Fine-Grained Subdomain Alignment and Feature Grouping-Based Cross-Domain SOH Estimation for Lithium-Ion Batteries Using a Patch Time Series CNN-Transformer Network

Xuanang Gui, Shu Zhang, Yuheng Cheng, Qianlong Wang, Tong Zhao, Huan Zhao, *Member, IEEE*, and Junhua Zhao, *Senior Member, IEEE*

***Abstract*—State-of-health (SOH) estimation plays a crucial role in ensuring the safe, reliable, and efficient management of lithium-ion batteries. However, the performance of data-driven models often degrades in cross-domain scenarios due to domain discrepancies and insufficient labeled target data. To address these challenges, we propose a novel framework—Fuzzy Cluster Subdomain Adaptation Patch Time Series CNN-Transformer (FCSA-PatchTSCT)—for cross-domain SOH estimation. This framework distinctively integrates a fine-grained subdomain alignment strategy using fuzzy clustering local maximum mean discrepancy for precise localized knowledge transfer, a patch-based network structure to enhance the local temporal feature extraction, and a Pearson correlation-guided homogeneous feature grouping mechanism to mitigate interference from variable heterogeneity. By addressing both temporal feature variations and domain shifts at a finer granularity, the framework enhances model generalization. It particularly excels in transferring knowledge to target batteries even when data are limited or derived from incomplete charging cycles. Comprehensive evaluations on diverse cross-domain transfer tasks, including those under practical partial charging conditions, confirm that FCSA-PatchTSCT achieves superior accuracy and stability. These results underscore its high potential for robust SOH estimation in real-world battery management systems.**

***Index Terms*—Feature grouping, lithium-ion batteries, patch-based modeling, state-of-health (SOH) estimation, subdomain adaptation.**

Received 20 May 2025; revised 28 August 2025 and 3 November 2025; accepted 18 November 2025. Date of publication 24 November 2025; date of current version 25 February 2026. This work was supported in part by the National Natural Science Foundation of China under Grant 72331009 and Grant 72171206; in part by the Guangdong Power Grid Company under Grant GDKIXM20231024; in part by The Hong Kong Polytechnic University under Grant P0051105; and in part by the Shenzhen Key Laboratory of Crowd Intelligence Empowered Low-Carbon Energy Network under Grant ZDSYS20220606100601002. Recommended for publication by Associate Editor M. Ferdowsi. *(Xuanang Gui and Shu Zhang contributed equally to this work.)* *(Corresponding authors: Xuanang Gui; Junhua Zhao.)*

Xuanang Gui, Shu Zhang, and Tong Zhao are with the School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen 518172, China (e-mail: xuananggui@link.cuhk.edu.cn; shu.zhang2@siat.ac.cn; tongzhao1@link.cuhk.edu.cn).

Yuheng Cheng and Junhua Zhao are with the School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen 518172, China, and also with the Center for Crowd Intelligence, Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen 518129, China (e-mail: yuhengcheng@link.cuhk.edu.cn; zhaojunhua@cuhk.edu.cn).

Qianlong Wang is with the Business School, Nankai University, Tianjin 300071, China (e-mail: wangqianlong21@mails.ucas.ac.cn).

Huan Zhao is with the Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Hong Kong (e-mail: huan-paul.zhao@polyu.edu.hk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPEL.2025.3636099.

Digital Object Identifier 10.1109/TPEL.2025.3636099

I. Introduction

With the global transition toward a low-carbon and sustainable energy structure, lithium-ion batteries (LIBs) have become central to electric vehicles (EVs), renewable energy storage systems, and portable electronic devices [1], owing to their high energy density, long cycle life, and low self-discharge rate. However, long-term usage inevitably leads to capacity degradation and performance deterioration, which in severe cases can pose safety hazards such as thermal runaway [2]. Therefore, accurately assessing the state-of-health (SOH) of LIBs is crucial for ensuring battery safety, extending lifespan, and optimizing battery management systems (BMS), creating a critical need for robust estimation methods that can generalize from controlled laboratory settings to the diverse and unpredictable conditions of real-world operation [3], [4]. It also supports applications such as second-life use and recycling, promoting sustainable development [5].

Recent research in SOH estimation for LIBs is broadly classified into three categories: experimental, model-based, and data-driven methods [6]. Experimental approaches, such as the open-circuit current method [7] and electrochemical impedance spectroscopy (EIS) [8], rely on specialized equipment for direct measurement. However, their complexity and time-consuming nature make them unsuitable for real-time estimation. Alternatively, model-based methods establish a relationship between battery signals and SOH by using equivalent circuit [9] or electrochemical models [10] to describe the battery's internal mechanisms.

While these model-based approaches offer good interpretability and can achieve high accuracy, often enhanced with observers [11] or filtering techniques [12], [13], their practical application is limited. Their primary drawback is the reliance on precise physical models that demand deep domain expertise, a significant barrier for complex or newly emerging battery

chemistries [14]. Consequently, their generalization capability is poor across diverse battery types and real-world operational scenarios where model assumptions may not holds.

With the rapid development of Big Data technology, data-driven SOH estimation methods have gained increasing attention from researchers. Unlike experimental and model-based methods, data-driven methods rely on machine learning or deep learning models to learn potential degradation patterns from historical operational data. Early research primarily employed traditional machine learning techniques, such as support vector machines (SVM) [15], support vector regression (SVR) [16], random forest (RF) [17], and gradient boosting decision trees (GBDT) [18], which have achieved notable success in SOH estimation. In recent years, deep learning (DL) methods have been widely applied, including convolutional neural networks (CNN) [19], recurrent neural networks (RNN) [20], long short-term memory networks (LSTM) [21], and Transformer models [22], [23]. These models, however, often generalize poorly. Distribution discrepancies between training and target batteries—such as limited operational history in new cells—can significantly degrade prediction accuracy [24].

TL methods transfer knowledge extracted from supervised source domain with similar degradation characteristics to the target domain. In the field of SOH estimation, early research mainly focused on TL methods based on pretrained fine-tuning [25]. In these methods, the target domain and source domain LIBs use the same deep network structure, with the network parameters pretrained on the source domain data transferred to the target domain network for initialization, followed by fine-tuning using target domain data. Shen et al. [26] used pretrained deep convolutional neural networks to transfer source-learned parameters to improve the capacity estimation for LIBs with limited training data. Huang et al. [27] utilized fine-tuning as a transfer learning strategy to adapt the model to different battery types, improving the accuracy and robustness of SOH estimation across various datasets. However, such approaches require labeled target data, which is often unavailable in practice [28].

Domain adaptation (DA) is a transfer learning method that achieves knowledge transfer from unlabeled target data by learning domain-invariant features in a shared subspace [29]. DA methods typically follow two main strategies [30]: minimizing statistical discrepancies or leveraging adversarial learning. Structure-based methods employ metrics like maximum mean discrepancy (MMD) [31] to directly minimize the statistical distance between domains. Alternatively, adversarial methods use a domain classifier that competes with the feature extractor, compelling it to learn domain-invariant features. These DA strategies have been successfully applied to SOH estimation for LIBs. For instance, Ma et al. [28] combined a CNN for feature extraction from raw charging voltage data with MMD to minimize domain discrepancy. Han et al. [32] integrated a domain adaptation layer into a deep LSTM network for degradation feature alignment across discrepant batteries. Meng et al. [33] employed a domain-adversarial neural network (DANN) for feature alignment, using EIS-derived health features extracted by an autoencoder. Yao et al. [34] introduced a semisupervised adversarial deep learning method, applying adversarial training to both labeled and unlabeled data to improve model generalization.

However, existing DA-based LIB SOH estimation methods face significant limitations hindering their practical application. First, their reliance on unsupervised global alignment is a key issue. Methods in [35], [36], for instance, use MMD-based or DANN-based approaches that overlook the distinct progression of battery degradation. This global approach misses crucial fine-grained temporal variations [37]. This limitation persists even in advanced subdomain methods like local MMD (LMMD) [38]. By relying on hard assignments to predefined stages, LMMD struggles to model the gradual transitions inherent in the continuous SOH degradation process, where stage boundaries are inherently fuzzy. Second, current architectures often fail to capture local dynamics and manage feature heterogeneity. While models like CNNs or transformers are common, they struggle with the subtle local patterns inherent in battery data [39], [40], [41]. Furthermore, the inherent heterogeneity across features, where variables show opposing degradation trends, creates channel interference and reduces accuracy in existing models [42], [43]. Collectively, these interconnected challenges—imprecise alignment for continuous processes, insensitivity to local patterns, and feature heterogeneity—impede the development of robust cross-domain SOH models.

To address the multifaceted challenges identified above, this article proposes a novel framework, FCSA-PatchTSCT, which integrates three synergistic components into a holistic solution. The core novelty lies in how these components work in concert to achieve robust cross-domain SOH estimation. The main contributions are as follows.

1) A cross-domain SOH estimation framework centered on a fuzzy clustering local maximum mean discrepancy (FCLMMD) method is proposed for subdomain adaptation. By leveraging fuzzy clustering to define soft subdomains based on battery degradation stages, FCLMMD achieves a fine-grained adaptation through precise alignment of their distributions. This targeted local alignment is complemented by the DANN framework, which provides robust global domain adaptation.
2) A PatchTSCT network framework is introduced, which employs a patch-based architecture to enhance temporal feature extraction. By dividing each time series into smaller patches and processing them independently, the model focuses on local characteristics to more effectively capture intricate temporal patterns. This approach significantly improves the learning of fine-grained features and boosts overall prediction performance.
3) A homogeneous feature grouping method based on Pearson correlation to tackle variable heterogeneity is developed. This method clusters highly correlated features into distinct groups that are fed into separate processing channels. This design effectively reduces interference among heterogeneous variables, enabling more robust feature extraction and an improved capture of relevant degradation information.

