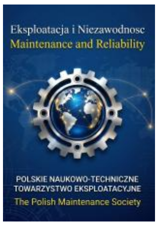



**Eksploatacja i Niezawodnosc - Maintenance and Reliability**

Volume 29 (2027), Issue 1

https://ein.org.pl

The Polish Maintenance Society

# Source-free joint SOC/SOH estimation for lithium-ion batteries via task-specific adapters and interaction attention

Ruhui Fan[1,2], Yi Lyu[1,2,*], Xu Xiao[3], Guiduo Duan[2]

[1] School of Computer, University of Electronic Science and Technology of China Zhongshan Institute, China
[2] School of Computer Science and Engineering, University of Electronic Science and Technology of China, China
[3] School of Automation, Guangdong University of Technology, China
* Corresponding Author: lvyi913001@163.com

*Received: 21 April 2026*
*Revised: 4 May 2026*
*Accepted: 25 June 2026*
*Online: 17 July 2026*



**Abstract**

In some practical applications, source domain data are unavailable during transfer stage due to constraints such as data privacy and commercial confidentiality. This limitation presents a major challenge for cross-condition estimation of lithium-ion batteries. To address this scenario, this paper proposes a Source-Free Joint Estimation framework for the simultaneous prediction of State of Charge (SOC) and State of Health (SOH). The framework disentangles general and task-specific representations through a shared backbone, two parallel task-specific modules, and a subsequent task interaction attention mechanism. During the transfer stage, the backbone is frozen, and only three lightweight task modules are fine-tuned using target-domain data. This training strategy enables parameter-efficient and accurate adaptation while mitigating catastrophic forgetting. Experimental results on the MIT dataset demonstrate that the proposed method achieves improved joint SOC and SOH estimation under diverse charging policies, highlighting its strong transferability in source-free settings.

**Keywords:** l*ithium-ion batteries, transfer learning, joint estimation, parameter-efficient fine-tuning*

*Article citation:*
Fan R, Lyu Y, Xiao X, Duan G, Source-free joint SOC/SOH estimation for lithium-ion batteries via task-specific adapters and interaction attention, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2027: 29(1) http://doi.org/10.17531/ein/225071

## Highlights

- Task interaction attention mechanism captures the deep intrinsic SOC and SOH coupling.
- PEFT enables source-free joint SOC/SOH estimation without catastrophic forgetting.
- Extensive experiments validate robustness across diverse transfer scenarios.

## 1. Introduction

The rapid growth of electric vehicles and renewable energy systems has driven the widespread adoption of lithium-ion batteries, valued for their high energy density and long cycle life [1-2]. For these applications, accurately estimating the State of Charge (SOC) and State of Health (SOH) is essential for ensuring system safety, extending battery lifespan, and enabling efficient energy management. SOC reflects the remaining usable capacity of a battery, whereas SOH quantifies its level of degradation. Given that these two metrics are inherently correlated and governed by similar electrochemical mechanisms, their joint estimation has become a valuable and significant area of research.

Current approaches for SOC and SOH estimation generally fall into two categories: model-based and data-driven. Model-based approaches—including electrochemical [3], equivalent-circuit [4], and empirical models [5]—rely on extensive domain knowledge to construct detailed mathematical representations of battery behavior. Although these methods benefit from strong theoretical foundations, they face notable practical limitations. In particular, accurately modeling batteries with complex and evolving degradation mechanisms remains difficult, often resulting in reduced predictive accuracy in real-world scenarios.

This limitation has driven a growing shift toward data-driven methods [6-7]. Enabled by advances in deep learning, these approaches are capable of automatically extracting informative features from historical data without relying on explicit domain knowledge. Consequently, various architectures have been explored for battery state estimation, such as dual-channel models [8], TCN-LSTM networks [9], adaptive attention mechanisms [10], and CNN-Transformer combinations [11]. To further meet practical testing demands and capture complex degradation dynamics, recent literature has introduced several advanced frameworks. Zhou et al. [12] developed a feature-enhanced ensemble learning method that utilizes partial discharging segments for rapid capacity estimation. Chen et al. [13] proposed a novel model-data-fusion strategy integrating a fractional-order RC circuit with a vision transformer network for highly accurate SOH prediction. Additionally, Bao et al. [14] designed a hybrid prognostic network combining a CNN, a variant LSTM (VLSTM), and a dimensional attention mechanism to improve SOH estimation.

Notably, the aforementioned methods are limited to estimating a single state: either the SOC or the SOH of the lithium-ion battery. In practical applications, however, SOC and SOH are not independent. SOC reflects the battery's short-term dynamic behavior, while SOH characterizes its long-term degradation process. The evolution of SOH inevitably affects SOC estimation accuracy, as battery aging alters its voltage-capacity characteristics. Conversely, inaccurate SOC estimation can lead to improper charging and discharging operations, thereby accelerating degradation and impacting SOH. If the coupling relationship between the two is ignored, prediction results are often unreliable and may even undermine the effectiveness of battery management system(BMS). Therefore, exploring joint SOC and SOH estimation methods not only enhances prediction accuracy and robustness but also provides a more comprehensive and scientific basis for full life-cycle management. For instance, Huang et al. [15] proposed the SOC–SOH Estimation Framework, which achieves parameter sharing through staged training and combines TCN with Bidirectional Gated Recurrent Unit in a charging encoder to capture both local and global features, resulting in more efficient and precise joint estimation. Li et al. [16] developed a joint estimation framework based on Bidirectional Long Short-Term Memory (Bi-LSTM), using Particle Swarm Optimization to tune the Bi-LSTM hyperparameters for high-precision estimation of SOH and terminal voltage, and then integrated an Extended Kalman Filter to further improve SOC prediction performance.

However, the generalization of these models is often compromised by the domain shift arising from varying operating conditions. Transfer learning has thus emerged as a key strategy to adapt models to new domains by transferring knowledge from a label-rich source [17-18]. Chen et al. [19] employed Maximum Mean Discrepancy (MMD) for distribution alignment, a technique Ma et al. [20] advanced by aligning both marginal and conditional feature distributions. Mu et al. [21] adapted this for multi-source domains. Alternatively, Dong et al. [22] proposed a method that uses sample entropy to identify transferable features, integrating them into a deep recurrent network.

Nevertheless, traditional transfer learning's reliance on source data conflicts with real-world constraints like data privacy, creating a source-free challenge. One approach is Source-Free Domain Adaptation (SFDA), an unsupervised method where a pre-trained model adapts to new data without target labels. Pioneering works have applied this to the battery domain: Shen et al. [23] achieved source-free SOC estimation, and Han et al. [24] accomplished source-free SOH estimation. However, SFDA's performance on complex regression tasks can be unstable due to its heavy reliance on the quality of pseudo-labels. In contrast, when target labels are available, supervised Fine-Tuning is a more direct and effective approach. Specifically, Parameter-Efficient Fine-Tuning (PEFT) techniques [25] have emerged as a highly attractive option. By adjusting only a small number of modules, PEFT effectively mitigates catastrophic forgetting and reduces computational costs, making it ideal for source-data-constrained scenarios.

However, whether SFDA or PEFT, existing source-free methods share a fundamental limitation: they treat SOC and SOH estimation as independent tasks. This neglects the inherent physical coupling between a battery's short-term dynamics and its long-term degradation, which ultimately compromises prediction accuracy and reliability. To address this issue, we propose a novel Source-free Joint Estimation Framework (SFJE) that integrates parameter-efficient fine-tuning with joint estimation of SOC and SOH. Our framework is built on a decoupled model with a shared feature extractor backbone. To preserve valuable knowledge learned from the source domain, this backbone remains frozen during the transfer stage. Our key innovation lies in three lightweight and tunable components constructed atop the frozen backbone: two parallel task-specific adapters for SOC and SOH, and a subsequent task interaction attention module. By fine-tuning only these modules, the framework achieves two objectives: the adapters enable parameter-efficient adaptation for each task, while the attention module explicitly captures their interdependence. This design directly overcomes the primary limitation of prior source-free methods—namely, their inability to exploit the intrinsic correlation between SOC and SOH. The main contributions of this paper are as follows:

1. We propose a Source-Free Joint SOC/SOH Estimation framework. By combining PEFT with a modular design, the framework freezes the pre-trained backbone and fine-tunes only lightweight adapters, enabling accurate joint SOC–SOH estimation without source data and effectively mitigating catastrophic forgetting.

2. We introduce a task interaction attention mechanism that models the intrinsic coupling between SOC and SOH. Through cross-attention–based information exchange, this module captures deep dependencies between short-term electrochemical dynamics and long-term degradation, representing a methodological advancement beyond independent-task source-free approaches.

3. We conduct extensive experimental validation on a public dataset, demonstrating the generality and robustness of SFJE. The proposed framework achieves improved SOC and SOH estimation under diverse charging policies and transfer scenarios, establishing its effectiveness and scalability for real-world source-free battery state estimation.

The remainder of this paper is organized as follows: Section 2 describes the dataset and analyzes the correlation between SOC and SOH. Section 3 details our proposed methodology. Section 4 describes the details of the experiments and presents the results. Finally, section 5 summarizes the paper.

## 2. Dataset

### 2.1. Dataset description

To validate the effectiveness of the proposed method, we employed the publicly available lithium-ion battery dataset released by MIT, Stanford University, and the Toyota Research Institute (referred to as the MIT dataset) [26]. This dataset comprises degradation data collected from batteries subjected to four distinct charging protocols. The batteries used in the experiments have a nominal capacity of 1.1 Ah and a nominal voltage of 3.3 V.

In this dataset, the charging protocols for all batteries are denoted as C1(Q1%)–C2. Specifically, this nomenclature indicates a multi-step constant current (CC) process: the battery is first charged at a current rate of C1 until it reaches Q1% of the total capacity, followed by a second charging stage at a rate of C2 until the SOC reaches 80%. Subsequently, a unified 1C constant current-constant voltage (CC-CV) charging strategy is applied to bring all batteries to 100% capacity. The overall charging profile is illustrated in Fig. 1.

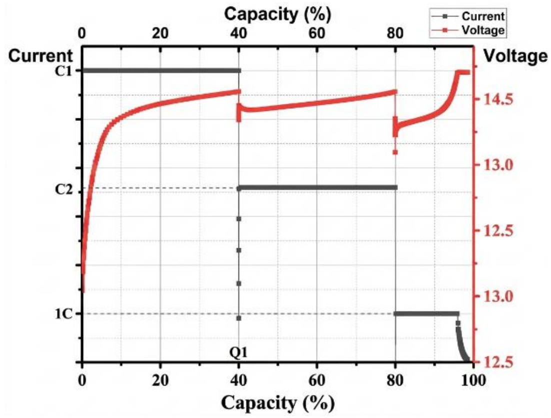



Figure 1. Variations of charging current and voltage against normalized capacity.

