	0.8	0.8	0.993	0.99	1.24	0.96
Attention-gated	0.47	0.55	0.982	0.44	0.68	0.995	0.48	0.61	0.991

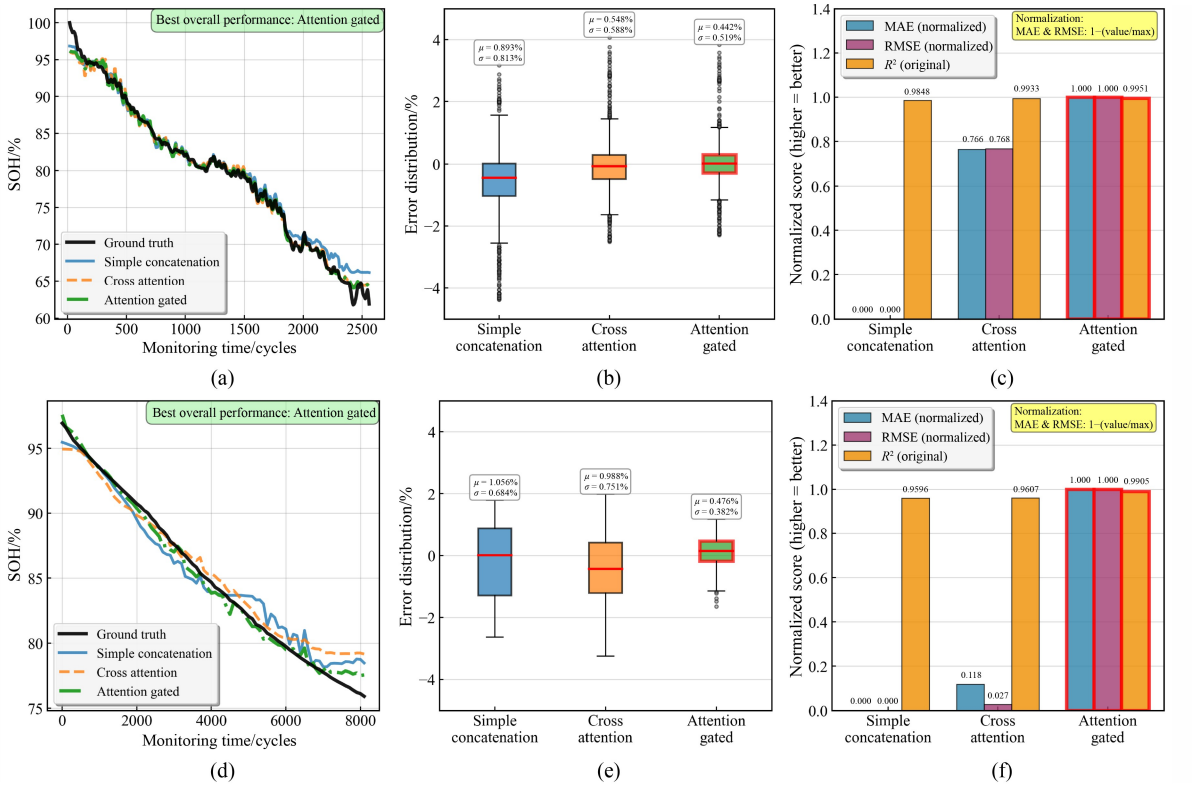

Fig. 6 Impact of feature weight allocation methods on prediction performance.

(a)–(c) CX2-3 dataset; (d)–(f) Oxford dataset. (MAE/RMSE (with percentage dimension) are normalized to a dimensionless 0–1 scale (higher score = better performance) to resolve dimension inconsistency and clarify differences between closely clustered values; R^2 (dimensionless) is displayed as the original value.)

complementarity (e.g., $R^2 = 0.984$ on CALCE dataset).

These results confirm that Attention-Gated fusion effectively balances inter-feature dependencies and noise suppression, offering a robust framework for multi-modal SOH prediction under diverse operating conditions. The consistent improvements across all datasets further demonstrate the generalizability and reliability of this method across different battery types and measurement conditions.

5.3.3 Attention weight distribution analysis

Analysis of attention weight distributions reveals distinct feature fusion strategies across the three datasets, as illustrated in Table 4 and Fig. 7.