By synergistically combining these three innovations, our framework provides a comprehensive solution that is demonstrably more accurate and generalizable than prior approaches. The article is structured as follows. Section II details the methodology, covering problem definition, method overview, and model

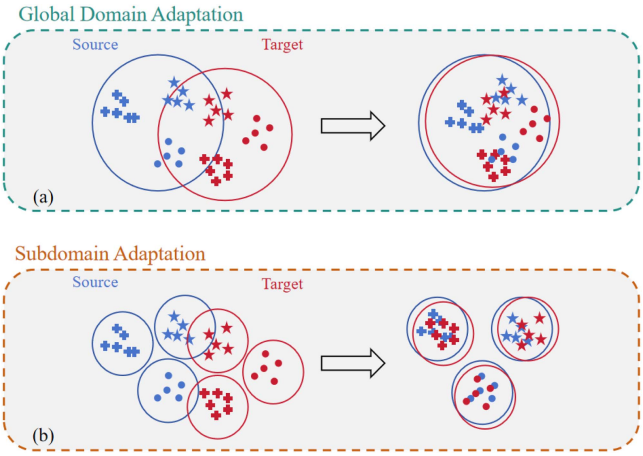

Fig. 1. Minimizing domain discrepancy through (a) global DA might destroy fine-grained information, whereas (b) subdomain adaptation aligns each corresponding pair of subdomains.

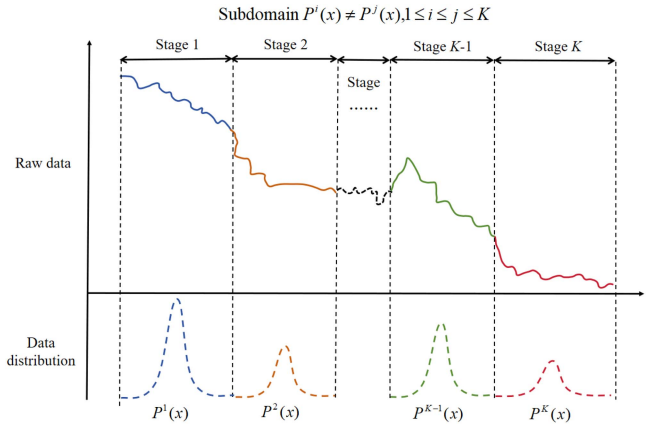

Fig. 2. Substructures of the domain.

development. Section III reports experimental results, including the evaluation of fine-grained domain adaptation and homogeneous feature grouping strategies, comparisons across model architectures, and assessment under nonfull charging conditions. Section IV concludes the study and discusses potential directions for future work.

II. Methodology

A. Problem Definition

1) Cross domain: This article explores a subdomain adaptation approach for the SOH estimation in cross-domain conditions. Let $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ denote the source domain data, where x_i^s is the ith sample in the source domain and y_i^s is the corresponding SOH label. N_s is the number of samples in the source domain, the sample meets both the marginal probability distribution $P_S(x)$ and the conditional probability distribution $P_S(y|x)$. Let $D_t = \{x_i^t\}_{i=1}^{N_t}$ denote the data from the target domain, where x_i^t is the ith sample of the target domain and N_t represents the total samples number of the target domain. The samples in the target domain do not have corresponding SOH labels. The marginal and conditional probability distributions of D_t are $P_T(x)$ and $P_T(y|x)$.

For cross-domain problem, the following basic assumptions are made. (a) samples in D_s are sufficiently labeled to enable valid predictor training. (b) D_t can only provide unlabeled data for use. (c) the marginal probability distribution of the source domain differs from that of the target domain, $P_S(x) \neq P_T(x)$, but the conditional probability distribution of the source domain is equivalent to that of the target domain $P_S(y|x) \approx P_T(y|x)$.

Based on the aforementioned assumptions, most studies have conducted a global alignment between the source and target domains, overlooking the presence of distinct substructures within each domain, which can lead to suboptimal feature alignment and reduced prediction accuracy, as shown in Fig. 1. In classification tasks, each sample category can be treated as a subdomain, with the number of categories corresponding to the number of subdomains. Subdomain adaptation, therefore, involves aligning the data from different categories separately [44]. However, subdomain adaptation has been rarely explored in SOH estimation tasks. This is because such tasks are fundamentally regression problems, where the distribution of time series exhibits continuous, time-varying properties, making it difficult to define subdomains. As a result, determining how to obtain substructure labels is a critical issue for implementing subdomain adaptation in regression tasks. Inspired by the degradation characteristics of LIBs, we discretize the degradation process into K stages, with each stage following the same distribution internally but differing across stages. Therefore, there are substructures in the domains, for $D = \{(x_i, y_i)\}_{i=1}^{N}$, i.e., $D = \{D^1, \ldots, D^K\}$. Data in the same stage follow the same distribution $P^i(x)$. For different stages $1 \leq i \neq j \leq K$ $P^i(x) \neq P^j(x)$ and $P^i(y|x) \neq P^j(y|x)$, as shown in Fig. 2.

2) Cross variable: In real-world data, cross-variable interactions exhibit both homogeneity and heterogeneity, which significantly impact SOH estimation in multidimensional time-series prediction. Homogeneous variables tend to display similar behavior, making it easier to model their interactions, as the features evolve in a predictable direction over time. However, heterogeneous variables show opposing behaviors, such as some features increasing while others decrease, leading to more complex relationships. This variation presents challenges, as the model must account for the opposing trends between features. As shown in Fig. 3, the negative correlation between feature 1, feature 2, and feature 3 and the positive correlation between Feature 2 and Feature 3 clearly demonstrate these cross-variable relationships.

B. Method Overview

In this article, a novel framework FCSA-PatchTSCT is proposed for cross-domain estimation considering the LIBs degradation in stages. Our proposed FCSA-PatchTSCT framework follows a systematic, multistage strategy. It first disentangles heterogeneous variables via feature grouping, then performs fine-grained alignment of degradation stages using fuzzy subdomain adaptation, and finally extracts robust local temporal

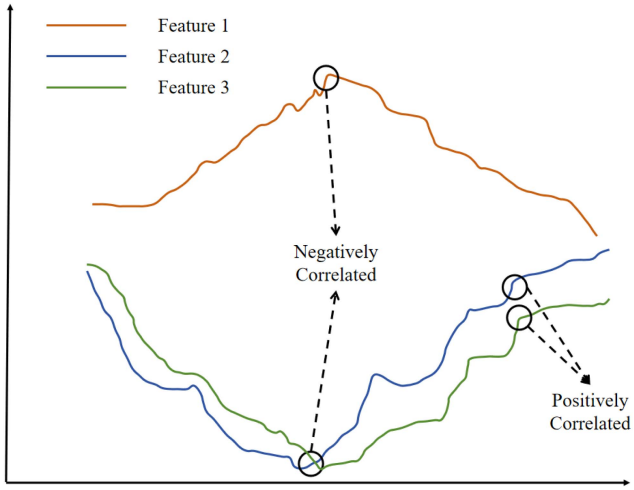

Fig. 3. Homogeneous and heterogeneous relationships between variables.

patterns with a patch-based network. The detailed process is shown in Fig. 4 and the steps are described as follows.

1) *Data acquisition and preprocessing:* First, a set of datasets from LIBs with charging voltage data are acquired, and the features of the charging curves are extracted to effectively reflect the SOH degradation trend.
2) *Model construction and training:* Next, the FCSA-PatchTSCT is constructed, while the loss function and the hyperparameter are set. The training data are divided into discrete substructure based on fuzzy clustered algorithm, and homogeneous features are grouped simultaneously based on their correlation coefficients. The processed data are fed into the model for training, and the optimal model is saved during the training process.
3) *Model testing:* Finally, the testing data are fed into the optimal model, the estimation results are used to evaluate the performance of the proposed method.

C. Model Development

1) Fuzzy substructure division and homogeneous feature grouping: To implement fine-grained subdomain adaptation in SOH estimation, the fuzzy C-means (FCM) clustering algorithm is utilized to obtain substructure labels and fuzzy stage memberships for LIBs degradation. Unlike traditional clustering algorithms, FCM assigns each data point to all clusters with a degree of membership, indicating how strongly the data point belongs to each cluster.

First of all, the number of clusters C is determined empirically. The original degradation sequence can be defined as $x = \{x_1, x_2, \ldots, x_n\}$, where n denotes the length of the degradation time series. The input to the FCM algorithm consists of vectors x_i, each composed of health features (e.g., charging time to a specific voltage) known to be sensitive to the battery's degradation state.

The core of our approach is leveraging FCM's soft memberships, which is critical for two reasons. First, they create functionally meaningful clusters corresponding to degradation stages rather than mere temporal segments. Second, they robustly model the gradual transitions between these stages, which enhances the alignment process, particularly when dealing with imbalanced subdomain distributions. The resulting fuzzy membership u_{ij} represents the likelihood that data point x_i belongs to cluster j, and is computed as

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - c_j\|^2}{\|x_i - c_k\|^2} \right)^{\frac{1}{m-1}}} \tag{1}$$

where $\| \bullet \|$ denotes the Euclidean distance norm between the sample x_i and the class center c_j. m represents the fuzzy factor. Then, each cluster's center point $C = \{c_1, c_2, \ldots c_C\}$ is computed based on the fuzzy membership u_{ij} of each data point x_i. The ith cluster's center point can be updated as follows:

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m x_i}{\sum_{i=1}^{N} x_{ij}^m}. \tag{2}$$

The FCM algorithm minimizes the objective function J by iteratively refining the membership values and cluster centers, and the objective function J along with its corresponding constraints can be described as follows:

$$J = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \|x_i - c_j\|^2, \quad 0 \le u_{ij} \le 1 \tag{3}$$

$$\sum_{j=1}^{C} u_{ij} = 1, \quad 0 \le i \le N \tag{4}$$

$$\sum_{i=1}^{N} u_{ij} \le N, \quad 0 \le j \le C. \tag{5}$$

Each cluster center is initially assigned randomly at the start of the iterative process, and the iteration ends when it reaches a local optimum. The convergence criterion is determined by the change in the membership matrix U between successive iterations, and the process is terminated once this change falls below a predefined threshold ϵ, as defined below:

$$\|U^{k+1} - U^k \le \epsilon\|. \tag{6}$$

Finally, the LIBs' substructure labels $i \in \{1, 2, \ldots C\}$ and fuzzy membership are obtained, which will be used as the input data of the subdomain adaptation network.