Given that the initial charging states vary across the different charging protocols, significant discrepancies exist in the data distributions between domains. To mitigate these domain shifts, we excluded the data corresponding to the 0-80% SOC range--where protocols differ the most--and utilized the charging segments subsequent to 80% capacity for our experiments [27].

Preliminary data analysis reveals that as the battery SOH declines, the duration of the constant current (CC) phase after reaching 80% SOC gradually converges to zero. In contrast, although the duration of the constant voltage (CV) phase decreases, it does not exhibit a vanishing trend and remains sufficiently observable. Consequently, we selected the CV charging data collected after the 80% SOC point for model training and testing [27].

While public fast-charging sessions may occasionally terminate before reaching high SOC levels, the practical relevance of this specific selection is well-supported by common Electric Vehicle (EV) usage patterns. Full charging events that include a complete CV stage routinely occur during widespread scenarios such as overnight residential charging or workplace destination charging [28]. Thus, this data segment remains universally accessible to the onboard Battery Management System (BMS) throughout the regular lifecycle of commercial EVs, ensuring the practical applicability of the proposed framework.

Given that the voltage remains quasi-static during this phase, we selected current and temperature, along with their first-order differentials, as the input features. Furthermore, Min-Max normalization was applied to scale all feature sequences to the range [0, 1] for each battery, as described by the following equation:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where x is the original feature value, $x_{min}$ and $x_{max}$ represent the minimum and maximum values of the feature sequence, respectively, and $x'$ denotes the normalized value.

### 2.2. Quantification of coupling between SOC and SOH

To quantitatively assess the non-linear coupling relationship between the continuous variables of SOC and SOH, this study employs Normalized Mutual Information (NMI) as the metric. Since NMI is defined for discrete variables, the continuous SOC and SOH data must first undergo discretization. Although neural network-based algorithms for the direct estimation of continuous mutual information have emerged recently [29], these methods typically require massive datasets and complex model training. In contrast, the traditional discretization-based approach adopted in this study is characterized by computational simplicity, robust results, and the absence of additional training requirements, rendering it a standard practice widely adopted in engineering applications [30].

#### 2.2.1. Data alignment and analysis set construction

The data in this study is organized into $N$ charge-discharge cycles. In the $n$-th cycle, the data comprises a scalar SOH value, $y_n$, and a time-series of SOC data containing $S$ sampling points. To facilitate mutual information calculation, we first construct an aligned analysis dataset $\mathcal{D}$. This process involves pairing each of the $S$ SOC observations within a cycle with the corresponding single SOH value $y_n$ of that cycle, and subsequently merging data from all $N$ cycles.

This yields a dataset containing $K = N \times S$ paired observations, denoted as $\mathcal{D} = \{(x_k, y_k) \mid k = 1, \ldots, K\}$, where $x_k$ represents SOC and $y_k$ represents SOH. This dataset serves as the basis for the subsequent calculations.

#### 2.2.2. Mutual information calculation

**1. Discretization**

Following data alignment, the continuous variables $X$ and $Y$ are discretized using the equal-width binning method to generate the discrete sets $X' = \{x'_k \mid k = 1, \ldots, K\}$ and $Y' = \{y'_k \mid k = 1, \ldots, K\}$ with $M$ bins. This transformation is implemented by applying the following mapping function $f(x)$ to each data point:

$$f(x) = \begin{cases} M, & if\ x = X_{max} \\ \left\lfloor \frac{x - X_{min}}{W} \right\rfloor + 1, & if\ x < X_{max} \end{cases} \tag{2}$$

where $X_{max}$ and $X_{min}$ represent the maximum and minimum values of the dataset, respectively, and $W = (X_{max} - X_{min})/M$ denotes the calculated width of the equal-width bins. The transformation for variable $Y'$ is performed in an identical manner to that of $X'$.

**2. Entropy**

Entropy quantifies the uncertainty associated with the entire discretized variable and is calculated as follows:

$$H(X') = -\sum_{i=1}^{M} p(x'_i)\, log p(x'_i) \tag{3}$$

where $p(x'_i)$ is the marginal probability of an SOC observation falling into the $i$-th discrete interval, estimated by:

$$p(x'_i) = \frac{1}{K}\sum_{k=1}^{K} \mathbb{N}(x'_k = i) \tag{4}$$

where $\mathbb{N}(\cdot)$ is the indicator function, which equals 1 if the condition is true and 0 otherwise.

**3. Mutual Information (MI)**

Mutual information quantifies the amount of information shared between the two discretized variables:

$$I(X', Y') = \sum_{i=1}^{M}\sum_{j=1}^{M} p(x'_i, y'_j) log \frac{p(x'_i, y'_j)}{p(x'_i)p(y'_j)} \tag{5}$$

where $p(x'_i, y'_j)$ is the joint probability of the SOC and SOH observation pair simultaneously falling into the corresponding discrete intervals $(i, j)$:

$$p(x'_i, y'_j) = \frac{1}{K}\sum_{k=1}^{K} \mathbb{N}(x'_k = i \wedge y'_k = j) \tag{6}$$

**4. Normalized Mutual Information (NMI)**

To obtain a standardized metric within the interval $[0, 1]$, we normalize the mutual information using the arithmetic mean of the entropies, yielding the final NMI value:

$$NMI(X',Y') = \frac{2\,I(X',Y')}{H(X') + H(Y')} \tag{7}$$

An NMI value of 0 indicates that the variables are completely independent, while a value of 1 implies a perfect correlation. This metric effectively quantifies the strength of the complex non-linear coupling between SOC and SOH.

Through the calculations described above, the obtained NMI value between SOC and SOH is approximately **0.6**. This quantitatively confirms the existence of a significant mutual influence between SOC and SOH. This strong coupling relationship serves as the fundamental premise for constructing the joint estimation model in this study. It demonstrates that during the battery aging process, these two state parameters do not evolve independently but are intricately interwoven. This finding implies that a joint model capable of simultaneously observing and learning the dynamic changes of both SOC and SOH can capture critical synergistic effects that independent models typically overlook, thereby achieving more precise joint state estimation.

## 3. Methodology

This section elaborates on the proposed cross-domain framework for joint SOC and SOH estimation. Adopting a transfer learning strategy, the method commences by pre-training a robust joint estimation model on a generic source domain dataset. To align with real-world deployment constraints, the framework then transitions into a strictly "source-free" adaptation phase. In this context, "source-free" does not imply complete independence from source data throughout the lifecycle, but rather the explicit exclusion of historical source data during target domain adaptation to protect data privacy and minimize communication overhead for edge-level BMS. Subsequently, with the source data completely discarded, the model is efficiently adapted to the specific operating conditions of the target domain by leveraging solely the pre-trained parameters and a limited amount of labeled data from the target domain through selective parameter fine-tuning.

### 3.1. SOC-SOH joint estimation framework

To achieve precise joint estimation of battery SOC and SOH, a multi-task learning model rooted in shared feature extraction and a task interaction attention mechanism was designed and implemented. The core philosophy of the model entails initially extracting a shared feature representation, rich in state information, from the input time-series data. Subsequently, features specific to SOC and SOH are decoupled and enhanced through task-specific adaptive modules and inter-task interaction modules. Ultimately, the estimated values for both states are derived via independent estimation heads.

As depicted in Fig. 2, the overall model architecture comprises four logically distinct core components:

1. **Shared Feature Extractor:** A deep network composed of LSTM layers and a Transformer Encoder, responsible for extracting a shared latent representation encapsulating comprehensive state information from the input sequence.

2. **Task-Specific Adapters:** Two parallel, lightweight Adapter modules designed to tailor the general features into task-specific representations for SOH and SOC, respectively.

3. **Task Interaction Attention Module:** A module based on multi-head attention mechanisms, facilitating explicit information exchange and fusion between the specific representations of the two tasks.

4. **State Estimation Heads:** Two independent prediction heads responsible for mapping the interaction-enhanced features to the final predicted values for SOH and SOC.

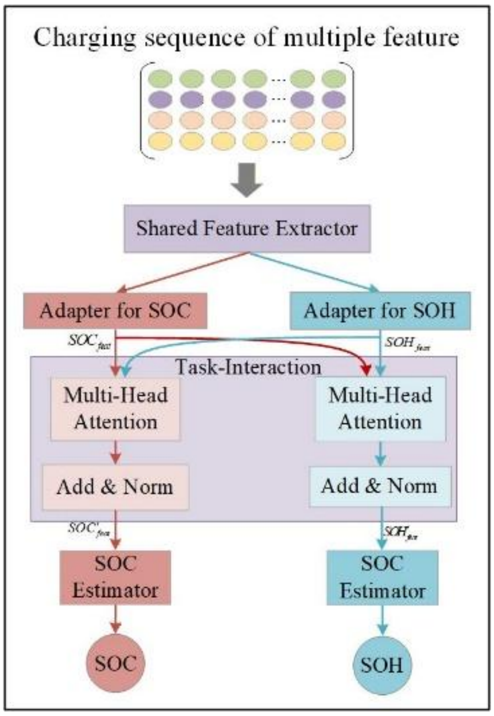



Figure 2. Network overall model diagram.

### 3.2. Shared feature extraction

The objective of this stage is to learn a high-dimensional shared feature representation from the raw input data that simultaneously encapsulates both SOC and SOH information. The structure of this module is illustrated in Fig. 3. Specifically, this module adopts a residual architecture that sequentially processes the input through a projection layer, LSTM units, and encoders, followed by layer normalization to stabilize the training process.

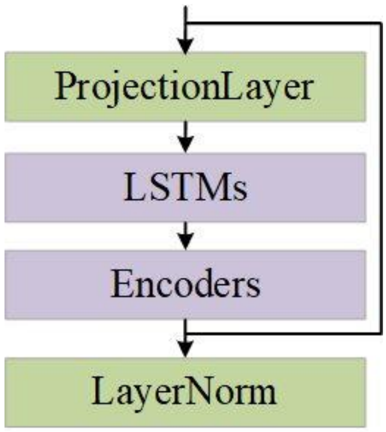



Figure 3. Network diagram of shared feature extractor.

Let the input sequence be denoted as $X \in \mathbb{R}^{B \times L \times C}$, where $B$ represents the batch size, $L$ denotes the sequence length, and $C$ is the dimension of the input features.

#### 3.2.1. Input projection

First, a linear layer is employed to project the raw, low-dimensional input features into the high-dimensional feature space of the model:

$$H_0 = Linear(X) \in \mathbb{R}^{B \times L \times d_{model}} \quad (8)$$

where $d_{model}$ represents the hidden dimension of the model.

#### 3.2.2. Local temporal feature extraction

To further capture the complex local temporal dependencies embedded within the battery sequence data, projected features are fed into an LSTM network [31-32]. Unlike convolutional networks, which are generally limited to processing fixed patterns, LSTMs leverage internal gating mechanisms--comprising input, forget, and output gates--to dynamically learn and retain long-term dependencies within time series, as shown in Fig. 4.