Table 4 Attention weight distribution analysis

Items	MIT	CX2-3	Oxford
SOH-range	0.80–0.97	0.61–1.00	0.75–0.97
TCN-weight-mean	13.05%	33.75%	54.79%
Transformer-weight-mean	93.84%	76.81%	61.83%
TCN-weight-std	$\sigma = 0.26\%$	$\sigma = 0.40\%$	$\sigma = 0.45\%$
Transformer-weight-std	$\sigma = 0.20\%$	$\sigma = 0.36\%$	$\sigma = 0.43\%$
TCN-SOH-correlation	−0.87	+0.89	+0.91
Transformer-SOH-correlation	+0.75	−0.82	−0.94

The MIT Dataset: The Transformer branch dominates, with an average weight of 93.84% ± 0.20% while the TCN branch contributes only 13.05% ± 0.26%,

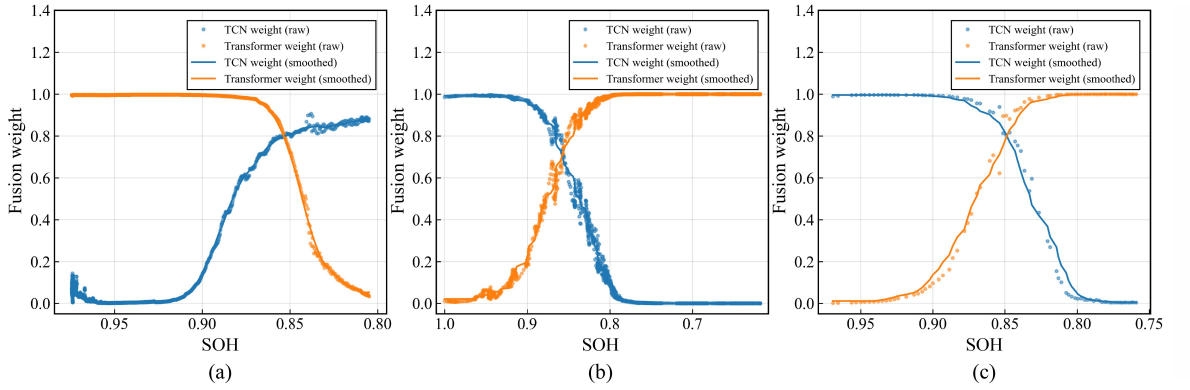

Fig. 7 Attention weight distribution analysis.
(a) MIT dataset; (b) CX2-3 dataset; (c) Oxford dataset.

demonstrating a strong dependency on global feature modeling.

The CALCE Dataset: Weight distribution becomes more balanced, with Transformer weight decreasing to 76.81% ± 0.36% and TCN weight increasing to 33.75% ± 0.40%, indicating that the model begins to prioritize local temporal features.

The Oxford Dataset: TCN weight further rises to 54.79% ± 0.45%, while Transformer weight decreases to 61.83% ± 0.43%, approaching a near-balanced collaboration between local and global feature extraction. Across the three datasets, TCN weight increases by 319.8%, highlighting a clear evolutionary pattern from Transformer-dominated global modeling toward local-global collaborative modeling.

Correlation analysis between attention weights and SOH provides further insights into adaptive fusion mechanisms.

In the MIT dataset, TCN weight demonstrates a strong negative correlation with SOH (r = −0.87), while Transformer weight shows a strong positive correlation (r = +0.75), indicating increased emphasis on local feature detection during degradation.

In the CALCE dataset, this pattern reverses, with TCN weight positively correlated with SOH (r = +0.89) and Transformer weight negatively correlated (r = −0.82), reflecting dynamic weight redistribution across different lifecycle stages.

In the Oxford dataset, the reversal intensifies, with TCN-SOH correlation reaching +0.91 and Transformer-SOH correlation strengthening to −0.94, displaying the most extreme collaborative differentiation.

The trends demonstrate that the model capability adjusts fusion strategies according to data characteristics, providing targeted local-global feature extraction mechanisms for accurate SOH prediction under diverse application conditions.

5.4 Cross-battery transfer learning

To evaluate the generalization ability of the proposed SOH estimation method across battery chemistries, transfer learning experiments were conducted to systematically assess model performance changes under cross-battery-domain training and parameter adaptation, leveraging feature distribution adaptation alignment to address inherent electrochemical differences across battery chemistries through a hierarchical transfer learning strategy.