Alongside subdomain division, we also perform homogeneous feature grouping based on the Pearson correlation coefficient. Specifically, for each feature in the dataset, we calculate the Pearson correlation coefficient with all other features. Features that show a high positive correlation are grouped together into homogeneous feature groups. This ensures that features which behave similarly over time are processed in the same way, reducing redundancy and improving the learning process by capturing the relevant patterns in the data. The Pearson correlation coefficient r is calculated as

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}} \tag{7}$$

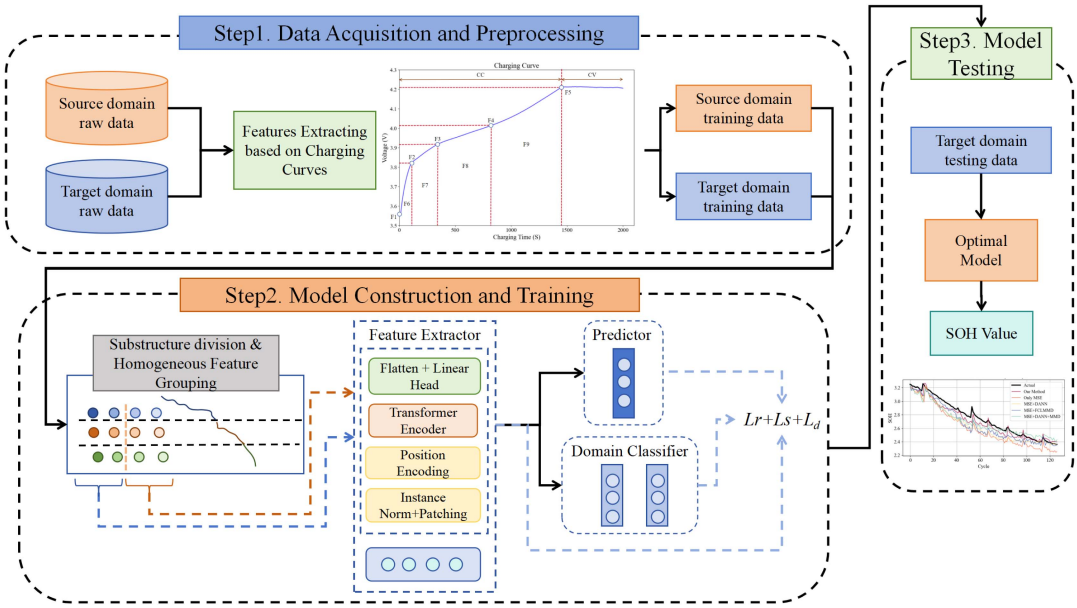

Fig. 4. Procedures of the proposed SOH estimation method.

Algorithm 1: Fuzzy C-Means Clustering Algorithm.

Input: The original degradation sequence $\{x_1, x_2, \ldots, x_N\}$, the number of clusters C
Output: Fuzzy membership matrix $U = \{u_{ij}\}_{C\times N}$
Randomly initialize fuzzy membership matrix $U = \{u_{ij}\}_{C\times N}$
repeat
  Compute the fuzzy centers $C = \{c_1, c_2, \ldots, c_i, \ldots, c_C\}$ using (2)
  Compute the objective function J
  Calculate the fuzzy membership matrix U
until $\|U^{k+1} - U^k\| < \epsilon$
Determine the stage label I_i of point x_i by calculating $I_j = \max\{u_{1i}, u_{2i}, \ldots, u_{Ci}\}$
Return fuzzy membership matrix U and stage label $\{I_i\}_{i=1}^N$

where X_i and Y_i are the values of the features X and Y. $\bar{X}$ and $\bar{Y}$ are the mean values of the respective features. Features with a correlation coefficient higher than a certain threshold are grouped together, while those with low correlation are treated as separate feature groups.

By combining fuzzy substructure division and homogeneous feature grouping, the data are processed in a more structured way, enabling the model to focus on the most relevant features while reducing noise and improving the accuracy of the SOH estimation.

2) SOH estimation based on FCSA-PatchTSCT: In this section, we propose a subdomain adaptation network, FCSA-PatchTSCT, which combines degradation substructure division and subdomain adaptation for estimating the SOH of LIBs,

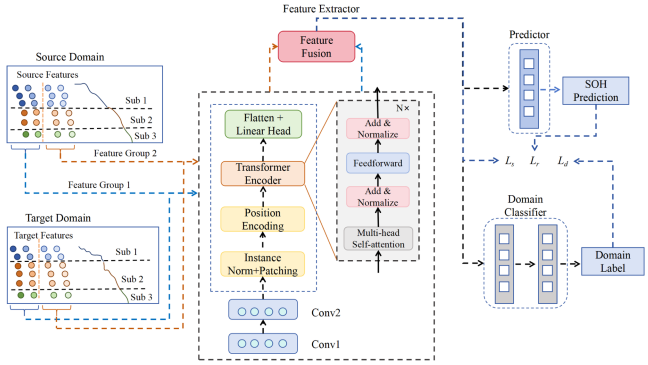

Fig. 5. Overall structure and details of FCSA-PatchTSCT.

taking into account degradation in stages. The network leverages fine-grained data alignment to effectively minimize the mismatch deviation inherent in global domain adaptation during the transfer process. The proposed FCSA-PatchTSCT consists of a feature extractor FE, a predictor P and a domain classfier D, and the framework is shown in Fig. 5.

1) Feature Extractor: The feature extractor is built upon a multichannel PatchTSCT architecture. The process begins by assigning each sample a subdomain label and partitioning its features into homogeneous groups based on Pearson correlation. These feature groups, along with their subdomain information, are then fed into separate parallel channels.

Within each channel, a patching operation first divides the input time series into smaller segments called patches to better capture local dynamics while still considering global information. Moreover, this process transforms the original input sequence length L into a reduced number of tokens, approximately L/S, where S is the stride (the time interval between the start of adjacent patches), resulting in a quadratic reduction in

the computation and memory usage of the self-attention maps, which significantly enhances the model's training efficiency. These patches are subsequently processed through a stack of CNN, multihead self-attention, and feedforward layers to independently extract features. Finally, the outputs from all channels are integrated by a feature fusion layer, creating a unified representation that effectively combines information from the different subdomains and homogeneous feature groups.

Crucially, this architectural design is explicitly linked to its corresponding performance gain. The effectiveness of the fine-grained domain alignment, which is driven by our FCLMMD loss metric, is critically dependent on the quality of this unified feature representation. By capturing local dynamics and disentangling features, the extractor produces a discriminative feature space where the alignment can be most meaningful and impactful.

FCLMMD improves upon existing methods in two key ways. First, unlike traditional MMD which only considers the global distribution, FCLMMD leverages the probability of each sample belonging to a specific subdomain, enabling a more precise data alignment. Second, it refines class-level methods like LMMD—which typically rely on confident, hard labels—by incorporating soft subdomain probabilities. This approach is better suited for real-world SOH scenarios, as it more robustly handles the label noise and ambiguous stage boundaries inherent in the continuous battery degradation process.

Given the source domain $D_s = \{(x_i^s, y_i^s, u_i^s)\}_{i=1}^{N_s}$ and the target domain $D_t = \{(x_i^t, u_i^t)\}_{i=1}^{N_t}$, where x_i^s and x_i^t are samples from the source and target domain, respectively, s_i^s and s_i^t are corresponding fuzzy memberships. D_s and D_t is divided into C subdomains $D_s^{(c)}$ and $D_t^{(c)}$, where $c = 1, 2, \ldots C$. For corresponding subdomains, the squared form of FCLMMD can be defined as follows:

$$d_{\mathcal{H}}(p,q) \triangleq E_c \| E_{p^{(c)}}\left[\phi(x^S)\right] - E_{q^{(c)}}\left[\phi(x^T)\right] \|_{\mathcal{H}}^2 \quad (8)$$

where $p^{(c)}$ and $q^{(c)}$ are distributions of $D_s^{(c)}$ and $D_t^{(c)}$, respectively. $\mathcal{H}$ represents the reproducing Hilbert space (RKHS) with a kernel function k, and $k(x^s, x^t) = <\phi(x^s), \phi(x^t)>$. $\phi(\bullet)$ denotes the function that mapping the original data to the RKHS. Assuming that each sample belongs to each category according to the weight w^c, the unbiased estimation of FCLMMD is calculated as

$$\hat{d}_{\mathcal{H}}(p,q) = \frac{1}{C}\sum_{c=1}^{C}\left\| \sum_{x_i^S \in D_S} w_i^{S(c)}\phi(x_i^S) - \sum_{x_j^T \in D_T} w_j^{T(c)}\phi(x_j^T)\right\|_{\mathcal{H}}^2 \quad (9)$$

where $w_i^{S(c)}$ and $w_j^{T(c)}$ are the weight of x_i^S and x_j^T belonging to class c, respectively. Both $\sum_{i=1}^{N_S} w_i^{S(c)}$ and $\sum_{i=1}^{N_T} w_i^{T(c)}$ are equal to 1. $\sum_{x_i \in D} w_i^{(c)}\phi(x_i)$ is a weighted sum of the category c. The weight of a sample x_i is calculated as

$$w_i^c = \frac{u_{ic}}{\sum_{(x_i, y_i, u_i) \in \mathcal{D}} u_{jc}} \quad (10)$$

where u_{ic} is the membership of the cth element of the vector u_i. Since the fuzzy membership function can effectively characterize the probability distribution of a sample x_i across different health stages, we use the fuzzy memberships $u_i^{S(c)}$ and $u_i^{T(c)}$ to calculate the sample weights $w_i^{S(c)}$ and $w_i^{T(c)}$. Based on FCLMMD, we designed a subdomain data structure information alignment loss, which is expressed as

$$L_F = \hat{d}_H\left(FE(X^s), FE(X^t)\right) \quad (11)$$

where $FE(X^s)$ and $FE(X^t)$ denote the source domain features and target domain features extracted from FE.

2) Predictor: The predictor P is an estimation network composed of two fully connected layers, and it performs SOH estimation by receiving the features obtained from the FE. The SOH estimation values can be calculated as $P(FE(X^s))$. To ensure the predictor delivers accurate results, the prediction loss L_p is computed using the mean squared error (MSE) between the true labels and the predicted labels. Since the target domain data are unlabeled, the loss for this part is defined as

$$L_P = \frac{1}{N_s}\sum_{i=1}^{N_s}\left(P\left(FE\left(x_i^s\right)\right) - y_i^s\right)^2. \quad (12)$$

3) Domain Classifier: The domain classifier D is designed to identify whether the extracted high-dimensional temporal features originate from the source domain or the target domain. Through adversarial learning, the feature extractor strives to confuse the domain discriminator. When the minimax game between the feature extractor and the domain discriminator reaches equilibrium, domain-invariant features are successfully learned. The loss for this component is represented by binary cross-entropy, and its calculation is as follows:

$$\begin{aligned} L_D = & -\frac{1}{N_s}\sum_{i=1}^{N_s}\log D(FE(x_i^s)) \\ & -\frac{1}{N_t}\sum_{i=1}^{N_t}\log(1 - D(FE(x_i^t))) \end{aligned} \quad (13)$$

where the value of $D(\cdot)$ is either 0 or 1, which can be used to distinguish the domain to which a sample belongs. By minimizing this part of the loss, domain label alignment can be achieved.