This capability enables the model to effectively identify critical events during the charging process while filtering out irrelevant information. Consequently, the module outputs a hidden state sequence $H_{lstm}$, which is enriched with contextual information:

$$H_{lstm} = LSTM(H_0) \in \mathbb{R}^{B \times L \times d_{model}} \quad (9)$$

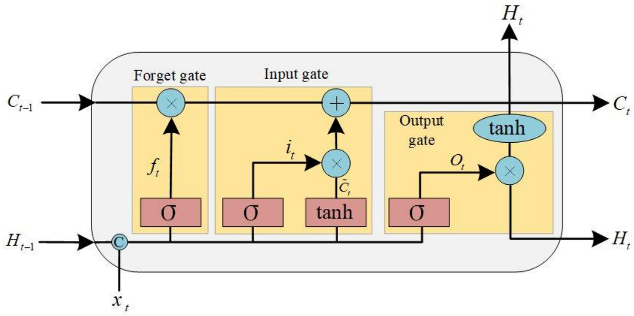



Figure 4. Network diagram of single-layer LSTM.

#### 3.2.3. Global dependency modeling

To capture global temporal dependencies of the sequence beyond the local features extracted by the LSTM, we stack multiple Transformer Encoder layers [33-34]. Each Encoder layer comprises a Multi-Head Attention sub-layer and a Feed-Forward Network sub-layer, as shown in Fig. 5. The self-attention mechanism is pivotal. It projects the input sequence into Query (Q), Key (K), and Value (V) spaces to compute attention scores for each element relative to all others in parallel.

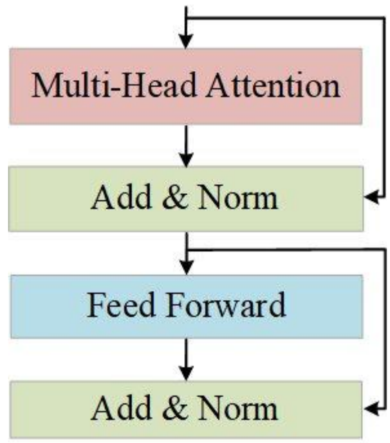



Figure 5. Network diagram of single-layer encoder.

This mechanism enables the model to dynamically assign varying levels of attention to different time steps, thereby directly capturing feature patterns—potentially spanning long temporal intervals—that are critical for battery state prediction. For instance, it can discern how current variations during the initial charging phase influence the thermal response at the end of the charging cycle. Following processing by the Encoder, each vector in the output feature sequence aggregates global context from the entire sequence, significantly enhancing representation capability:

$$H_{enc} = TransformerEncoder(H_{lstm}) \in \mathbb{R}^{B \times L \times d_{model}} \quad (10)$$

### 3.2.4. Feature fusion and output

Finally, we employ a residual connection and Layer Normalization to fuse the global features with the local features, yielding the final shared feature representation:

$$F_{share} = LayerNorm(H_{enc}) + Dropout(H_{lstm}) \in R\mathbb{R}^{B \times L \times d_{model}} (11)$$

## 3.3. Task-specific adaptive module

Despite the coupling relationship between SOC and SOH, significant differences exist regarding their sensitivity to input features and their dynamic evolution characteristics. To enhance the model's feature adaptability in multi-task scenarios, we introduce an Adapter module to facilitate task-specific lightweight parameter tuning.

Fig. 6 depicts the detailed bottleneck structure of the Adapter. It employs a bottleneck design where the input is first projected into a lower-dimensional space via a down-projection layer, processed by a GELU activation, and then restored to the original dimension via an up-projection layer. A residual connection fuses the input with the processed features before Layer Normalization. The Adapter operation, composed of down-projection and up-projection layers, is defined as:

$$Z = GELU(W_{down} F_{share}) \quad (12)$$

$$\tilde{h} = W_{up} Z + F_{share} \quad (13)$$

where $W_{down} \in \mathbb{R}^{d_{model} \times d_{hidden}}$, $W_{up} \in \mathbb{R}^{d_{hidden} \times d_{model}}$, $d_{hidden} \ll d_{model}$.

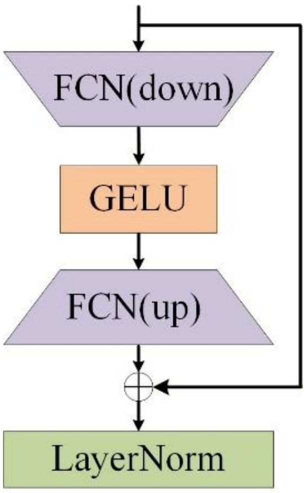



Figure 6. Network diagram of Adaptive module.

This architecture achieves task-specific feature mapping through non-linear compression and residual reconstruction. Consequently, the task-specific representations are obtained as:

$$F_{SOC} = Adapter_{SOC}(F_{share}) \quad (14)$$

$$F_{SOH} = Adapter_{SOH}(F_{share}) \quad (15)$$

This design enables the formation of distinguishable sub-streams for each task within the shared feature space while maintaining parameter efficiency.

## 3.4. Task-Interaction Attention mechanism (TIA)

This stage is pivotal for achieving the effective fusion of SOC and SOH information, thereby enhancing the performance of the joint estimation. We designed a symmetrical bidirectional cross-attention module to facilitate the bidirectional flow of information between the SOC and SOH feature representations, as shown in Fig. 2. The fundamental building block of this architecture is the Multi-Head Attention (MHA) mechanism [35].

### 3.4.1. MHA mechanism

The MHA mechanism operates by projecting the Query ($Q$), Key ($K$), and Value ($V$) into multiple distinct subspaces to compute attention in parallel. This enables the model to simultaneously attend to information from different positions and representational aspects. For a given set of inputs $Q$, $K$, and $V$, the computation process of MHA is detailed as follows.

**1. Linear Projection**

First, $Q$, $K$, and $V$ are projected into $h$ distinct heads via learnable linear transformation matrices. For the $i$-th head, the projections are defined as:

$$Q_i = Q \times W_i^Q, \qquad K_i = K \times W_i^K, \qquad V_i = V \times W_i^V \quad (16)$$

**2. Parallel Attention Calculation**

Within each head, Scaled Dot-Product Attention is computed in parallel:

$$head_i = Attention(Q_i, K_i, V_i) = softmax\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i \quad (17)$$

where $d_k$ denotes the dimension of each head, utilized to scale the dot products to maintain gradient stability.

**3. Concatenation and Final Projection**

The outputs from all $h$ heads are concatenated and subsequently projected back to the original model dimension via a linear transformation matrix $W^O$, yielding the final output of the MHA module:

$$MHA(Q, K, V) = Concat(head_1, \cdots, head_h) W^O \quad (18)$$

### 3.4.2. SOC features attending to SOH features

In this phase, the task-specific SOC features, $F_{SOC}$, serve as the *Query* to interrogate aggregate relevant information encapsulated within the task-specific SOH features, $F_{SOH}$, which simultaneously function as both *Key* and *Value*.

Conceptually, this process can be interpreted as the model identifying and retrieving the global SOH information that is most pertinent to the feature representation at each SOC time step. The calculation is defined as follows:

$$A_{SOC,SOH} = MHA(F_{SOC}, F_{SOH}, F_{SOH}) \tag{19}$$

### 3.4.3. SOH features attending to SOC features

Symmetrically, the SOH features $F_{SOH}$ are used as the query to attend to and aggregate relevant information from the SOC features $F_{SOC}$. The calculation is defined as follows:

$$A_{SOH,SOC} = MHA(F_{SOH}, F_{SOC}, F_{SOC}) \tag{20}$$

### 3.4.4. Feature fusion

Upon completion of the two independent cross-attention computations, the respective attention outputs are fused with the original task-specific features via Residual Connections and Layer Normalization. This step is critical for maintaining the stability of the information flow and mitigating the vanishing gradient problem.

**Final SOC Feature Update:**

$$F'_{SOC} = LayerNorm\left(F_{SOC} + Dropout\left(A_{SOC,SOH}\right)\right) \tag{21}$$

**Final SOH Feature Update:**

$$F'_{SOH} = LayerNorm\left(F_{SOH} + Dropout\left(A_{SOH,SOC}\right)\right) \tag{22}$$

Through this symmetrical bidirectional interaction process, the updated features $F'_{SOC}$ and $F'_{SOH}$ not only preserve the unique information of their respective tasks but also mutually integrate critical contextual information from the counterpart task via the cross-attention mechanism. This achieves synergistic feature enhancement, establishing a solid foundation for the subsequent joint estimation.

### 3.5. Estimator

Following the cross-interaction enhancement of the features, the high-dimensional representations are mapped to the final predicted values via two independent estimation heads. The detailed architecture is illustrated in Fig. 7. Specifically, the model employs a dual-branch structure to decouple the regression tasks for SOC and SOH. Each branch consists of a Tanh layer for non-linear feature transformation, followed by a linear projection layer to reduce dimensionality, and finally, a Sigmoid activation to constrain the predicted states within the normalized range of [0, 1].

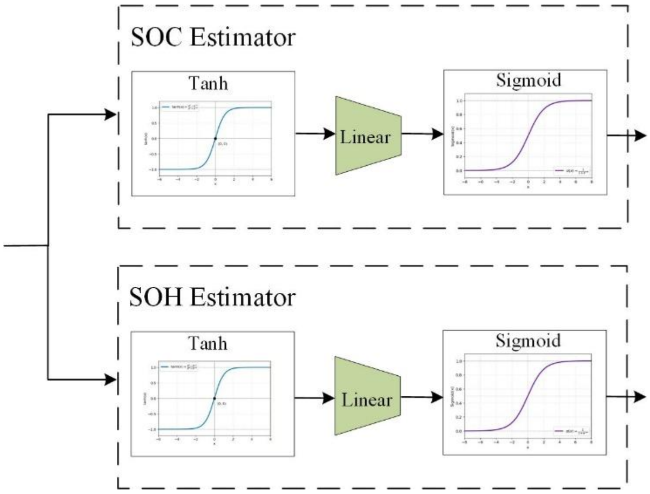



Figure 7. Architecture of the dual-head estimator for joint SOC and SOH prediction.

### 3.5.1. SOC Estimation

The interaction-enhanced SOC features, $F'_{SOC}$, first undergo Layer Normalization before being fed into the SOC estimation head. This head projects the hidden-dimension feature at each time step to an SOC sequence matching the original data dimensionality:

$$y_{soc} = \sigma(W_{soc} tanh(F'_{soc}) + b_{soc}) \tag{23}$$

where $\sigma(\cdot)$ denotes the sigmoid activation function.