5.4.1 Transfer learning methodology

The proposed strategy employs a hierarchical approach where low-level feature extractors capture universal temporal patterns, while high-level classifiers learn domain-specific mappings. The method freezes the first N layers to preserve universal representations, and fine-tunes subsequent layers using 30% of target domain data. Optimal frozen-layer configurations are empirically determined: 6 layers for CX2-3 dataset and 5 layers for the Oxford dataset (Figs. 8(b) and 8(d)), reflecting variations in feature distribution similarity between source and target domains.

5.4.2 Cross-domain feature distribution analysis

Voltage feature space analysis reveals significant electrochemical differences across datasets: MIT exhibits a positive linear slope (K = 0.986), whereas CALCE and Oxford show negative slopes (K = −0.958 and K = −0.985, respectively). These differences produce sign reversals in the $\Delta Q(V)$ versus $Q(V)^2$ relationship and distinct statistical baselines for $\sigma_{\Delta Q}$ and $\sigma_{\delta\Delta Q}$. Despite these variations, all four extracted features maintain strong correlation with SOH (r > 0.95) across

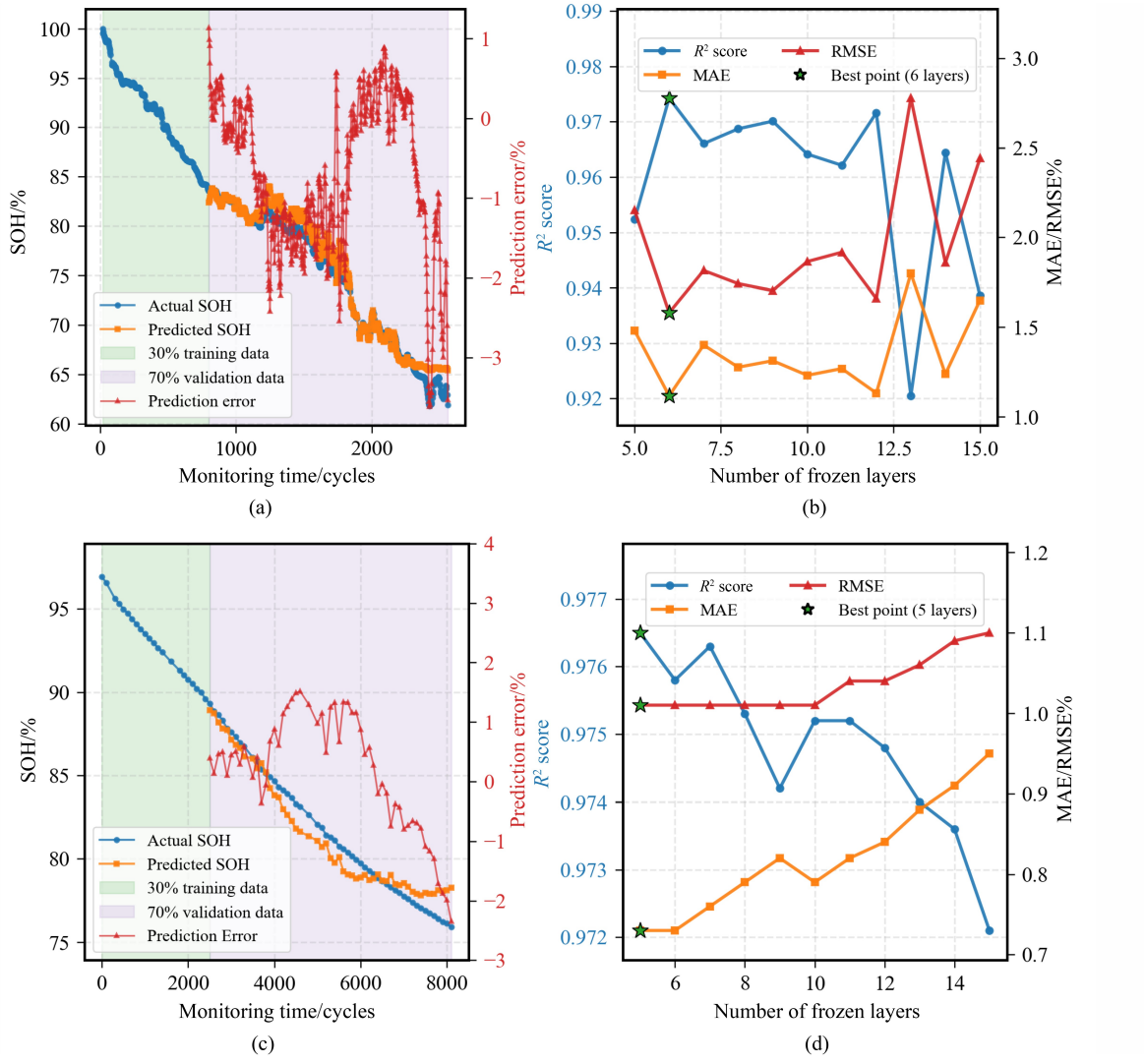

Fig. 8 Transfer learning results across battery types.
(a)–(b) CX2-3 dataset; (c)–(d) Oxford dataset.

datasets, confirming their universal relevance as health indicators.