3) Training process and optimization objective: To bridge the gap between the source and target domains and enhance the prediction accuracy, the loss function is defined as follows:

$$L_t = L_P + \alpha L_F + \beta L_D \quad (14)$$

where α and β are tradeoff parameters. The Adam optimizer is utilized to minimize the total loss. In this framework, DANN provides a coarse global alignment while FCLMMD performs fine-grained subdomain alignment, and potential conflicts between these two objectives are naturally resolved through the unified weighted loss, where α and β explicitly control the tradeoff.

To facilitate clarity and consistency in the following formulations, Table I summarizes the key parameters used throughout the equations, along with their concise descriptions.

TABLE I
KEY PARAMETERS IN THE EQUATIONS

Parameters	Descriptions
D_s, D_t	Source domain data and target domain data
x_i^s, x_i^t	The ith sample in the source domain and target domain
C	The number of clusters
J	FCM objective function
U	Membership matrix in FCM
r	Pearson correlation coefficient between two features
$D_s^{(c)}$, $D_t^{(c)}$	Subdomains of source and target domain for cluster c
$p^{(c)}$, $q^{(c)}$	Distributions of source and target subdomain c
$\mathcal{H}$	Reproducing Kernel Hilbert Space (RKHS)
$\phi(\bullet)$	Mapping function from original space to RKHS
w_i^c	Normalized sample weight in cluster c
$FE(\bullet)$	Feature extractor
L_F	Subdomain alignment loss based on FCLMMD
L_P	MSE loss between predicted and true SOH in the source domain
L_D	Domain label alignment loss based on DANN

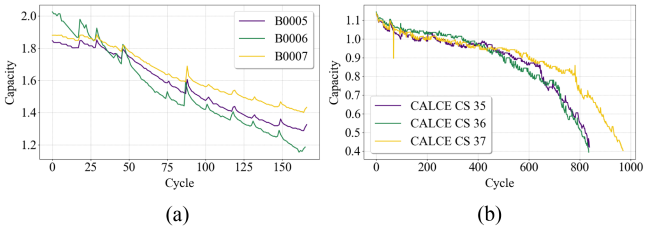

Fig. 6. Capacity degradation curves of LIBs datasets. (a) LIBs dataset from NASA. (b) LIBs dataset from CALCE.

III. EXPERIMENTS

A. Set Up

To evaluate the performance of the proposed method, two public datasets from NASA Prognostics Center of Excellence [45] and Center for Advanced Life Cycle Engineering (CALCE) [46] are used in this article. The batteries in both datasets are of the Lithium Cobalt Oxide (LCO) type, a chemistry primarily utilized in consumer electronics due to its high energy density. The NASA dataset consists of 18650 LIBs undergoing multiple cycles of charge, discharge, and impedance under experimental conditions at 24°C. For this research, the B0005, B0006 and B0007 LIBs datasets are utilized. During the charging process, the LIBs were charged at 1.5-A constant current (CC) until the battery voltage reached 4.2 V and were continued at a 4.2 V constant voltage (CV) until the current dropped to 20 mA. Then, the LIBs discharge at a constant CC level of 2 A until the voltages dropped to 2.7, 2.5, and 2.2 V, respectively.

The CALCE datasets are collected by the U.S. Arbin BT2000 battery test system. In this study, LIBs numbered CS2_35, CS2_36, and CS2_37 are selected for research. The charge–discharge protocol of these LIBs is similar to that of the NASA dataset.

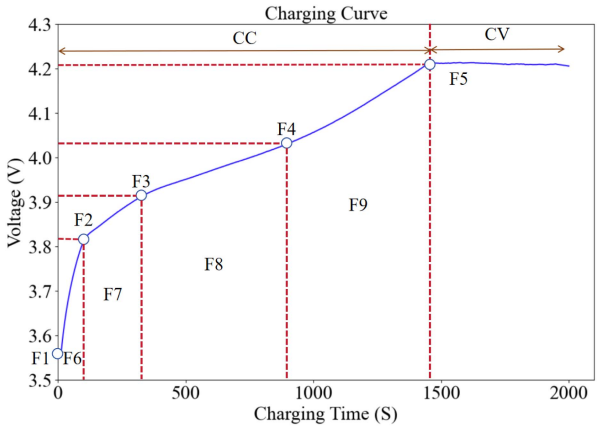

Fig. 7. Graphical representation of the features-extracted.

TABLE II
EXPERIMENTAL INFORMATION OF THE TRANSFER LEARNING TASKS

Task	Source Domain Data	Target Domain Data
A	B0006, B0007	B0005
B	B0005, B0007	B0006
C	B0005, B0006	B0007
D	CALCE_CS_2_36, CALCE_CS_2_37	CALCE_CS_2_35
E	CALCE_CS_2_35, CALCE_CS_2_37	CALCE_CS_2_36
F	CALCE_CS_2_35, CALCE_CS_2_36	CALCE_CS_2_37

The capacity curves of these two datasets are shown in Fig. 6 and the SOH of LIBs is defined as the following equation:

$$\text{SOH} = \frac{C_c}{C_{\text{new}}} \tag{15}$$

where C_c and C_{new} are the rated and current capacity, respectively. The degradation features in our study are obtained by using the voltage curve from the CC charging process as inputs for the model. As shown in Fig. 7, we extracted nine features from the geometrical analysis of the CC charging curve. Feature F1 is defined as the initial voltage, while the remaining features are derived from the subsequent voltage curve, which is segmented in a 2:3:5 ratio based on voltage values. As the battery ages, its capacity would diminish, and the time intervals required to reach the same voltage would be decreased. Therefore, features F_2–F_5 capture the time stamps of specific voltage points. Based on Faraday's law, the magnetic flux is the integral of voltage over time, which represents the power supply capability for a given system. The trapezoidal rule is applied to approximate features F_6–F_9.

We evaluated our method across six transfer learning tasks, each involving training on the source domain and testing on the target domain to assess cross-domain generalization. Each dataset was partitioned by selecting the complete degradation trajectories of two batteries as the source domain and one battery as the target domain. The detailed configuration of these transfer tasks is summarized in Table II.

TABLE III
EXPERIMENTAL INFORMATION OF THE TRANSFER LEARNING TASKS UNDER PRACTICAL CHARGING CONDITIONS

Task	Source Domain Data	Target Domain Data
G	B01, B02	B03
H	B02, B03	B01
I	B01, B03	B02

To further validate the practicality of our method under real-world partial charging conditions, we additionally employ the Xi'an Jiaotong University (XJTU) battery dataset processed by [47]. This dataset contains data from 6 batches of Lithium Nickel Manganese Cobalt Oxide (NMC) batteries, a chemistry prevalent in modern EVs and ESS. From the first batch of this dataset, we select three representative batteries for experiments. The health features for these batteries are adopted directly from the benchmark study [47], where they are derived not from full charging curves but from specific, short voltage intervals. These three batteries are then used to construct three challenging cross-domain transfer tasks, as shown in Table III, providing a rigorous evaluation of our method's robustness.

To perform fine-grained subdomain alignment, we apply FCM clustering to the extracted degradation features. In our experiments, the number of clusters (i.e., subdomains) is empirically set to three, corresponding to the early, middle, and late stages of battery degradation. Each sample is assigned to a subdomain based on its highest membership value, while the complete membership vector is retained as soft subdomain information. These subdomain labels are used to guide the subdomain adaptation process, allowing the model to better align local degradation dynamics between source and target domains.

To quantitatively evaluate the performance of the proposed method, we adopt three commonly used regression metrics: Root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_t^i - S_p^i)^2} \tag{16}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|S_t^i - S_p^i\right| \tag{17}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(S_t^i - S_p^i)^2}{\sum_{i=1}^{n}\left(S_t^i - \sum_{i=1}^{n} S_p^i\right)^2} \tag{18}$$

where S_t^i is the true SOH value of ith sample, and S_p^i is the predicted SOH value of ith sample. Meanwhile, hyperparameter tuning was performed to determine the optimal values for α and β, which control the weights of the fine-grained subdomain alignment loss and the domain adversarial loss, respectively. A grid search was conducted over the following ranges: $\alpha \in \{0.001, 0.01, 0.1, 1, 10\}$, $\beta \in \{0.1, 1, 10, 100, 1000\}$. The average root mean square error (RMSE) across all six cross-domain transfer tasks was used as the selection criterion. Fig. 8 illustrates the model's sensitivity to these hyperparameters, and the results provide empirical insight into managing the tradeoffs between the global (DANN) and subdomain (FCLMMD) adaptation objectives. As shown in figures, both very small and very large values for α and β lead to increased RMSE. Low values (e.g., $\alpha = 0.001$) result in insufficient domain alignment, causing underadaptation. Conversely, excessively high values (e.g., $\alpha = 10$ or $\beta = 1000$) can cause the adaptation objectives to dominate the loss function. This creates an optimization conflict, forcing the feature extractor to prioritize domain invariance at the expense of learning features discriminative for the primary SOH prediction task, a phenomenon known as negative transfer.

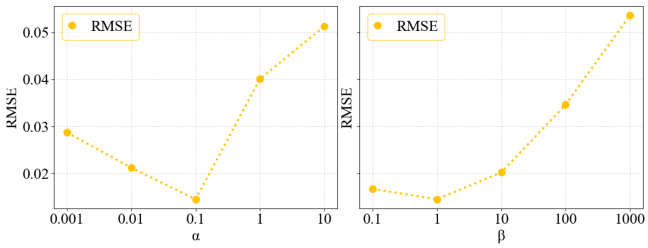

Fig. 8. Hyperparameter sensitivity test of our framework.

Based on this analysis, we selected $\alpha = 0.1$ and $\beta = 1$ as the optimal values for all subsequent experiments. These values represent the empirically found sweet spot that best balances the adaptation and prediction objectives, thereby maximizing the knowledge transfer while minimizing optimization conflicts.