### 3.5.2. SOH Estimation

The interaction-enhanced SOH features, $F'_{SOH}$, are first projected along the model dimension through a linear transformation. Subsequently, average pooling is performed over the sequence length $L$ to aggregate temporal information across the entire time window. Finally, the pooled representation is passed through an estimator to produce a single scalar SOH prediction.

The SOH estimation module achieves a holistic health assessment through sequence compression and linear regression, formulated as follows:

$$z_{SOH} = \frac{1}{L}\sum_{t=1}^{L}\left(W_{soh}\widetilde{H}_{SOH,t} + b_{soh}\right) \tag{24}$$

$$\hat{y}_{SOH} = \sigma(W_{soh} tanh(z_{SOH}) + b_{soh}) \tag{25}$$

where $\sigma(\cdot)$ denotes the sigmoid activation function.

This sequence averaging mechanism essentially functions as a temporal aggregation of information regarding the battery's operation over the full cycle, enabling the model to reflect degradation trends from a global perspective.

### 3.5. Cross-domain transfer and supervised fine-tuning strategy

This study employs a supervised fine-tuning strategy within a transfer learning paradigm to achieve effective model adaptation from the source domain to the target domain. The core premise of this strategy is to utilize a model pre-trained on a source-domain dataset as the foundational baseline and subsequently refine it using labeled target-domain data. This process enables the model to adjust to the specific statistical characteristics and operating conditions of the target domain.

To ensure efficient target-domain adaptation while preserving the knowledge learned from the source domain, we adopt a selective parameter fine-tuning strategy. During the fine-tuning phase, different components of the model are treated distinctly as follows.

**Frozen Parameters:** The parameters of the Shared Feature Extractor and the two Estimation Heads are frozen, ensuring their weights remain invariant during the fine-tuning process. This operation aims to preserve the general feature extraction and output mapping capabilities acquired from the source domain, which are hypothesized to be domain-invariant.

**Trainable Parameters:** The parameters of the two Task-Specific Adapters and the Task Interaction Attention module are set to a trainable state. These modules are considered domain-specific; fine-tuning them allows the model to efficiently adapt to the subtle nuances of task features in the target domain and the unique correlation between SOC and SOH, without altering the core model architecture.

During the supervised fine-tuning phase in the target domain, we utilize the target domain training dataset $D_{train}^{T}$, to optimize the model's trainable parameters. For a given data batch $\{X^T, Y_{soc}, Y_{soh}\}$ sampled from $D_{train}^{T}$, the forward propagation of the model yields the estimated sequences $\hat{Y}_{soc}^{T}$ and $\hat{Y}_{soh}^{T}$.

Root Mean Square Error (RMSE), which is calculated as Equation 26, is employed as the loss function to supervise the training of both tasks.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{26}$$

For SOH estimation, $n$ equals the batch size, $y_i$ represents the ground truth SOH value, and $\hat{y}_i$ denotes the predicted SOH value.

For SOC estimation, given its nature as a sequence prediction task, the error must be computed across all $n = B \times L$ time steps, where $B$ is the batch size and $L$ is the sequence length, $y_i$ and $\hat{y}_i$ correspond to the ground truth and predicted SOC values at the respective time step.

The overall optimization objective of the model is defined as the weighted sum of the losses from the two tasks:

$$\min_{\theta} \mathcal{L}_{total} = \alpha \cdot \text{RMSE}(\hat{Y}_{soc}, Y_{soc}) + \beta \cdot \text{RMSE}(\hat{Y}_{soh}, Y_{soh}) \tag{27}$$

where, θ represents the set of all trainable parameters within the model, and $\alpha$ and $\beta$ are hyperparameters designed to balance the relative importance of the two tasks during the fine-tuning process. The model fine-tuning is performed using the Stochastic Gradient Descent (SGD) optimizer with momentum. The whole process is shown in Algorithm **1**.

Algorithm **1** Training Procedure of SFJE

**Require** : Training target dataset $\mathcal{D}^t = \{(x_i, y_i^{soc}, y_i^{soh})\}_{i=1}^{n_t}$, trained source model $M_s$
**Ensure**: Trained target model $M_t$
1: Initialize target model $M_t$ with parameters from $M_s$
2: Freeze shared feature extractor and estimators in $M_t$
3: For $epoch = 1$ to $max_epoch$ do
4: For each mini-batch $(X, Y_{soc}, Y_{soh})$ sampled from $\mathcal{D}^t$ do
5: **Forward Propagation**:
6: Obtain predictions: $\hat{Y}_{soc}, \hat{Y}_{soh} = M_t(X)$
7: **Loss Calculation**:
8: Compute task-specific losses:
9: $L_{soc} = \text{RMSE}(Y_{soc}, \hat{Y}_{soc})$
10: $L_{soh} = \text{RMSE}(Y_{soh}, \hat{Y}_{soh})$
11: Compute joint weighted loss:
12: $L_{total} = \alpha \cdot L_{soc} + \beta \cdot L_{soh}$
13: **Back propagation**:
14: Update trainable parameters of $M_t$ by minimizing $L_{total}$
15: end for
16: end for
17: return Trained target model $M_t$

## 4. Experiments and results analysis

This section aims to conduct a comprehensive and in-depth evaluation of the performance of the proposed cross-domain

SOC-SOH joint estimation framework through a series of systematic experiments. First, the evaluation metrics and the overall experimental design employed in this study are detailed. Subsequently, the dual superiority of the proposed model--in terms of both estimation accuracy and domain adaptability--is validated through comparative analysis against state-of-the-art methods across source domain performance and cross-domain transfer scenarios. Finally, ablation studies are conducted to investigate the specific contributions of the core modules, followed by an uncertainty analysis to further verify the model's robustness and the reliability of the prediction results.

### 4.1. Experimental design

To systematically assess the model's cross-condition generalization capability and transfer adaptability, we adhered to a rigorous transfer learning experimental paradigm. We selected four representative charging protocols with significant distinctness from the dataset to serve as the experimental subjects, as detailed in Table 1.

Table 1. The selected charging protocols used for battery dataset construction.

| ID | Charging Protocol |
|---|---|
| D1 | 5.4C(60%)--3C |
| D2 | 7C(40%)--3C |
| D3 | 8C(35%)--3.6C |
| D4 | 6C(40%)--3C |

To simulate the practical deployment scenario where a model transitions from known operating conditions to unknown, novel conditions, we designed a series of closed-loop cross-domain transfer tasks. As illustrated in Table 2, the four aforementioned charging protocols are combined in pairs, alternating roles between the Source Domain and the Target Domain.

Table 2. Matrix of the cross-domain transfer tasks.

| Target / Source | D1 | D2 | D3 | D4 |
|---|---|---|---|---|
| D1 | - | √ | √ | √ |
| D2 | √ | - | √ | √ |
| D3 | √ | √ | - | √ |
| D4 | √ | √ | √ | - |

Specifically, each transfer task comprises two distinct stages: pre-training (on the Source Domain) and supervised fine-tuning (on the Target Domain). This rigorous matrix design, encompassing $4 \times 3 = 12$ distinct transfer tasks, enables a comprehensive and unbiased validation of the proposed algorithm's superiority and stability in handling variations in operating conditions.

To comprehensively quantify the estimation accuracy of the model, we employ two widely established evaluation metrics:

**RMSE**: As a metric highly sensitive to large prediction errors, RMSE serves as the primary indicator for assessing the model's overall performance, and is calculated as shown in Equation **26**.

**Mean Absolute Error (MAE)**: Provide a more intuitive reflection of the average magnitude of prediction errors and calculated as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{28}$$

where $N$ denotes the total number of samples, $y_i$ represents the ground truth value, and $\hat{y}_i$ is the predicted value.

### 4.2. Hyperparameter settings

To ensure the fairness and reproducibility of the experiments, all baseline models underwent rigorous hyperparameter tuning. The key hyperparameter settings for the proposed model are detailed in Table 3. Unless otherwise specified in the subsequent parameter sensitivity analysis section, this configuration is consistently applied across all experiments.

Table 3. Detailed hyperparameter settings for the training and architecture of SFJE.

| Parameter | Value |
|---|---|
| Learning Rate | 0.0006 |
| Batch Size | 16 |
| Training Epochs | 200 |
| Embedding Dimension $d_{model}$ | 256 |
| Adapter Dimension $d_m$ | 50 |
| Number of Attention Heads $n_{heads}$ | 8 |
| Number of Encoder Layers $e_{layers}$ | 2 |
| α | 1 |
| β | 3 |

### 4.3. Computational cost and parameter efficiency

A core advantage of the proposed framework is its parameter efficiency during the transfer process. To substantiate this, the computational overhead and model size were explicitly quantified, as summarized in Table 4.

As reported in Table 4, the total parameter count of the multi-task framework is approximately 2.59 million. By freezing the heavy backbone network during transfer learning, only the specific task-related modules (including the lightweight Adapters and the TIA module) are fine-tuned. The total trainable parameters are restricted to 315,392, representing only about 12.16% of the overall model size. This highly localized fine-tuning strategy drastically mitigates the computational cost and memory footprint required for target-domain adaptation compared to full fine-tuning. Furthermore, the framework achieves an average inference latency of 2.94 ms. This lightweight characteristic and microsecond-level response time confirm the framework's viability for real-time deployment in on-board BMS, where computational resources are strictly constrained.

Table 4. Quantification of model parameters and inference time.

| Metric | Value |
|---|---|
| Total Model Parameters | 2593713 |
| Total Trainable Parameters | 315392 |
| Adapter Parameters | 51200 |
| Ratio of Trainable Parameters | 12.16% |
| Average Inference Latency | 2.94 ms |

### 4.4. Experimental results and analysis

#### 4.4.1. Performance comparison in the source domain

To evaluate the intrinsic effectiveness of the proposed SFJE backbone, we conducted a comparative performance analysis against various baselines (LSTM [32,36], CNN [37-38], GRU [39-40], and Transformer [41-42]) across four distinct source domain operating conditions (D1-D4).

Table 5 details the RMSE results from these experiments. The following insights can be drawn from the analysis:

1. **Overall Performance**: Our model achieves the lowest average RMSE for both SOC (0.2355%) and SOH (0.3093%) tasks. Unlike single-stream baselines, which exhibit significant volatility across domains, our method demonstrates superior comprehensive performance and generalization capability.

2. **Complex Conditions**: Under conditions D1 and D2, our model significantly outperforms all baselines. Given the drastic temporal heterogeneity visualized in Fig. 8, this superiority stems from the SFJE framework's synergy of local temporal modeling (LSTM) and global dependency capturing (Encoder), enabling it to effectively handle complex dynamic characteristics.

3. **Local-Dominant Conditions**: Conversely, simpler baselines occasionally yield optimal results in D3 and D4. As suggested by Fig. 8, these scenarios likely rely on simple temporal patterns that align with lighter inductive biases. Nevertheless, our model remains highly competitive in these cases, avoiding the severe performance drops observed in other baselines.