5.4.3 Transfer performance and analysis

Cross-domain results demonstrate the reasonable cross-chemistry generalization potential. For MIT→CX2-3, MAE increases from 0.44% to 1.11% (R^2 = 0.974), while MIT→Oxford exhibits MAE increase from 0.48% to 0.79% (R^2 = 0.975). The performance reduction primarily stems from inherent feature distribution differences rather than methodological limitations, as evidenced by the preserved high feature-SOH correlations. This approach offers material-agnostic design applicable to scenarios lacking detailed electrochemical parameters, although some cross-chemistry performance degradation remains inevitable due to fundamental material behavioral differences (Table 5).

Table 5 Cross-cell type transfer learning results

Source → Target	MAE/%	RMSE/%	R^2
MIT → CX2-3	1.11	1.58	0.974
MIT → Oxford_Cell8	0.79	1.01	0.975

6 Conclusion and future work

This paper presents a parallel TCN-Transformer hybrid architecture for lithium-ion battery SOH estimation under dynamic operating conditions. The key contributions include:

1) Dynamic feature extraction framework: Four highly correlated health features (correlation > 0.95) are derived from irregular discharge patterns using dual polarization modeling and standardized voltage segmentation, enabling accurate representation of battery degradation.

2) Parallel deep learning architecture: The architecture synergistically combines TCN's multi-scale temporal pattern recognition with Transformer's global dependency modeling, capturing both local and long-term degradation patterns.

3) Attention-gated fusion mechanism: Adaptive feature integration is achieved through attention-based, with systematic weight evolution demonstrating the model's capability to prioritize targeted features.

Experimental validation across MIT, CALCE, and Oxford datasets confirms the superior performance of the proposed method, achieving RMSE below 1% in all cases. The maximum MAE is only 0.48% and maximum RMSE is only 0.68%, representing substantial improvements over conventional approaches. Attention weight analysis reveals correlations with aging mechanisms, providing interpretability for model decisions through systematic transformation from global dominance (MIT: 93.84% Transformer weight) toward local-global collaborative modeling (Oxford: 54.79% TCN weight), consistent with battery aging mechanisms.

The feature extraction framework proposed herein is designed for direct integration into electric vehicle (EV) battery management systems (BMS). While the four extracted health features consistently exhibit high correlation (> 0.95) with SOH across the three studied datasets (MIT, CALCE, Oxford), further validation is required to confirm applicability across all EV scenarios. The current study focuses on LFP and LCO battery chemistries. Future work should extend validation to NCM ternary batteries, which dominate modern EV applications and exhibit distinct aging mechanisms that may affect feature extraction and model performance. Additionally, investigation of multi-cell battery pack health assessment and extreme-temperature operation will further enhance practical applicability for next-generation ESSs.

Overall, the proposed methodology provides a high-accuracy, interpretable framework for battery health monitoring in dynamic environments, offering a robust foundation for intelligent BMS implementation.

Acknowledgements This work was financially supported by the Major Science and Technology Projects for Independent Innovation of China FAW Group Co., Ltd. (Grant Nos. 20230301016ZD and 20230301021ZD), and the 2024 Key Laboratory Open Project Fund of the Key Laboratory of MIIT for Intelligent Products Testing and Reliability (IPQERT (KL)-2024-01).

Competing Interests The authors declare that they have no competing interests.

Electronic supplementary material Supplementary material is available in the online version of this article at https://doi.org/10.1007/s11708-026-1046-4 and is accessible for authorized users.