B. Validation of Fine-Grained Domain Adaptation Strategy

To investigate the effectiveness of proposed transfer learning strategies and various loss components on cross-domain SOH estimation, we conducted experiments by comparing our method with five methods with different loss configurations. The first method $Loss_M$ just includes the prediction loss and serves as a baseline without any transfer learning strategies. The second method $Loss_D$ adds the domain adversarial training to encourage domain-invariant feature learning. The third method $Loss_F$ includes the prediction loss and the loss item based on FCLMMD to achieve fine-grained domain adaptation. The last method $Loss_{DM}$ integrates standard MMD with adversarial loss to jointly align global feature distributions. The final method $Loss_{DLM}$ combines the prediction loss, adversarial loss, and LMMD, aiming to simultaneously enforce domain-invariant representation learning and classwise conditional distribution alignment.

From the comparison results in Fig. 9, we observe several clear trends that underscore the importance of domain adaptation in cross-domain SOH estimation.

First, the baseline model $Loss_M$, trained only with prediction loss, performs significantly worse across all metrics. This confirms its limited capacity to generalize across distribution shifts and highlights the fundamental necessity of domain adaptation techniques for this task.

Second, when comparing the transfer learning strategies, our complete framework consistently achieves the best performance. This superiority stems from our method's unique ability to perform stage-aware alignment that is tailored for continuous degradation processes. While both $Loss_{DM}$ and $Loss_{DLM}$

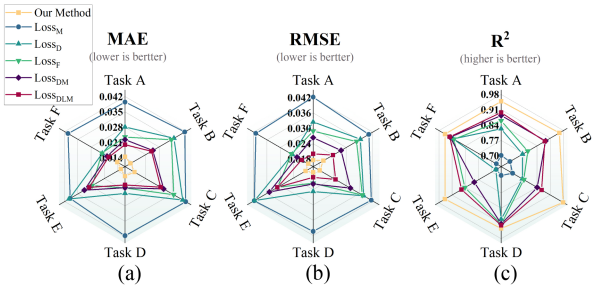

Fig. 9. Performance comparison under different loss configurations on cross-domain SOH estimation tasks. (a) MAE. (b) RMSE. (c) R^2.

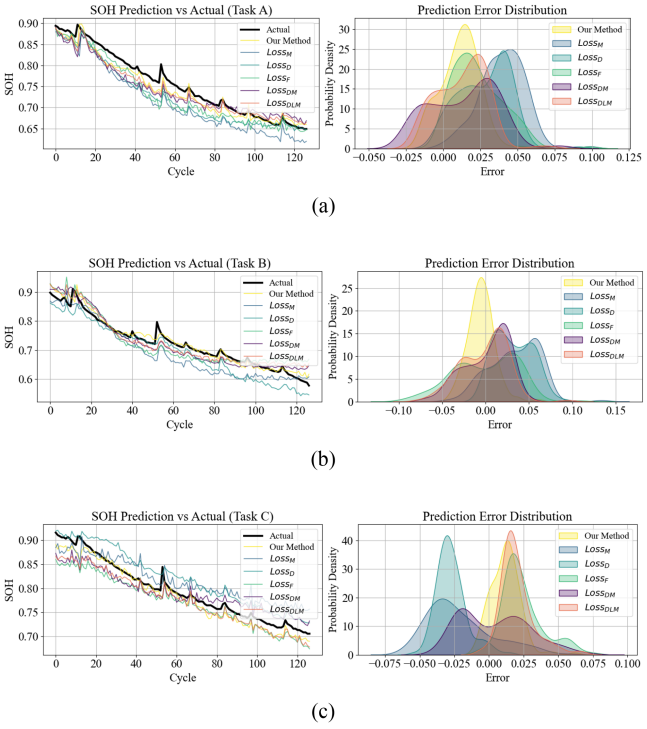

Fig. 10. Prediction curves and error curves of the comparison results between different loss items for NASA Dataset. (a) Task A. (b) Task B. (c) Task C.

outperform simple DANN, they have inherent limitations. The $Loss_{DM}$ approach, which combines two global alignment metrics, still fails to address the critical issue of matching fine-grained degradation stages. The $Loss_{DLM}$ method improves upon this by incorporating LMMD for subdomain alignment, yet its reliance on hard, discrete class assignments is a poor fit for the continuous and gradual nature of battery aging. Data points in transitional phases are often misaligned. In contrast, our FCLMMD's use of soft, fuzzy memberships accurately models these ambiguous stage boundaries, leading to a smoother and more physically meaningful alignment that ultimately yields higher accuracy.

The visual results in Figs. 10 and 11 underscore these quantitative findings. Our method's predicted SOH values track the ground truth more closely, and its error distributions are markedly more concentrated around zero. For a real-world

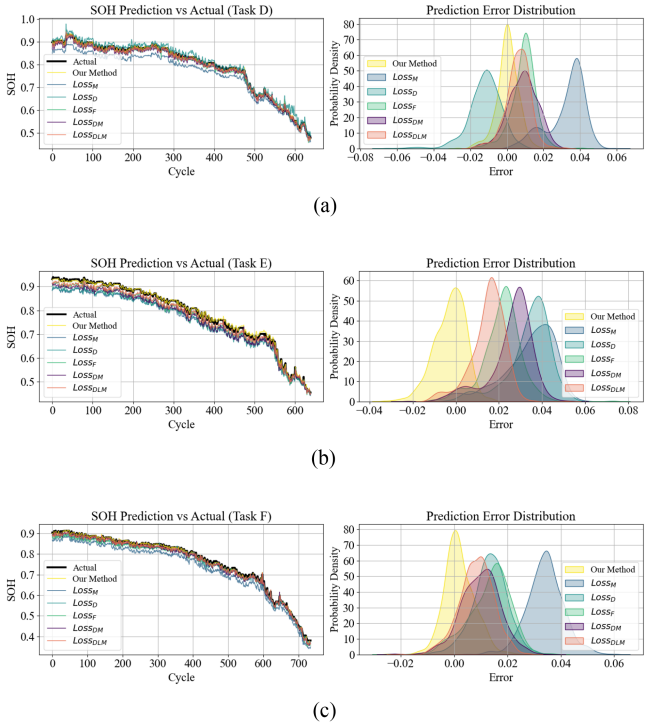

Fig. 11. Prediction curves and error curves of the comparison results between different loss items for CALCE Dataset. (a) Task D. (b) Task E. (c) Task F.

BMS, this implies a significant increase in reliability. A consistently low error minimizes the risk of dangerously inaccurate estimates, which is critical for downstream tasks like remaining useful life (RUL) prediction and ensuring operational safety.

C. Comparisons With Different Model Architecture

In this section, to evaluate the effectiveness of the proposed PatchTSCT, we compare it against four baseline architectures, including standalone CNN and transformer models, as well as two sequential hybrid architectures: CNN-LSTM and CNN-Transformer. Notably, the CNN-Transformer baseline can be viewed as a variant of our method without the patch-based representation module. This comparison allows us to isolate the contribution of the patch mechanism in enhancing temporal feature modeling and overall prediction performance.

Table IV presents the comparative performance, where our proposed PatchTSCT architecture consistently achieves state-of-the-art results across all tasks. This superior performance, further visualized in Figs. 12 and 13, stems directly from its specialized design for temporal feature modeling.

The analysis reveals two key insights. First, the patch-based representation is crucial for capturing local dynamics. Standalone CNN and transformer architectures exhibit broader error spreads because their pointwise or global receptive fields often miss the subtle, fine-grained patterns inherent in battery degradation. Then, while hybrid models like CNN-LSTM and CNN-Transformer improve upon single architectures, our full

TABLE IV
COMPARISON RESULTS OF DIFFERENT MODEL STRUCTURES

Task	Method	RMSE	MAE	R2
A	CNN	0.0328	0.0302	0.8193
	Transformer	0.0387	0.0334	0.7489
	CNN-LSTM	0.0137	0.0158	0.9591
	CNN-Transformer	0.0246	0.0204	0.8984
	Our Method	**0.0175**	**0.0145**	**0.9484**
B	CNN	0.0360	0.0290	0.8225
	Transformer	0.0439	0.0400	0.7362
	CNN-LSTM	0.0334	0.0279	0.8475
	CNN-Transformer	0.0296	0.0246	0.8802
	Our Method	**0.0162**	**0.0125**	**0.9642**
C	CNN	0.0289	0.0263	0.7844
	Transformer	0.0294	0.0254	0.7772
	CNN-LSTM	0.0244	0.0190	0.8461
	CNN-Transformer	0.0217	0.0189	0.8782
	Our Method	**0.0160**	**0.0129**	**0.9342**
D	CNN	0.0130	0.0093	0.9594
	Transformer	0.0178	0.0120	0.9457
	CNN-LSTM	0.0102	0.0071	0.9751
	CNN-Transformer	0.0088	0.0079	0.9816
	Our Method	**0.0035**	**0.0025**	**0.9971**
E	CNN	0.0144	0.0111	0.9614
	Transformer	0.0162	0.0137	0.9510
	CNN-LSTM	0.0105	0.0083	0.9796
	CNN-Transformer	0.0089	0.0067	0.9852
	Our Method	**0.0043**	**0.0033**	**0.9965**
F	CNN	0.0127	0.0096	0.9724
	Transformer	0.0179	0.0159	0.9232
	CNN-LSTM	0.0095	0.0059	0.9844
	CNN-Transformer	0.0083	0.0053	0.9882
	Our Method	**0.0035**	**0.0026**	**0.9979**

PatchTSCT model still demonstrates superior estimation performance. This is because the patch-based input forces the network to learn robust local features first, a more effective strategy than the sequential feature passing in standard hybrids. The result is a more concentrated error distribution, as shown in the right panels of the figures.

This architectural advantage has significant implications for real-world BMS applications. An architecture adept at extracting meaningful patterns from smaller data segments, like PatchTSCT, is inherently more robust to the fragmented and imperfect data streams caused by noisy sensors and partial charging cycles. This leads to more reliable and consistent SOH estimation under nonideal operational conditions.

Beyond prediction accuracy, a critical aspect for practical application is the model's computational efficiency, which determines its feasibility for deployment on resource-constrained BMS hardware. To provide a thorough evaluation, we adopt several indicators: floating-point operations (FLOPs), training time, the number of parameters, and storage size. In this context, FLOPs refers to the total count of FLOPs required for a single forward pass (presented in millions in Table V), directly reflecting the network's computational workload. Training time measures the model's overall efficiency during the offline learning phase. The number of parameters and the derived storage size from the model file indicate the actual memory footprint required on the device.