Synthesizing these results, our rationale for selecting this hybrid architecture is not to surpass simpler models in every specific condition, but to establish a robust general framework. It achieves a higher performance ceiling in complex scenarios and stable overall performance across diverse situations. This confirms that integrating multiple inductive biases is essential for handling multi-scale dynamics, serving as an ideal foundation for cross-domain transfer research.

Table 5. Quantitative performance evaluation of the SFJE and baseline models on the source domain (RMSE, %).

| Source | LSTM | | CNN | | GRU | | Transformer | | LSTM+Transformer(Ours) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | SOC | SOH | SOC | SOH | SOC | SOH | SOC | SOH | SOC | SOH |
| D1 | 0.2982 | 0.2437 | 0.3535 | 1.1532 | 0.3355 | 0.4879 | 0.2427 | 0.3244 | **0.1625** | **0.2090** |
| D2 | 0.3738 | 0.3802 | 0.3821 | 0.6322 | 0.4320 | 0.2847 | 0.4161 | 0.5216 | **0.2066** | **0.2820** |
| D3 | 0.3573 | **0.3336** | 0.4025 | 0.4541 | 0.3531 | 0.3857 | 0.3983 | 0.5621 | **0.3457** | 0.4107 |
| D4 | **0.2182** | 0.2994 | 0.2787 | 1.1988 | 0.2334 | 0.3017 | 0.2582 | **0.2973** | 0.2272 | 0.3353 |
| AVG | 0.3119 | 0.3142 | 0.3542 | 0.8596 | 0.3385 | 0.3650 | 0.3288 | 0.4263 | **0.2355** | **0.3093** |

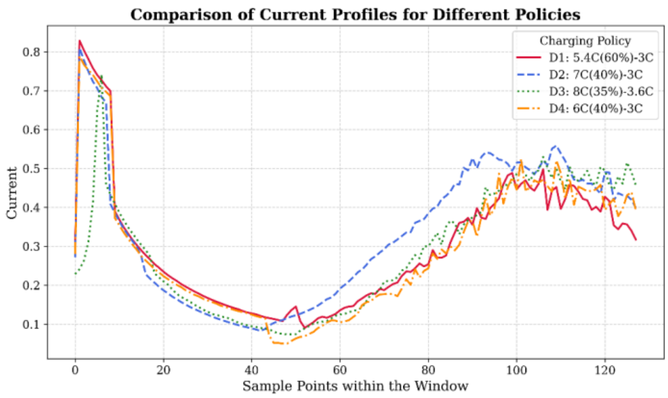



Figure 8. Detailed comparison of charging current profiles generated by different charging policies.

#### 4.4.2. Cross-domain transfer performance comparison

To comprehensively evaluate the effectiveness of the SFJE framework for cross-domain transfer tasks, we systematically compared it against several recent state-of-the-art (SOTA) approaches published in top-tier journals, which represent mainstream domain adaptation algorithms. First, we employed Source-Only as a non-adaptive baseline. Second, we selected DANN [43] and MMD [35] to represent classic Unsupervised Domain Adaptation (UDA) methods. Finally, we included SHOT [44], which is a state-of-the-art Source-Free UDA method. Table 6 shows comparison of these methods' characteristics.

Table 7 and Table 8 illustrate the RMSE and MAE results for all methods, which are evaluated on the complete set of 12 cross-domain transfer tasks in Table 2, covering all possible source-target combinations to fully verify the adaptability across different operating conditions.

**Source-Only**: This baseline applies the source-trained model directly to the target domain. It serves as a lower bound to quantify the initial domain shift.

**DANN**: This adversarial method employs a domain discriminator to align feature distributions. It enables the model to learn domain-invariant representations by confusing the discriminator.

**MMD**: This statistical method aligns global distributions by minimizing the distance between the sample means of the source and target domains in a kernel space.

**SHOT**: We adapted this source-free method for regression tasks. It fine-tunes the model using pseudo-labels and consistency regularization derived from feature space clustering.

Table 6. Comprehensive comparison of methodological characteristics between SFJE and existing approaches.

| Feature | Source-Only | DANN | SHOT | MMD | SFJE(Ours) |
|---|---|---|---|---|---|
| Target Labels | √ | × | × | √ | √ |
| Source Data | √ | √ | × | √ | × |

Through an in-depth horizontal and vertical comparison of the performance metrics across these 12 transfer tasks, we derive the following core conclusions focusing on three dimensions: comprehensive model accuracy, the necessity of transfer learning, and task balance.

First of all, the SFJE method achieves the best performance in terms of both accuracy and balance. As shown in Table 7 and Table 8, SFJE yields the lowest average errors among all methods, with an RMSE of 0.2818% for SOC, 0.4394% for SOH, and an MAE of 0.2105% for SOC, 0.4277% for SOH. These results validate the effectiveness of our Supervised Selective Fine-tuning Strategy. They demonstrate that SFJE maintains robust estimation capabilities even when facing significant domain shifts. Fig. 9 displays the prediction curves for the D1 to D2 task, visually confirming the superiority of our method.

Secondly, the results further substantiate the importance of domain adaptation and the advantages of SFJE methods. The high average errors observed in the Source-Only baseline clearly reveal the existence of significant domain shifts within cross-condition battery data. Although traditional unsupervised adaptation methods, such as DANN, MMD and SHOT, mitigated this issue to a certain extent, they often exhibited a distinct performance imbalance between SOC and SOH tasks. In contrast, SFJE achieved simultaneous improvements in both metrics, yielding average errors significantly lower than all unsupervised baselines. This demonstrates that when labeled samples are available, supervised fine-tuning offers substantial advantages for feature alignment and synergistic task optimization.

Crucially, SFJE achieves a highly balanced and robust performance across joint estimation tasks. In multi-task coupled scenarios, traditional transfer methods often struggle with performance biases arising from task conflicts. For instance, DANN and MMD show SOC degradation in specific transfer scenarios, while SHOT tends to improve SOH estimation at the cost of SOC accuracy.

In contrast, SFJE simultaneously improves the estimation efficacy of both tasks across the majority of transfer combinations, exhibiting exceptional task coordination capabilities. Although it may not achieve the absolute optimal performance in isolated scenarios, such as RMSE and MAE for SOC in D3 to D2 and D4 to D1, SFJE maintains a leading position among all methods with minimal error variance and slight performance fluctuations, as illustrated in Fig. 10.

Table 7. Estimation performance of the proposed and compared methods on the target domains (RMSE, %).

| Source | Target | Source-Only | | DANN | | MMD | | SHOT | | SFJE(Ours) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | SOC | SOH | SOC | SOH | SOC | SOH | SOC | SOH | SOC | SOH |
| | D2 | 0.4459 | 0.3391 | 0.3819 | 0.3584 | 0.3815 | 0.3562 | 1.4506 | 0.4105 | **0.2989** | **0.3259** |
| | D3 | 1.3869 | 8.0153 | 1.2662 | 7.3117 | 1.2664 | 7.3118 | 1.3779 | 0.6787 | **0.3958** | **0.4410** |
| D1 | D4 | 0.2419 | 0.5417 | 0.2110 | 0.4403 | 0.2107 | 0.4403 | 1.5881 | 0.7415 | **0.1579** | **0.4256** |
| | D1 | 0.2715 | 0.4485 | 0.2741 | 0.4745 | 0.2741 | 0.4763 | 1.6015 | 0.5835 | **0.2101** | **0.4336** |
| D2 | D2 | 1.2948 | 5.5438 | 1.2337 | 4.9079 | 1.2338 | 4.9110 | 1.6079 | 0.8370 | **0.3615** | **0.5641** |
| | D3 | 0.3807 | 0.7856 | 0.3787 | 0.8042 | 0.3791 | 0.8052 | 1.6321 | 0.6626 | **0.2115** | **0.3881** |
| | D1 | 0.3277 | 0.9025 | 0.3487 | 0.9684 | 0.3489 | 0.9689 | 1.8535 | 0.8919 | **0.2673** | **06216** |
| D3 | D2 | 0.2784 | 0.4366 | **0.2621** | 0.4795 | 0.2622 | 0.4812 | 1.7264 | 0.5214 | 0.3038 | **0.2726** |
| | D4 | 0.2977 | 1.1315 | 0.3167 | 1.2355 | 0.3168 | 1.2342 | 1.9008 | 1.0206 | **0.2721** | **0.6043** |
| | D1 | 0.3860 | 0.2324 | 0.340 | 1.3427 | 0.3441 | 1.3444 | 1.8318 | 0.5660 | **0.2560** | **0.3157** |
| D4 | D2 | 0.4274 | 0.5649 | 0.3645 | 0.4949 | 0.3644 | 0.4958 | 2.1796 | 0.5358 | **0.2932** | **0.4819** |
| | D3 | 1.1239 | 3.0154 | 0.9471 | 2.6475 | 0.9476 | 2.6515 | 2.1295 | 0.7238 | **0.3536** | **0.3985** |
| **AVG** | | 0.5719 | 1.9131 | 0.5274 | 1.7889 | 0.5275 | 1.7897 | 1.7400 | 0.6811 | **0.2818** | **0.4394** |

Table 8. Estimation performance of the proposed and compared methods on the target domains (MAE, %).

| Source | Target | Source-Only | | DANN | | MMD | | SHOT | | SFJE(Ours) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | SOC | SOH | SOC | SOH | SOC | SOH | SOC | SOH | SOC | SOH |
| | D2 | 0.3609 | 0.3257 | 0.3058 | 0.3474 | 0.3054 | 0.3453 | 0.8816 | 0.4019 | **0.2237** | **0.3197** |
| D1 | D3 | 0.8362 | 8.0061 | 0.7607 | 7.3030 | 0.7609 | 7.3030 | 0.8701 | 0.6687 | **0.2730** | **0.4320** |
| | D4 | 0.1740 | 0.5304 | 0.1531 | 0.4322 | 0.1529 | 0.4323 | 0.9488 | 0.7326 | **0.1144** | **0.4171** |
| | D1 | 0.2124 | 0.4350 | 0.2103 | 0.4631 | 0.2103 | 0.4640 | 0.9916 | 0.5678 | **0.1546** | **0.4219** |
| D2 | D2 | 0.9527 | 5.5308 | 0.9046 | 4.8960 | 0.9044 | 4.8990 | 1.0208 | 0.8218 | **0.2559** | **0.5542** |
| | D3 | 0.2698 | 0.7726 | 0.2589 | 0.7925 | 0.2592 | 0.7936 | 0.9997 | 0.6528 | **0.1562** | **0.3683** |
| | D1 | 0.2278 | 0.8821 | 0.2422 | 0.9438 | 0.2424 | 0.9446 | 1.2015 | 0.8546 | **0.2068** | **0.6129** |
| D3 | D2 | 0.2123 | 0.4305 | **0.2015** | 0.4710 | 0.2017 | 0.4728 | 1.0720 | 0.5088 | 0.2458 | **0.2572** |
| | D4 | 0.2176 | 1.1259 | 0.2317 | 1.2298 | 0.2317 | 1.2285 | 1.2234 | 1.0096 | **0.2126** | **0.5886** |
| | D1 | 0.2586 | 1.1814 | **0.2314** | 1.2915 | **0.2314** | 1.2932 | 1.0489 | 0.5437 | 0.2565 | **0.3852** |
| D4 | D2 | 0.3533 | 0.5573 | 0.3001 | 0.4862 | 0.2999 | 0.4871 | 1.3084 | 0.5232 | **0.2319** | **0.4752** |
| | D3 | 0.8819 | 2.9978 | 0.7307 | 2.6330 | 0.7312 | 2.6371 | 1.3303 | 0.7058 | **0.1943** | **0.3006** |
| **AVG** | | 0.4131 | 1.8980 | 0.3776 | 1.7741 | 0.3776 | 1.7750 | 1.0748 | 0.6659 | **0.2105** | **0.4277** |