References

1. Lu J, Xiong R, Tian J, et al. Deep learning to estimate lithium-ion battery state of health without additional degradation experiments. Nature Communications, 2023, 14(1): 2760
2. Li X, Yu D, Søren Byg V, et al. The development of machine learning-based remaining useful life prediction for lithium-ion batteries. Journal of Energy Chemistry, 2023, 82: 103–121
3. Chou J H, Wang F K, Lo S C. Predicting future capacity of lithium-ion batteries using transfer learning method. Journal of Energy Storage, 2023, 71: 108120
4. Zhao J, Han X, Ouyang M, et al. Specialized deep neural networks for battery health prognostics: Opportunities and challenges. Journal of Energy Chemistry, 2023, 87: 416–438
5. Wen P, Ye Z S, Li Y, et al. Physics-informed neural networks for prognostics and health management of lithium-ion batteries. IEEE Transactions on Intelligent Vehicles, 2024, 9(1): 2276–2289
6. Lin M, Wu D, Chen S, et al. Battery health prognosis based on sliding window sampling of charging curves and independently recurrent neural network. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1–9
7. Sun Y, Xie H, Diao Q, et al. A novel SOH estimation method with attentional feature fusion considering differential temperature features for lithium-ion batteries. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1–11
8. Attia P M, Bills A, Brosa Planella F, et al. Review—"Knees" in lithium-ion battery aging trajectories. Journal of the Electrochemical Society, 2022, 169(6): 060517
9. Deng Z, Hu X, Li P, et al. Data-driven battery state of health estimation based on random partial charging data. IEEE Transactions on Power Electronics, 2022, 37(5): 5021–5031
10. Deng Z, Hu X, Lin X, et al. General discharge voltage information enabled health evaluation for lithium-ion batteries. IEEE/ASME Transactions on Mechatronics, 2021, 26(3): 1295–1306
11. Guo F, Huang G, Zhang W, et al. State of health estimation method for lithium batteries based on electrochemical impedance spectroscopy and pseudo-image feature extraction. Measurement, 2023, 220: 113412
12. Gao Z, Jin Y, Zhang Y, et al. Static EIS multi-frequency feature points combined with WOA-BP neural network for Li-ion battery SOH estimation. Measurement, 2025, 253: 117780
13. Zhang C, Tu L, Zhou Z, et al. Estimating lithium-ion battery health using hybrid attention networks and multisource data. IEEE Sensors Journal, 2024, 24(19): 30375–30385
14. Xiao R, Hu Y, Jia X, et al. A novel estimation of state of charge for the lithium-ion battery in electric vehicle without open circuit voltage experiment. Energy, 2022, 243: 123072
15. S V, Che H S, Selvaraj J, et al. State of health (SOH) estimation methods for second life lithium-ion battery-review and