As summarized in Table V, the proposed FCSA-PatchTSCT achieves an excellent balance between performance and efficiency. While it is naturally more resource-intensive than standalone CNN or Transformer models, it demonstrates a

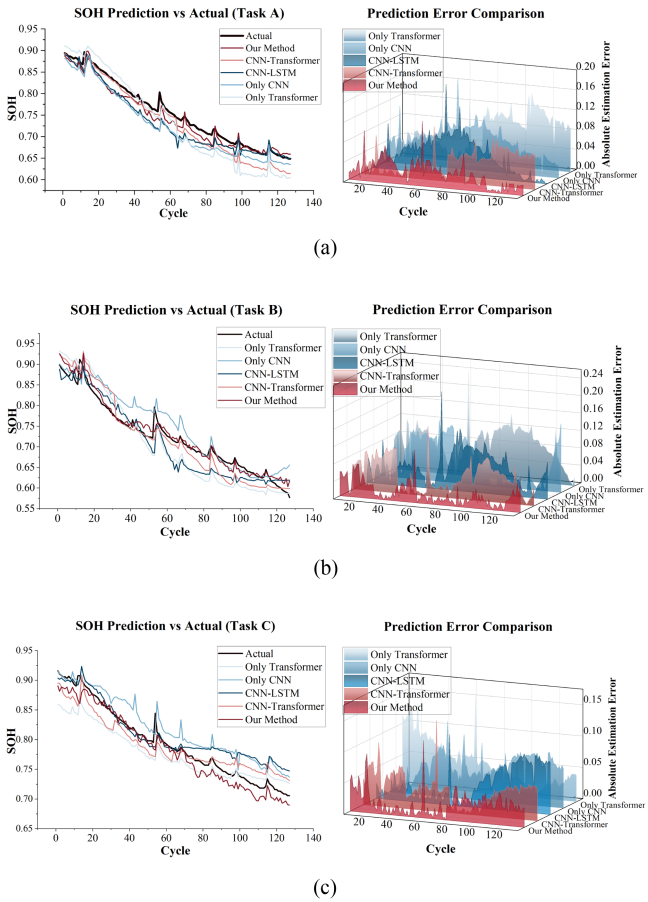

Fig. 12. Prediction curves and error curves of the comparison results between the proposed method and other network structures for NASA Dataset. (a) Task A. (b) Task B. (c) Task C.

TABLE V
COMPREHENSIVE EVALUATION OF THE PROPOSED METHOD

Methods	Performance index			
	FLOPs (Million)	Training time (min)	Parameters	Storage (MB)
CNN	0.614	4.549	50471	0.179
Transformer	1.399	11.848	109441	0.433
CNN-LSTM	1.021	9.087	110977	0.430
CNN-Transformer	1.998	16.048	176193	0.689
Our Method	1.546	14.781	163252	0.638

crucial advantage over the conventional hybrid architectures. Benefiting from the patch-based representation which effectively reduces redundant temporal dependencies, our framework is not only more accurate but also significantly more efficient than the standard CNN-Transformer baseline, requiring fewer FLOPs and a shorter training time. This efficiency gain is a direct result of the patch-based design, which mitigates the Transformer's inherent quadratic complexity. By allowing the self-attention mechanism to operate on a short sequence of patches instead of a long sequence of individual time points, both memory and computational demands are substantially reduced.

Most importantly, the practicality for deployment must be viewed in the context of the specific task. SOH is a slow-varying parameter, with BMS updates typically required on the scale

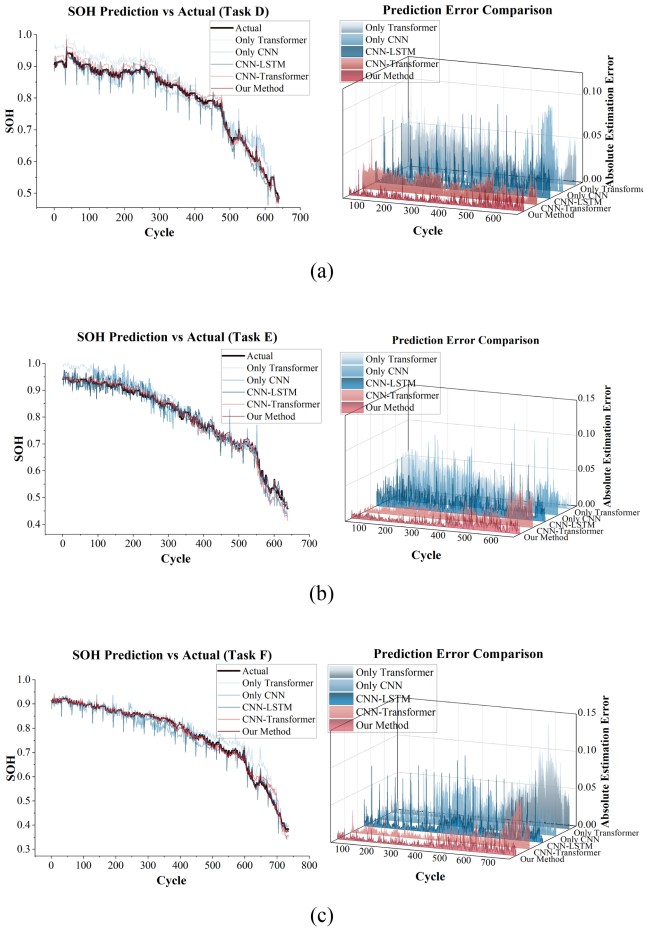

Fig. 13. Prediction curves and error curves of the comparison results between the proposed method and other network structures for CALCE Dataset. (a) Task D. (b) Task E. (c) Task F.

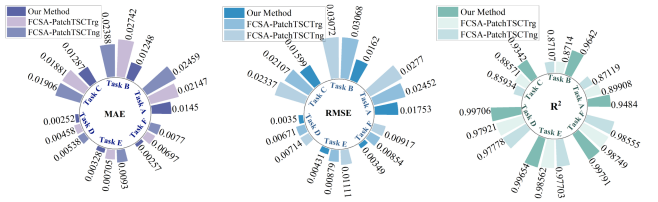

Fig. 14. Performance comparison under different group method on cross-domain SOH estimation tasks.

of seconds or even minutes, not milliseconds. Therefore, our model's modest computational cost—1.546 million FLOPs per inference—translates to an estimated latency well within the 8–16 ms range on a typical automotive-grade microcontroller. This is several orders of magnitude faster than the required SOH update interval. While classical methods like equivalent circuit models (ECMs) and Kalman filters are computationally simpler, their reliance on precise physical models often limits their accuracy and generalization across diverse operating conditions. Our framework, in contrast, accepts a manageable increase in deployment cost for a substantial gain in prediction accuracy and robustness to domain shifts. This analysis confirms that the proposed framework achieves a more favorable tradeoff between prediction performance and computational demand, establishing it as a robust and practical solution for embedded SOH estimation tasks.

While our analysis demonstrates feasibility, further optimization can ensure scalability across a wider range of BMS hardware. Several TinyML techniques are well-suited for this purpose. For instance, model quantization can convert the model to an 8-bit integer (INT8) representation, significantly reducing its memory footprint and accelerating inference on compatible hardware. Additionally, network pruning can reduce model complexity by removing redundant parameters, while knowledge distillation allows the transfer of knowledge from our larger model to a much smaller, highly efficient student network designed for on-device execution. These optimization strategies, implemented via frameworks like TensorFlow Lite for microcontrollers, form a clear roadmap for future work, aiming to deliver our framework's high accuracy in a power and memory-efficient package for real-world BMS.

D. Validation of Homogeneous Feature Grouping Method

To validate the effectiveness of homogeneous feature grouping strategy, comparison experiments with no homogeneous feature grouping strategy and different grouping strategy are conducted. Different grouping strategy are listed as follows:

1) $FCSA-PatchTSCT_{ng}$, which processes all features collectively without grouping and relies solely on the attention mechanism to model dependencies;
2) $FCSA-PatchTSCT_{rg}$, which involves randomly partitioning features into groups.
3) Proposed grouping strategy, which divides features into homogeneous groups based on Pearson correlation coefficients.

All configurations share identical model architecture and hyperparameters.

As illustrated in Fig. 14, the results clearly demonstrate the critical importance of feature grouping. The baseline model, which processes all features collectively, performs poorly. This suggests that tasking a single network with simultaneously resolving feature heterogeneity and modeling temporal patterns leads to suboptimal learning due to interference from conflicting feature trends. Our explicit grouping strategy simplifies this complex task, allowing the network to learn more effectively.

Furthermore, our proposed Pearson correlation-based grouping consistently outperforms the random grouping approach. While random grouping provides some benefit by separating features, it does not guarantee dynamic coherence within each group. In contrast, our correlation-based method ensures that each parallel channel of the PatchTSCT architecture processes only temporally similar features. This targeted design allows the multichannel architecture to function as intended, leading to a more accurate and robust SOH estimation.