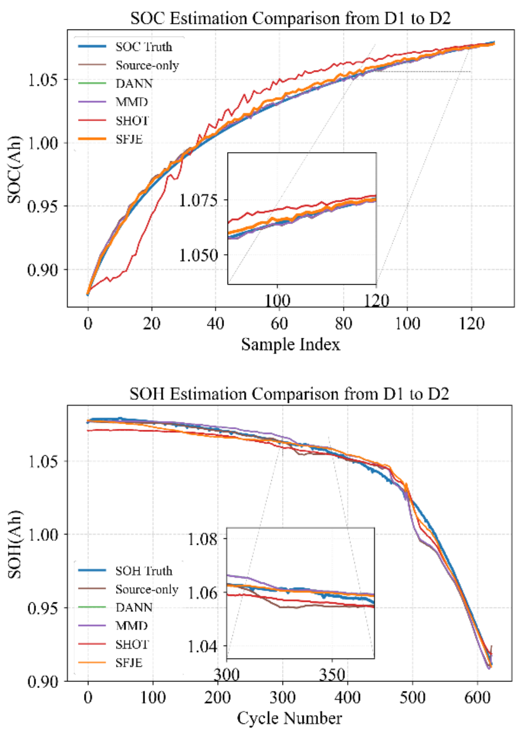



Figure 9. Comparison of the joint SOC and SOH prediction curves across different models of Transfer Task D1 to D2.

This phenomenon highlights the structural resilience of the SFJE framework in complex transfer scenarios. Rather than prioritizing one task at the expense of the other, our approach preserves the integrity of the shared feature representation. This implies that the model successfully harmonizes the competing demands of joint estimation, ensuring consistent and reliable performance despite substantial distributional divergence in the target domain.

Mechanistically, this enhanced performance is driven by the synergy between the TIA mechanism and the Fine-tuning strategy. The mechanism explicitly captures intrinsic correlations between SOC and SOH at the feature level, while the fine-tuning strategy facilitates a precise calibration of the shared feature space. Together, these components allow the model to simultaneously refine its representations for both tasks during transfer, thereby achieving an optimal equilibrium for joint estimation.

In conclusion, SFJE not only achieves high average accuracy but, more critically, maintains exceptional robustness and task equilibrium in multi-task transfer scenarios. The framework effectively addresses the challenges of negative transfer and task bias often observed in traditional cross-domain methods. These results validate SFJE as a precise and unified framework, capable of providing stable estimation even under complex, conflicting transfer conditions.

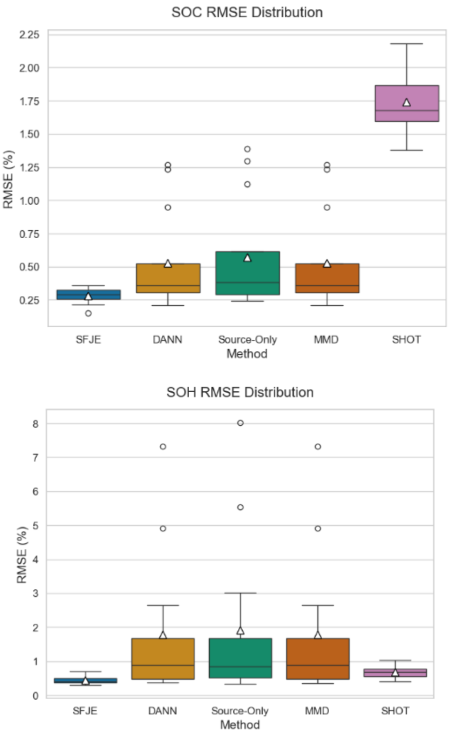



Figure 10. Statistical distribution of the average RMSE for SOC and SOH estimation across all transfer tasks.

## 4.5. Ablation studies

To thoroughly investigate the specific contributions of the individual core components within the proposed framework, we conducted a comprehensive series of ablation studies. Taking the complete architecture as the baseline, we constructed degraded variants by selectively removing or bypassing specific key modules. All variants were evaluated under identical experimental settings.

We devised the following model variants, with the experimental design detailed in Table 9.

**onlyShared**: Excludes both the task-specific Adapter modules and the TIA module, retaining only the shared feature extractor

**noAdapter**: Removes the task-specific Adapter modules; the shared features are fed directly into the TIA module.

**noTIA**: Removes the TIA module; the features processed by the Adapters are directly forwarded to their respective estimators.

**selfAtten**: Replaces the proposed TIA with a standard self-attention fusion mechanism to serve as an attention variant baseline.

Table 9. Configuration details of the ablation model variants.

| Module | onlyShared | noAdapter | noTIA | **selfAtten** | SFJE |
|---|---|---|---|---|---|
| Adapter | × | × | √ | √ | √ |
| TIA module | × | √ | × | × | √ |
| Self-Attention | × | × | × | √ | × |

Using the transfer task with D4 as the source domain and D1, D2, and D3 as target domains, we evaluated the performance of these variants. The results are detailed in Table 10, and Fig. 11 illustrates the comparative performance across different variants. The following observations can be drawn from the results:

**SFJE Achieves Optimal Performance**: The SFJE model delivers the lowest average RMSE for both SOC (0.3009%) and SOH (0.3987%). Unlike ablated variants, it maintains robust precision across all target domains. This validates that integrating task interaction with selective fine-tuning is essential for establishing a robust architecture capable of handling complex domain shifts.

**TIA mechanism is Central to Balance**: The interaction module is pivotal for equilibrium. Removing it (noTIA) provides a quantitative measure of task interference, causing a severe performance skew: SOH error rises from 0.3987% to 0.5431% while SOC degrades from 0.3009% to 0.3179%. This confirms the mechanism's ability to suppress these conflicts and the seesaw effect, ensuring that one task is not optimized at the expense of the other.

**TIA Outperforms Standard Self-Attention**: Compared to SFJE, the selfAtten variant suffers a severe SOC accuracy drop to 0.4547%. This confirms that standard self-attention causes "over-fusion" by mixing features indiscriminately, whereas our TIA avoids this by selectively extracting complementary context while preserving task-specific identities.

**Adapters are Crucial for Single-Task Precision**: Adapters are key to specific feature alignment. The noAdapter variant suffers a simultaneous accuracy drop in both tasks, with average SOH error rising to 0.5735%. This indicates that shared features alone are insufficient, and Adapters are critical for fine-tuning representations to match distinct target domain distributions.

Table 10. Quantitative ablation study results for SOC and SOH estimation across target domains (RMSE, %).

| Source | Target | onlyShared | | noAdapter | | noTIA | | selfAtten | | SFJE(Ours) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | SOC | SOH | SOC | SOH | SOC | SOH | SOC | SOH | SOC | SOH |
| D4 | D1 | 0.3280 | 0.4939 | 0.2704 | 0.4257 | 0.2602 | 0.4616 | 0.4950 | 0.3950 | **0.2560** | **0.3157** |
| | D2 | 0.3081 | 0.6659 | 0.2942 | 0.6349 | 0.3068 | 0.6615 | 0.3797 | 0.5243 | **0.2932** | **0.4819** |
| | D3 | 0.4627 | 1.0506 | 0.4037 | 0.6599 | 0.3867 | 0.5063 | 0.4893 | 0.4123 | **0.3536** | **0.3985** |
| **AVG** | | 0.3663 | 0.7368 | 0.3228 | 0.5735 | 0.3179 | 0.5431 | 0.4547 | 0.4439 | **0.3009** | **0.3987** |

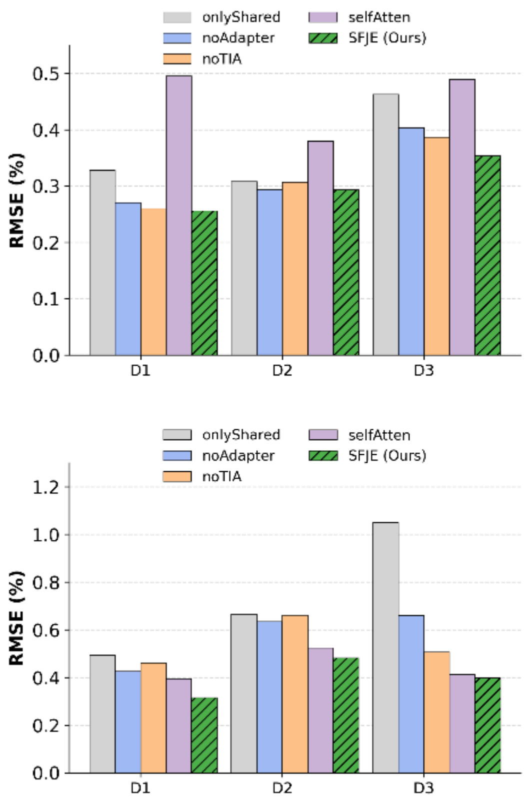



Figure 11. Bar chart comparison of SOC and SOH estimation accuracy in ablation studies.

### 4.6. Uncertainty analysis

To assess the reliability of the model's predictions, we incorporated an uncertainty analysis framework [45]. By enabling the Dropout layers during the inference phase a technique known as Monte Carlo Dropout [46] we perform multiple stochastic forward passes for a given input sample to generate a predictive distribution. The mean of this distribution serves as the final prediction, while its standard deviation functions as a quantitative metric for model uncertainty.

Fig. 12 illustrates the test results and corresponding 95% confidence intervals for the D1 to D2 transfer scenario. The data confirms that the model maintains high stability and confidence for both short-term SOC and long-term SOH predictions. For the SOC sequence, the mean prediction closely tracks the ground truth with consistently narrow uncertainty bounds. Likewise, for the SOH trajectory, the model accurately captures the degradation trend with tight confidence intervals across most cycles. This indicates that the model has effectively learned the underlying long-term aging mechanisms.