challenges. Applied Energy, 2024, 369: 123542
16. Wei Z, Sun X, Wang J, et al. State of health estimation of lithium-ion batteries based on multihealth features fusion and improved group method of data handling. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1–15
17. Ardakani A H, Abdollahian S A, Abdollahi F. NARX transformer: A dynamic model for leveraging multicycle data in long-term battery state of health estimation. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1–8
18. Greenbank S, Howey D. Automated feature extraction and selection for data-driven models of rapid battery capacity fade and end of life. IEEE Transactions on Industrial Informatics, 2022, 18(5): 2965–2973
19. Deng Z, Lin X, Cai J, et al. Battery health estimation with degradation pattern recognition and transfer learning. Journal of Power Sources, 2022, 525: 231027
20. Huang H, Meng J, Wang Y, et al. A comprehensively optimized lithium-ion battery state-of-health estimator based on local coulomb counting curve. Applied Energy, 2022, 322: 119469
21. Meng J, Cai L, Luo G, et al. Lithium-ion battery state of health estimation with short-term current pulse test and support vector machine. Microelectronics and Reliability, 2018, 88–90: 1216–1220
22. Sun J, Ren S, Shang Y L, et al. A novel fault prediction method based on convolutional neural network and long short-term memory with correlation coefficient for lithium-ion battery. Journal of Energy Storage, 2023, 62: 106811
23. Hou J, Su T N, Gao T, et al. Early prediction of battery lifetime for lithium-ion batteries based on a hybrid clustered CNN model. Energy, 2025, 319: 134992
24. Lin M, You Y, Meng J, et al. lithium-ion batteries SOH estimation with multimodal multilinear feature fusion. IEEE Transactions on Energy Conversion, 2023, 38(4): 2959–2968
25. He Y, Bai W, Wang L, et al. SOH estimation for lithium-ion batteries: An improved GPR optimization method based on the developed feature extraction. Journal of Energy Storage, 2024, 83: 110678
26. Wei L, Sun Y, Diao Q, et al. State of health estimation of lithium-ion batteries based on stacked-LSTM transfer learning with bayesian optimization and multiple features. IEEE Sensors Journal, 2024, 24(22): 37607–37619
27. Shao K, Zhai C, Zhang C, et al. An FE-S-BiLSTM and heatmap-based state-of-health estimation method for lithium-ion batteries. IEEE Sensors Journal, 2025, 25(10): 17727–17738
28. Cui Y, Chen Y. Prognostics of lithium-ion batteries based on capacity regeneration analysis and long short-term memory network. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1–13
29. Xue K, Yang J, Yang M, et al. An improved generic hybrid prognostic method for RUL prediction based on PF-LSTM learning. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1–21
30. Feng J, Cai F, Zhan X, et al. A novel state-of-health prediction and assessment strategies for high-capacity mining lithium-ion batteries based on multi-indicator. Journal of the Electrochemical Society, 2024, 171(5): 050514
31. Xia F, Tang C, Chen J. Online two-dimensional filter for anti-interference aging features extraction to accurately predict the battery health. Measurement, 2024, 234: 114758
32. Marri I, Petkovski E, Cristaldi L, et al. Comparing machine learning strategies for SOH estimation of lithium-ion batteries using a feature-based approach. Energies, 2023, 16(11): 4423
33. Xie Q, Liu R, Huang J, et al. Residual life prediction of lithium-ion batteries based on data preprocessing and a priori knowledge-assisted CNN-LSTM. Energy, 2023, 281: 128232
34. Gao J, Yang D, Wang S, et al. State of health estimation of lithium-ion batteries based on Mixers-bidirectional temporal convolutional neural network. Journal of Energy Storage, 2023, 73: 109248
35. Li L, Li Y J, Mao R Z, et al. Remaining useful life prediction for lithium-ion batteries with a hybrid model based on TCN-GRU-DNN and dual attention mechanism. IEEE Transactions on Transportation Electrification, 2023, 9(3): 4726–4740
36. Zhao W, Ding W, Zhang S, et al. A deep learning approach incorporating attention mechanism and transfer learning for lithium-ion battery lifespan prediction. Journal of Energy Storage, 2024, 75: 109647
37. Akbar K, Zou Y, Awais Q, et al. A machine learning-based robust state of health (SOH) prediction model for electric vehicle batteries. Electronics, 2022, 11(8): 1216
38. He T, Gong Z. State of health estimation for lithium-ion batteries using a hybrid neural network model with multi-scale convolutional attention mechanism. Journal of Power Sources, 2024, 609: 234680
39. Zhang C, Tu L, Yang Z, et al. A CMMOG-based lithium-battery SOH estimation method using multi-task learning framework. Journal of Energy Storage, 2025, 107: 114884
40. Chen L, Xie S, Lopes A M, et al. A new SOH estimation method for lithium-ion batteries based on model-data-fusion. Energy, 2024, 286: 129597
41. Zhao W, Ding W, Zhang S, et al. Enhancing lithium-ion battery lifespan early prediction using a multi-branch vision transformer model. Energy, 2024, 302: 131816
42. Bao X, Chen L, Lopes A M, et al. Hybrid deep neural network with dimension attention for state-of-health estimation of Lithium-ion batteries. Energy, 2023, 278: 127734
43. Qian C, Xu B, Xia Q, et al. SOH prediction for lithium-ion batteries by using historical state and future load information with an AM-seq2seq model. Applied Energy, 2023, 336: 120793
44. Ayankoso S, Olejnik P. Time-series machine learning techniques for modeling and identification of mechatronic systems with friction: A review and real application. Electronics, 2023, 12(17): 3669
45. Attia P M, Grover A, Jin N, et al. Closed-loop optimization of fast-charging protocols for batteries with machine learning. Nature, 2020, 578(7795): 397–402
46. Bai T, Wang H. Convolutional transformer-based multiview information perception framework for lithium-ion battery state-

of-health estimation. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1–12

47. Xiong W, Mo Y, Yan C. Online state-of-health estimation for second-use lithium-ion batteries based on weighted least squares support vector machine. IEEE Access: Practical Innovations, Open Solutions, 2021, 9: 1870–1881
48. Sun X, Wei Z, Li Y, et al. New insights into estimating lithium-ion battery cell health leveraging physics-informed machine learning. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 1–15