This finding has significant practical implications. Real-world battery degradation involves complex and evolving relationships between health indicators, which can exhibit highly correlated

TABLE VI
COMPARISON RESULTS OF DIFFERENT UNSUPERVISED DA METHODS

Task	Method	RMSE	MAE	R2
A	CTSGAN	0.0213	0.0179	0.9257
	AT-GPR	0.0231	0.0187	0.9131
	VarLSTM-TL	0.0199	0.0177	0.9352
	DDAN	0.0208	0.0167	0.9292
	MTR-GPR	0.0261	0.0203	0.8886
	Our Method	**0.0175**	**0.0145**	**0.9484**
B	CTSGAN	0.0234	0.0177	0.9284
	AT-GPR	0.0251	0.0218	0.9190
	VarLSTM-TL	0.0200	0.0157	0.9479
	DDAN	0.0216	0.0175	0.9391
	MTR-GPR	0.0229	0.0195	0.9316
	Our Method	**0.0162**	**0.0125**	**0.9642**
C	CTSGAN	0.0195	0.0182	0.9049
	AT-GPR	0.0175	0.0152	0.9237
	VarLSTM-TL	0.0166	0.0132	0.9314
	DDAN	0.0184	0.0168	0.9157
	MTR-GPR	0.0179	0.0159	0.9201
	Our Method	**0.0160**	**0.0129**	**0.9342**
D	CTSGAN	0.0110	0.0076	0.9782
	AT-GPR	0.0081	0.0067	0.9881
	VarLSTM-TL	0.0121	0.0102	0.9724
	DDAN	0.0088	0.0073	0.9861
	MTR-GPR	0.0072	0.0055	0.9907
	Our Method	**0.0035**	**0.0025**	**0.9971**
E	CTSGAN	0.0133	0.0099	0.9741
	AT-GPR	0.0085	0.0070	0.9895
	VarLSTM-TL	0.0094	0.0081	0.9853
	DDAN	0.0143	0.0108	0.9681
	MTR-GPR	0.0108	0.0089	0.9829
	Our Method	**0.0043**	**0.0033**	**0.9965**
F	CTSGAN	0.0172	0.0115	0.9556
	AT-GPR	0.0088	0.0063	0.9882
	VarLSTM-TL	0.0183	0.0127	0.9495
	DDAN	0.0110	0.0076	0.9816
	MTR-GPR	0.0122	0.0093	0.9781
	Our Method	**0.0035**	**0.0026**	**0.9979**
G	CTSGAN	0.0092	0.0078	0.9659
	AT-GPR	0.0085	0.0075	0.9709
	VarLSTM-TL	0.0068	0.0057	0.9811
	DDAN	0.0083	0.0070	0.9720
	MTR-GPR	0.0069	0.0058	0.9805
	Our Method	**0.0054**	**0.0046**	**0.9882**
H	CTSGAN	0.0073	0.0062	0.9782
	AT-GPR	0.0068	0.0055	0.9839
	VarLSTM-TL	0.0094	0.0083	0.9688
	DDAN	0.0062	0.0054	0.9866
	MTR-GPR	0.0093	0.0082	0.9696
	Our Method	**0.0053**	**0.0038**	**0.9901**
I	CTSGAN	0.0096	0.0088	0.9697
	AT-GPR	0.0084	0.0072	0.9728
	VarLSTM-TL	0.0075	0.0059	0.9783
	DDAN	0.0072	0.0052	0.9828
	MTR-GPR	0.0092	0.0086	0.9711
	Our Method	**0.0057**	**0.0047**	**0.9891**

TABLE VII
COMPARISON RESULTS OF DIFFERENT SUPERVISED DL METHODS

Task	Method	RMSE	MAE	R2
A	TCN-transformer	0.0217	0.0181	0.9231
	SHMM-transformer-BiGRU	0.0223	0.0182	0.9187
	BC-PINN	0.0220	0.0178	0.9210
	PINN-LTB	0.0193	0.0162	0.9388
	Our Method	**0.0175**	**0.0145**	**0.9484**
B	TCN-transformer	0.0201	0.0168	0.9473
	SHMM-transformer-BiGRU	0.0218	0.0171	0.9381
	BC-PINN	0.0210	0.0172	0.9420
	PINN-LTB	0.0193	0.0158	0.9512
	Our Method	**0.0162**	**0.0125**	**0.9642**
C	TCN-transformer	0.0184	0.0166	0.9157
	SHMM-transformer-BiGRU	0.0179	0.0159	0.9201
	BC-PINN	0.0173	0.0163	0.9253
	PINN-LTB	0.0169	0.0159	0.9290
	Our Method	**0.0160**	**0.0129**	**0.9342**
D	TCN-transformer	0.0089	0.0066	0.9859
	SHMM-transformer-BiGRU	0.0071	0.0057	0.9909
	BC-PINN	0.0111	0.0092	0.9830
	PINN-LTB	0.0065	0.0043	0.9923
	Our Method	**0.0035**	**0.0025**	**0.9971**
E	TCN-transformer	0.0101	0.0081	0.9851
	SHMM-transformer-BiGRU	0.0109	0.0084	0.9828
	BC-PINN	0.0129	0.0102	0.9758
	PINN-LTB	0.0081	0.0066	0.9903
	Our Method	**0.0043**	**0.0033**	**0.9965**
F	TCN-transformer	0.0127	0.0086	0.9757
	SHMM-transformer-BiGRU	0.0098	0.0069	0.9853
	BC-PINN	0.0123	0.0084	0.9770
	PINN-LTB	0.0065	0.0045	0.9934
	Our Method	**0.0035**	**0.0026**	**0.9979**
G	TCN-transformer	0.0081	0.0070	0.9736
	SHMM-transformer-BiGRU	0.0067	0.0059	0.9816
	BC-PINN	0.0076	0.0068	0.9764
	PINN-LTB	0.0070	0.0059	0.9804
	Our Method	**0.0054**	**0.0046**	**0.9882**
H	TCN-transformer	0.0095	0.0084	0.9685
	SHMM-transformer-BiGRU	0.0089	0.0078	0.9722
	BC-PINN	0.0078	0.0067	0.9785
	PINN-LTB	0.0072	0.0063	0.9817
	Our Method	**0.0053**	**0.0038**	**0.9901**
I	TCN-transformer	0.0110	0.0103	0.9598
	SHMM-transformer-BiGRU	0.0097	0.0090	0.9687
	BC-PINN	0.0077	0.0072	0.9803
	PINN-LTB	0.0069	0.0057	0.9834
	Our Method	**0.0057**	**0.0047**	**0.9891**

or even opposing trends. A strategy that intelligently groups features by their dynamic behavior, rather than processing them monolithically, is therefore inherently more effective. It allows the model to better capture the true underlying degradation dynamics, making it more robust and adaptable throughout the battery's lifecycle.

E. Comparative Analysis With State-of-Art Methods

To provide a comprehensive validation, we benchmarked our framework against both state-of-the-art unsupervised DA methods and advanced supervised DL models.

1) Comparison With Unsupervised DA Methods: In this section, our proposed method is compared with some state-of-the-art unsupervised DA methods, including Lu's conditional time series generative adversarial network (CTSGAN) [48], Mondal's adaptive transfer Gaussian processes regression (AT-GPR) [49], Kim's variational LSTM with transfer learning (VarLSTM-TL) [50], Ni's deep domain adaptation network (DDAN) [24], and Sheng's multidomain transfer Gaussian process regression (MTR-GPR) [51]. As shown in Table VI, our method consistently and significantly outperformed all these baseline models across all nine transfer tasks.

This superior performance stems from the synergistic effect of our framework's three core components. First, the fine-grained subdomain alignment (FCLMMD) enables more precise knowledge transfer by aligning corresponding degradation stages,

which is critical under variable operating conditions. Second, the patch-based temporal modeling (PatchTSCT) excels at capturing robust local patterns even from fragmented data segments. Finally, the homogeneous feature grouping minimizes interference from conflicting feature trends. This integrated approach allows our model to learn a more robust and accurate representation of battery health from the challenging, incomplete data. These results also validate that our method is not reliant on full charging curve data; the strong performance on the XJTU dataset confirms the framework's effectiveness in handling the partial and irregular charging cycles commonly encountered in practical scenarios.

2) Comparison with supervised DL methods: To further contextualize this state-of-the-art unsupervised performance, we conducted a demanding benchmark against five advanced supervised DL models. These baselines were selected to represent recent advancements and include transformer variants, namely temporal convolutional network (TCN) with a transformer (TCN-transformer) [52] and segmented hidden Markov model transformer-BiGRU (SHMM-transformer-BiGRU) [53], as well as Physics-Informed Neural Networks (PINNs), such as Bayesian calibrated PINN (BC-PINN) [54] and PINN-Lasso-Transformer-BiLSTM (PINN-LTB) [55]. It is critical to note the training protocol: these supervised baselines were trained on the full source domain dataset and were additionally fine-tuned on the first 20% of labeled data from the target domain—a significant informational advantage.

The results of this benchmark are presented in Table VII. The data reveals that our unsupervised framework demonstrates exceptional performance, even when compared against these supervised models. Despite the informational advantage held by the baselines, our proposed method outperforms the supervised models in a majority of tasks and achieves highly comparable accuracy in the remaining tasks. This outcome underscores the high data efficiency of our framework, an advantage rooted in the synergy of its core components.

Our framework's effectiveness, unlike standard fine-tuning, stems from a more synergistic knowledge transfer mechanism. The core FCLMMD strategy performs a fine-grained alignment, leveraging the full source domain to compensate for the limited information in the target's lifecycle. This alignment is built upon a robust feature foundation, where the PatchTSCT network extracts high-quality local features and homogeneous feature grouping minimizes interference from conflicting trends. This integrated approach results in a far more data-efficient and robust transfer learning pipeline, allowing it to attain a level of accuracy that rivals or surpasses models requiring explicit target domain labels. This capability is of significant practical value, as it validates the framework's effectiveness in scenarios where labeled data for new batteries is unavailable or impractical to obtain.

IV. Conclusion

In this article, we proposed FCSA-PatchTSCT, a novel cross-domain framework for LIBs' SOH estimation. The proposed framework integrates three key components to address the challenges of domain shift, temporal modeling, and feature heterogeneity in real-world SOH estimation tasks.

First, we introduced a fine-grained subdomain adaptation strategy based on FCLMMD, which divides the original data into multiple fuzzy subdomains and aligns their distributions. This enables more precise and localized knowledge transfer across domains. Second, we designed the PatchTSCT network, a patch-based time series network architecture that segments univariate sequences into subseries-level patches. This representation enhances model's ability to extract local temporal characteristics, and improves feature learning capacity compared to conventional pointwise input methods. Finally, we proposed a homogeneous feature grouping mechanism that leverages Pearson correlation to cluster variables and independently process them in parallel channels. This reduces interference among heterogeneous features and facilitates more effective feature extraction. Extensive experiments on multiple public datasets, including scenarios with partial charging data, demonstrate that our framework consistently outperforms state-of-the-art baselines in both prediction accuracy and generalization ability. The ablation studies further validate the effectiveness of each proposed module, confirming its robustness and practical applicability for real-world BMS.

While FCSA-PatchTSCT performs strongly, its reliance on domain similarity and current evaluation scope present limitations. Future work will enhance robustness and scalability for large-scale systems and greater domain shifts. Key directions include strategies for leveraging multiple source domains, particularly with scarce or distinct target data. Furthermore, we will focus on optimizing the framework for on-device deployment using TinyML techniques and validating its performance on large-scale datasets to advance practical applicability, aiming for more reliable, adaptable battery management solutions.

References

[1] G. Vennam, A. Sahoo, and S. Ahmed, "A survey on lithium-ion battery internal and external degradation modeling and state of health estimation," *J. Energy Storage*, vol. 52, 2022, Art. no. 104720.