In summary, the model simultaneously achieves high precision and confidence at both microscopic and macroscopic scales. The consistently low uncertainty highlights the framework's robustness and generalization capabilities, establishing it as a reliable tool for battery state estimation.

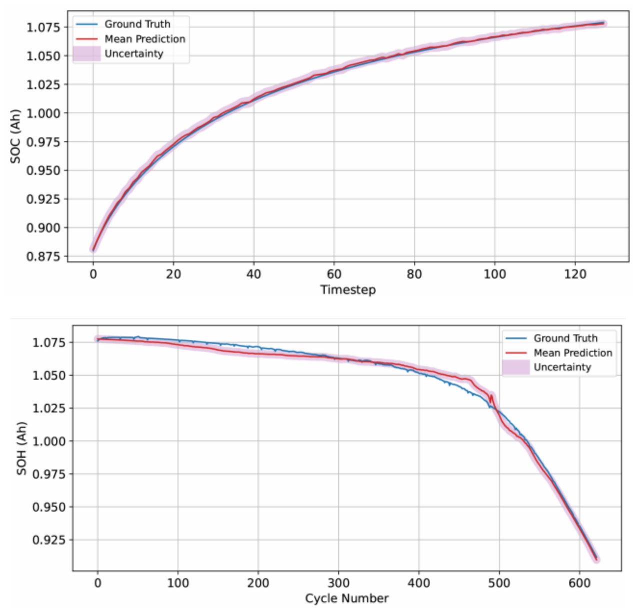



Figure 12. Visualization of the uncertainty quantification for SOC and SOH joint predictions on Transfer Task D1 to D2.

### 4.7. Parameter sensitivity analysis

To ensure the robustness, efficiency, and reliability of the proposed multi-task framework, comprehensive sensitivity analyses were conducted to determine the optimal settings for three critical hyperparameters: the bottleneck dimension ($d_m$) of the Task-Specific Adapter, the dropout rate ($p$), and the number of Monte Carlo forward passes ($T$).

#### 4.7.1. Bottleneck dimension ($d_m$) of the task-specific adapter

The Adapter aims to balance parameter efficiency with feature representation capability. As shown in Fig. 13, the joint estimation errors exhibited a distinct U-shaped trajectory when evaluated across d_m∈{10, 30, 50, 80, 100}.

Narrower dimensions ($d_m$≤30) caused an excessive information bottleneck, leading to underfitting and high estimation errors. Conversely, wider dimensions ($d_m$≥80) introduced parameter redundancy, which disrupted the multi-task learning equilibrium and increased the risk of overfitting. Therefore, $d_m$=50 was selected as the optimal trade-off, maximizing feature representation capability while preserving a lightweight structure.

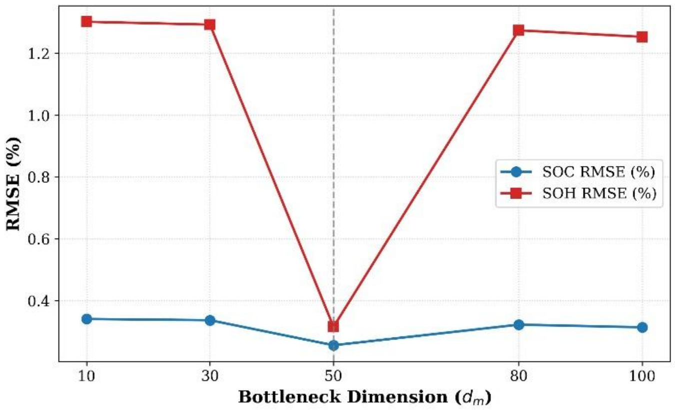
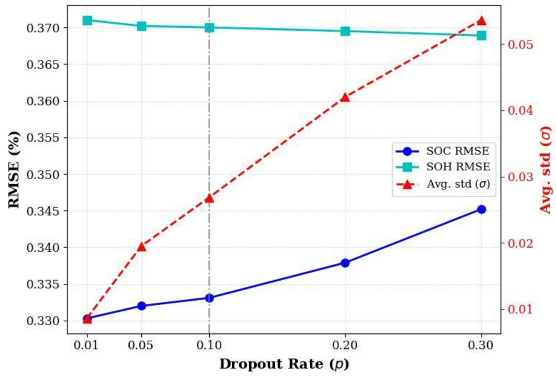



Figure 13. The sensitivity of $d_m$.

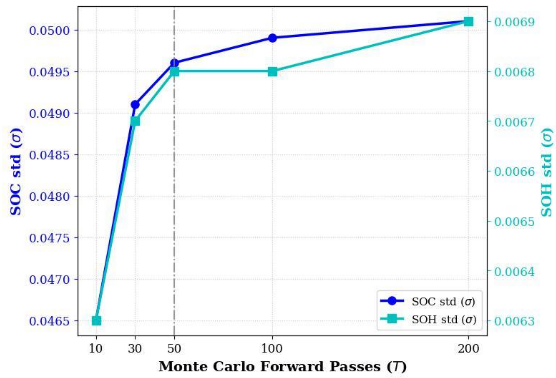
### 4.7.2. Hyperparameters of Monte Carlo Dropout

1) Dropout Rate ($\boldsymbol{p}$): The selection of $p$ involves a critical trade-off between predictive accuracy (RMSE) and uncertainty calibration (standard deviation, $\sigma$). As illustrated in Fig. 14(a), the empirical results tested across $p \in \{0.01, 0.05, 0.1, 0.2, 0.3\}$ showed that $\sigma$ strictly increases with $p$. Setting $p$ too low ($p \leq 0.05$) yielded an overconfident model with excessively narrow uncertainty bounds. Conversely, excessively high dropout rates ($p \geq 0.2$) severely degraded the SOC estimation accuracy due to a loss of network representation capacity. Consequently, $p = 0.1$ was established as the optimal multi-task sweet spot, preserving robust joint accuracy while maintaining well-calibrated epistemic uncertainty.

2) **Monte Carlo Forward Passes *(T)***: The hyperparameter T dictates the balance between statistical convergence and computational overhead. As demonstrated in Fig. 14(b), an evaluation across T ∈{10,30,50,100,200} revealed a clear convergence knee point at T = 50. Below this threshold, the predicted standard deviations for both SOC and SOH exhibited instability. Beyond T = 50, further increases yielded mathematically negligible variations in uncertainty metrics while linearly inflating the inference time. To ensure reliable uncertainty quantification without violating the real-time execution constraints of practical BMS, T = 50 was adopted as the optimal operational threshold.




(a) Dropout rate




(b) MC passes

Figure 14. Sensitivity analysis of the dropout rate and MC passes.

## 5. Conclusion

This paper introduces the SFJE framework, a multi-task learning approach designed for the joint estimation of SOC and SOH under cross-condition scenarios. By incorporating a shared feature extractor, task-specific adapters, and a TIA module, the framework enables effective information fusion while achieving a balanced performance across tasks. Ablation studies validate the contribution of each component, with the TIA mechanism playing a particularly critical role in mitigating task conflicts and enhancing overall accuracy. In addition, robustness and uncertainty analyses show that SFJE delivers stable and reliable predictions under diverse operating conditions, demonstrating its strong generalization capability in complex multi-task settings.

Despite the improvements in accuracy and robustness, the proposed framework still has limitations in terms of computational complexity and the breadth of validation. Future work will therefore proceed in three main directions. First, we will pursue model lightweighting by incorporating knowledge distillation and pruning techniques to improve deployment efficiency on embedded BMS. Second, to address the current reliance on a single public dataset, we are constructing a proprietary dataset to validate the framework's effectiveness across diverse battery chemical compositions, and plan to extend it to support the joint estimation of additional critical states, such as RUL. Finally, leveraging predictive uncertainty to enable risk-aware, active battery management constitutes an important direction for our subsequent research.

**Acknowledgment**

This research was funded by The Social Public Welfare and Basic Research Project of Zhongshan City: 2025B4008.

**References**

1. Li D, Deng J, Zhang Z, Wang Z, Zhou L, Liu P. Battery Safety Risk Assessment in Real-World Electric Vehicles Based on Abnormal Internal Resistance Using Proposed Robust Estimation Method and Hybrid Neural Networks. *IEEE Transactions on Power Electronics* 2023; 38(6): 7661-7673. https://doi.org/10.1109/tpel.2023.3241938.
2. Wen J, Zhao D, Zhang C. An overview of electricity powered vehicles: Lithium-ion battery energy storage density and energy conversion efficiency. *Renewable Energy* 2020; 162: 1629-1648. https://doi.org/10.1016/j.renene.2020.09.055.
3. Xia B, Qin Z, Fu H. Rapid estimation of battery state of health using partial electrochemical impedance spectra and interpretable machine learning. *Journal of Power Sources* 2024; 603: 234413. https://doi.org/10.1016/j.jpowsour.2024.234413.
4. Tao Z, Zhao Z, Wang C, Huang L, Jie H, Li H, Hao Q, Zhou Y, See KY. State of charge estimation of lithium batteries: Review for equivalent circuit model methods. *Measurement* 2024; 236: 115148. https://doi.org/10.1016/j.measurement.2024.115148.
5. Bavand A, Khajehoddin SA, Ardakani M, Tabesh A. Online Estimations of Li-Ion Battery SOC and SOH Applicable to Partial Charge/Discharge. *IEEE Transactions on Transportation Electrification* 2022; 8(3): 3673-3685. https://doi.org/10.1109/tte.2022.3162164.
6. Lyu Y, Wen Z, Chen A. A novel transfer learning approach based on deep degradation feature adaptive alignment for remaining useful life prediction with multi-condition data. *Journal of Intelligent Manufacturing* 2025; 36(1): 619-637. https://doi.org/10.1007/s10845-023-02264-4.
7. Luo X, Bu W, Liang H, Zheng M. Convolutional Neural Network - Gated Recurrent Unit combined with Error Correction for Lithium Battery State of Health Estimation. *Eksploatacja i Niezawodność – Maintenance and Reliability* 2025; 27(4): 202184. https://doi.org/10.17531/ein/202184.
8. Zhang Q, Wan G, Li C, Li J, Liu X, Li M. State of charge estimation for Li-ion battery during dynamic driving process based on dual-channel deep learning methods and conditional judgement. *Energy* 2024; 294: 130948. https://doi.org/10.1016/j.energy.2024.130948.
9. Li C, Han X, Zhang Q, Li M, Rao Z, Liao W, Liu X, Liu X, Li G. State-of-health and remaining-useful-life estimations of lithium-ion battery based on temporal convolutional network-long short-term memory. *Journal of Energy Storage* 2023; 74: 109498. https://doi.org/10.1016/j.est.2023.109498.
10. Yang K, Tang Y, Zhang S, Zhang Z. A deep learning approach to state of charge estimation of lithium-ion batteries based on dual-stage attention mechanism. *Energy* 2022; 244: 123233. https://doi.org/10.1016/j.energy.2022.123233.
11. Gu X, See KW, Li P, Shan K, Wang Y, Zhao L, Lim KC, Zhang N. A novel state-of-health estimation for the lithium-ion battery using a convolutional neural network and transformer model. *Energy* 2023; 262: 125501. https://doi.org/10.1016/j.energy.2022.125501.
12. Zhou Z, Zhang C, Chen S, Zhang Y, Wang L. Feature-enhanced ensemble learning for accurate capacity estimation of lithium-ion batteries using partial discharging segments in initial stage based on second-order voltage derivatives. *Green Energy and Intelligent Transportation* 2026; 5(3): 100388. https://doi.org/10.1016/j.geits.2025.100388.
13. Chen L, Xie S, Lopes AM, Li H, Bao X, Zhang C, Li P. A new SOH estimation method for Lithium-ion batteries based on model-data-fusion. *Energy* 2024; 286: 129597. https://doi.org/10.1016/j.energy.2023.129597.
14. Bao X, Chen L, Lopes AM, Li X, Xie S, Li P, Chen YQ. Hybrid deep neural network with dimension attention for state-of-health estimation of Lithium-ion Batteries. *Energy* 2023; 278: 127734. https://doi.org/10.1016/j.energy.2023.127734.

15. Huang H, Bian C, Wu M, An D, Yang S. A novel integrated SOC–SOH estimation framework for whole-life-cycle lithium-ion batteries. *Energy* 2024; 288: 129801. https://doi.org/10.1016/j.energy.2023.129801.
16. Li S, Jiang Z, Zhu Z, Jiang W, Ma Y, Sang X, Yang S. A framework of joint SOC and SOH estimation for lithium-ion batteries: Using BiLSTM as a battery model. *Journal of Power Sources* 2025; 635: 236342. https://doi.org/10.1016/j.jpowsour.2025.236342.
17. Lyu Y, Shen Z, Zhou N, Wen Z, Chen C. A feature separation transfer network with contrastive metric for remaining useful life prediction under different working conditions. *Reliability Engineering & System Safety* 2025; 256: 110790. https://doi.org/10.1016/j.ress.2024.110790.
18. Lyu Y, Xiao X, Fan R, Chen C. Multisource Feature Separation and Weighted Network for Cross-Conditional Capacity Estimation of Lithium Batteries. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 2025; 55(10): 6842-6856. https://doi.org/10.1109/tsmc.2025.3585260.
19. Chen X, Sun T, Lai X, Zheng Y, Han X. Transfer learning strategies for lithium-ion battery capacity estimation under domain shift differences. *Journal of Energy Storage* 2024; 90: 111860. https://doi.org/10.1016/j.est.2024.111860.
20. Ma Y, Shan C, Gao J, Chen H. Multiple health indicators fusion-based health prognostic for lithium-ion battery using transfer learning and hybrid deep learning method. *Reliability Engineering & System Safety* 2023; 229: 108818. https://doi.org/10.1016/j.ress.2022.108818.
21. Mu D, Zhang W, Lin B, Lin C, Lu Y. Multi-source transfer network for cross-domain state of charge estimation of lithium-ion batteries. *Journal of Energy Storage* 2025; 123: 116636. https://doi.org/10.1016/j.est.2025.116636.
22. Dong G, Hua N, Chen H, Lou Y. Deep Transfer Learning Enabled State of Health Estimation of Lithium-Ion Battery Using Voltage Sample Entropy Under Fast Charging Profiles. *IEEE Transactions on Transportation Electrification* 2025; 11(1): 3703-3714. https://doi.org/10.1109/tte.2024.3445344.
23. Shen L, Li J, Zuo L, Zhu L, Shen HT. Source-Free Cross-Domain State of Charge Estimation of Lithium-Ion Batteries at Different Ambient Temperatures. *IEEE Transactions on Power Electronics* 2023; 38(6): 6851-6862. https://doi.org/10.1109/tpel.2023.3251568.
24. Han T, Yue S, Yang P, Zhou R, Yu J. Source-Free Dynamic Weighted Federated Transfer Learning for State-of-Health Estimation of Lithium-Ion Batteries With Data Privacy. *IEEE Transactions on Power Electronics* 2024; 39(11): 15085-15100. https://doi.org/10.1109/tpel.2024.3434751.
25. Zou W, Shen Z, Li Q, Ge J, Li C, Chen X, Shen X, Huang L, Luo B. Experimental Evaluation of Parameter-Efficient Fine-Tuning for Software Engineering Tasks. *ACM Transactions on Software Engineering and Methodology* 2025; 34(7): 1-34. https://doi.org/10.1145/3722107.
26. Severson KA, Attia PM, Jin N, Perkins N, Jiang B, Yang Z, Chen MH, Aykol M, Herring PK, Fraggedakis D, Bazant MZ, Harris SJ, Chueh WC, Braatz RD. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy* 2019; 4(5): 383-391. https://doi.org/10.1038/s41560-019-0356-8.
27. Wang Y, Guo S, Cui Y, Deng L, Zhao L, Li J, Wang Z. A comprehensive review of machine learning-based state of health estimation for lithium-ion batteries: data, features, algorithms, and future challenges. *Renewable and Sustainable Energy Reviews* 2025; 224: 116125. https://doi.org/10.1016/j.rser.2025.116125.
28. Purushothaman D, Narayanamoorthi R, Ramachandramurthy VK. Charging ahead: Unlocking the potential of constant voltage and constant current modes in WPT for EVs. *Journal of Energy Storage* 2024; 96: 112603. https://doi.org/10.1016/j.est.2024.112603.
29. Yu J, Yin H, Xia X, Chen T, Li J, Huang Z. Self-Supervised Learning for Recommender Systems: A Survey. *IEEE Transactions on Knowledge and Data Engineering* 2024; 36(1): 335-355. https://doi.org/10.1109/tkde.2023.3282907.
30. Yin X, Pan T, Tian J, Ni L, Lao L. Voltage-fault diagnosis for battery pack in electric vehicles using mutual information. *Journal of Power Sources* 2024; 608: 234636. https://doi.org/10.1016/j.jpowsour.2024.234636.
31. Peng S, Wang Y, Tang A, Jiang Y, Kan J, Pecht M. State of health estimation joint improved grey wolf optimization algorithm and LSTM using partial discharging health features for lithium-ion batteries. *Energy* 2025; 315: 134293. https://doi.org/10.1016/j.energy.2024.134293.
32. Zhang C, Luo L, Yang Z, Du B, Zhou Z, Wu J, Chen L. Flexible method for estimating the state of health of lithium-ion batteries using partial charging segments. *Energy* 2024; 295: 131009. https://doi.org/10.1016/j.energy.2024.131009.
33. Pang X, Yu P, Yin C, Liu W, Zheng Z. Transformer and bidirectional gated recurrent unit hybrid network with attention mechanisms for lithium-ion battery state-of-health estimation. *Journal of Power Sources* 2025; 658: 238370. https://doi.org/10.1016/j.jpowsour.

2025.238370.

34. Wei Y, Grau G, Wu D. Sheet resistance prediction of laser induced graphitic carbon with transformer encoder-enabled contrastive learning. *Journal of Intelligent Manufacturing* 2025; 36(3): 1983-1997. https://doi.org/10.1007/s10845-024-02333-2.
35. Li F, Yu Y, Yuan X, Ren G. State-of-health estimation for lithium-ion batteries using unsupervised deep subdomain adaptation. *Energy* 2025; 324: 135862. https://doi.org/10.1016/j.energy.2025.135862.
36. Xu H, Xu Q, Duanmu F, Shen J, Jin L, Gou B, Wu F, Zhang W. State-of-Charge Estimation of Lithium-Ion Batteries Based on EKF Integrated With PSO-LSTM for Electric Vehicles. *IEEE Transactions on Transportation Electrification* 2025; 11(1): 2311-2321. https://doi.org/10.1109/tte.2024.3421260.
37. Yu M, Zhu Y, Gu X, Zhang Q, Shang Y. Incremental Expansion Analysis and State-of-Health Estimation for Lithium-Ion Batteries. *IEEE Transactions on Industrial Electronics* 2026; 73(1): 546-557. https://doi.org/10.1109/tie.2025.3587113.
38. Fang P, Sui X, Zhang A, Wang D, Yin L. Fusion model based RUL prediction method of lithium-ion battery under working conditions. *Eksploatacja i Niezawodność – Maintenance and Reliability* 2024; 26(3): 186537. https://doi.org/10.17531/ein/186537.
39. Liu X, Gao Z, Tian J, Wei Z, Fang C, Wang P. State of Health Estimation for Lithium-Ion Batteries Using Voltage Curves Reconstruction by Conditional Generative Adversarial Network. *IEEE Transactions on Transportation Electrification* 2024; 10(4): 10557-10567. https://doi.org/10.1109/tte.2024.3386212.
40. Qi W, Qin W, Yun Z. Closed-loop state of charge estimation of Li-ion batteries based on deep learning and robust adaptive Kalman filter. *Energy* 2024; 307: 132805. https://doi.org/10.1016/j.energy.2024.132805.
41. Fan X, Li B, Xie Z, Hao Y, Tang Q, Hu X. An improved Transformer incorporating fuzzy information entropy and average input strategy for SOC estimation of lithium-ion battery. *Energy* 2025; 330: 136953. https://doi.org/10.1016/j.energy.2025.136953.
42. Zhao SY, Ou K, Gu XX, Dan ZM, Zhang JJ, Wang YX. A novel transformer-embedded lithium-ion battery model for joint estimation of state-of-charge and state-of-health. *Rare Metals* 2024; 43(11): 5637-5651. https://doi.org/10.1007/s12598-024-02942-z.
43. Meng J, Hu D, Lin M, Peng J, Wu J, Stroe DI. A Domain-Adversarial Neural Network for Transferable Lithium-Ion Battery State-of-Health Estimation. *IEEE Transactions on Transportation Electrification* 2025; 11(3): 7732-7742. https://doi.org/10.1109/tte.2025.3530536.
44. Fang Y, Yap PT, Lin W, Zhu H, Liu M. Source-free unsupervised domain adaptation: A survey. *Neural Networks* 2024; 174: 106230. https://doi.org/10.1016/j.neunet.2024.106230.
45. Lyu Y, Zhang Q, Chen A, Wen Z. Interval Prediction of Remaining Useful Life based on Convolutional Auto-Encode and Lower Upper Bound Estimation. *Eksploatacja i Niezawodność – Maintenance and Reliability* 2023; 25(2): 165811. https://doi.org/10.17531/ein/165811.
46. Zhu R, Hu J, Peng W. Bayesian calibrated physics-informed neural networks for second-life battery SOH estimation. *Reliability Engineering & System Safety* 2025; 264: 111432. https://doi.org/10.1016/j.ress.2025.111432.