[2] L. Feng, L. Jiang, J. Liu, Z. Wang, Z. Wei, and Q. Wang, "Dynamic overcharge investigations of lithium ion batteries with different state of health," *J. Power Sources*, vol. 507, 2021, Art. no. 230262.

[3] T. Han, J. Tian, C. Chung, and Y.-M. Wei, "Challenges and opportunities for battery health estimation: Bridging laboratory research and real-world applications," *J. Energy Chem.*, vol. 89, pp. 434–436, 2024.

[4] M. Jiang et al., "Advances in battery state estimation of battery management system in electric vehicles," *J. Power Sources*, vol. 612, 2024, Art. no. 234781.

[5] Z. Deng, X. Hu, P. Li, X. Lin, and X. Bian, "Data-driven battery state of health estimation based on random partial charging data," *IEEE Trans. Power Electron.*, vol. 37, no. 5, pp. 5021–5031, May 2022.

[6] G. Nuroldayeva, Y. Serik, D. Adair, B. Uzakbaiuly, and Z. Bakenov, "State of health estimation methods for lithium-ion batteries," *Int. J. Energy Res.*, vol. 2023, no. 1, 2023, Art. no. 4297545.

[7] S. Tong, M. P. Klein, and J. W. Park, "On-line optimization of battery open circuit voltage for improved state-of-charge and state-of-health estimation," *J. Power Sources*, vol. 293, pp. 416–428, 2015.

[8] Y. Fu, J. Xu, M. Shi, and X. Mei, "A fast impedance calculation-based battery state-of-health estimation method," *IEEE Trans. Ind. Electron.*, vol. 69, no. 7, pp. 7019–7028, Jul. 2022.

[9] S. Amir, M. Gulzar, M. O. Tarar, I. H. Naqvi, N. A. Zaffar, and M. G. Pecht, "Dynamic equivalent circuit model to estimate state-of-health of lithium-ion batteries," *IEEE Access*, vol. 10, pp. 18279–18288, 2022.

[10] Y. Gao, K. Liu, C. Zhu, X. Zhang, and D. Zhang, "Co-estimation of state-of-charge and state-of-health for lithium-ion batteries using an enhanced electrochemical model," *IEEE Trans. Ind. Electron.*, vol. 69, no. 3, pp. 2684–2696, Mar. 2022.

[11] Y. Liu et al., "A nonlinear observer SOC estimation method based on electrochemical model for lithium-ion battery," *IEEE Trans. Ind. Appl.*, vol. 57, no. 1, pp. 1094–1104, Jan./Feb. 2021.

[12] Z. Huang, M. Best, J. Knowles, and A. Fly, "Adaptive piecewise equivalent circuit model with SOC/SOH estimation based on extended Kalman filter," *IEEE Trans. Energy Convers.*, vol. 38, no. 2, pp. 959–970, Jun. 2023.

[13] G. Vennam and A. Sahoo, "A dynamic SOH-coupled lithium-ion cell model for state and parameter estimation," *IEEE Trans. Energy Convers.*, vol. 38, no. 2, pp. 1186–1196, Jun. 2023.

[14] X. Gu et al., "A novel state-of-health estimation for the lithium-ion battery using a convolutional neural network and transformer model," *Energy*, vol. 262, 2023, Art. no. 125501.

[15] X. Feng et al., "Online state-of-health estimation for Li-ion battery using partial charging segment based on support vector machine," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8583–8592, Sep. 2019.

[16] D. Yang, Y. Wang, R. Pan, R. Chen, and Z. Chen, "State-of-health estimation for the lithium-ion battery based on support vector regression," *Appl. Energy*, vol. 227, pp. 273–283, 2018.

[17] K. S. Mawonou, A. Eddahech, D. Dumur, D. Beauvois, and E. Godoy, "State-of-health estimators coupled to a random forest approach for lithium-ion battery aging factor ranking," *J. Power Sources*, vol. 484, 2021, Art. no. 229154.

[18] R. Pan, T. Liu, W. Huang, Y. Wang, D. Yang, and J. Chen, "State of health estimation for lithium-ion batteries based on two-stage features extraction and gradient boosting decision tree," *Energy*, vol. 285, 2023, Art. no. 129460.

[19] M. Lin, Y. You, J. Meng, W. Wang, J. Wu, and D.-I. Stroe, "Lithium-ion batteries SOH estimation with multimodal multilinear feature fusion," *IEEE Trans. Energy Convers.*, vol. 38, no. 4, pp. 2959–2968, Dec. 2023.

[20] M. Jiao, D. Wang, and J. Qiu, "A GRU-RNN based momentum optimized algorithm for SOC estimation," *J. Power Sources*, vol. 459, 2020, Art. no. 228051.

[21] Y. Ma, C. Shan, J. Gao, and H. Chen, "A novel method for state of health estimation of lithium-ion batteries based on improved LSTM and health indicators extraction," *Energy*, vol. 251, 2022, Art. no. 123973.

[22] L. Chen, S. Xie, A. M. Lopes, and X. Bao, "A vision transformer-based deep neural network for state of health estimation of lithium-ion batteries," *Int. J. Elect. Power Energy Syst.*, vol. 152, 2023, Art. no. 109233.

[23] K. Luo, H. Zheng, and Z. Shi, "A simple feature extraction method for estimating the whole life cycle state of health of lithium-ion batteries using transformer-based neural network," *J. Power Sources*, vol. 576, 2023, Art. no. 233139.

[24] Z. Ni, B. Li, and Y. Yang, "Deep domain adaptation network for transfer learning of state of charge estimation among batteries," *J. Energy Storage*, vol. 61, 2023, Art. no. 106812.

[25] L. Song, X. Gui, J. Du, Z. Fan, M. Li, and L. Guo, "A novel transfer learning approach for state-of-health prediction of lithium-ion batteries in the absence of run to failure data," *IEEE Trans. Instrum. Meas.*, vol. 73, 2024, Art. no. 2529712.

[26] S. Shen, M. Sadoughi, M. Li, Z. Wang, and C. Hu, "Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries," *Appl. Energy*, vol. 260, 2020, Art. no. 114296.

[27] K. Huang, K. Yao, Y. Guo, and Z. Lv, "State of health estimation of lithium-ion batteries based on fine-tuning or rebuilding transfer learning strategies combined with new features mining," *Energy*, vol. 282, 2023, Art. no. 128739.

[28] G. Ma et al., "A transfer learning-based method for personalized state of health estimation of lithium-ion batteries," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 1, pp. 759–769, Jan. 2024.

[29] J. Li, Z. Yu, Z. Du, L. Zhu, and H. T. Shen, "A comprehensive survey on source-free domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 8, pp. 5743–5762, Aug. 2024.

[30] X. Ma, T. Zhang, and C. Xu, "GCAN: Graph convolutional adversarial network for unsupervised domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 8258–8268.

[31] A. Schrab, I. Kim, M. Albert, B. Laurent, B. Guedj, and A. Gretton, "MMD aggregated two-sample test," *J. Mach. Learn. Res.*, vol. 24, no. 194, pp. 1–81, 2023.

[32] T. Han, Z. Wang, and H. Meng, "End-to-end capacity estimation of lithium-ion batteries with an enhanced long short-term memory network considering domain adaptation," *J. Power Sources*, vol. 520, 2022, Art. no. 230823.

[33] J. Meng, D. Hu, M. Lin, J. Peng, J. Wu, and D.-I. Stroe, "A domain-adversarial neural network for transferable lithium-ion battery state of health estimation," *IEEE Trans. Transp. Electrific.*, vol. 11, no. 3, pp. 7732–7742, Jun. 2025.

[34] J. Yao, Z. Chang, T. Han, and J. Tian, "Semi-supervised adversarial deep learning for capacity estimation of battery energy storage systems," *Energy*, vol. 294, 2024, Art. no. 130882.

[35] C. Liu, Z. Deng, X. Zhang, H. Bao, and D. Cheng, "Battery state of health estimation across electrochemistry and working conditions based on domain adaptation," *Energy*, vol. 297, 2024, Art. no. 131294.

[36] X. Qiu, Y. Bai, and S. Wang, "A novel unsupervised domain adaptation-based method for lithium-ion batteries state of health prognostic," *J. Energy Storage*, vol. 75, 2024, Art. no. 109358.

[37] Y. Ding, M. Jia, and Y. Cao, "Remaining useful life estimation under multiple operating conditions via deep subdomain adaptation," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 3516711.

[38] Y. Zhang, K. Feng, J. Ji, K. Yu, Z. Ren, and Z. Liu, "Dynamic model-assisted bearing remaining useful life prediction using the cross-domain transformer network," *IEEE/ASME Trans. Mechatron.*, vol. 28, no. 2, pp. 1070–1080, Apr. 2023.

[39] A. Mohan and A. Thosar, "RNN and CNN based ensemble models for state-of-health prediction of Li-ion batteries," in *Proc. IEEE Int. Conf. Intell. Syst., Smart Green Technol.*, 2024, pp. 128–132.

[40] J. Kim, H. Kim, H. Kim, D. Lee, and S. Yoon, "A comprehensive survey of deep learning for time series forecasting: Architectural diversity and open challenges," *Artif. Intell. Rev.*, vol. 58, no. 7, pp. 1–95, 2025.

[41] X. Chen, Y. Qin, W. Zhao, Q. Yang, N. Cai, and K. Wu, "A self-attention knowledge domain adaptation network for commercial lithium-ion batteries state-of-health estimation under shallow cycles," *J. Energy Storage*, vol. 86, 2024, Art. no. 111197.

[42] Q. Huang et al., "CrossGNN: Confronting noisy multivariate time series via cross interaction refinement," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, vol. 36, pp. 46885–46902.

[43] Z. Wei, X. Sun, Y. Li, W. Liu, C. Liu, and H. Lu, "A joint estimation method for the SOC and SOH of lithium-ion batteries based on AR-ECM and data-driven model fusion," *Electronics*, vol. 14, no. 7, 2025, Art. no. 1290.

[44] Y. Zhu et al., "Deep subdomain adaptation network for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 4, pp. 1713–1722, Apr. 2021.

[45] H. H. Goh, Z. Lan, D. Zhang, W. Dai, T. A. Kurniawan, and K. C. Goh, "Estimation of the state of health (SOH) of batteries using discrete curvature feature extraction," *J. Energy Storage*, vol. 50, 2022, Art. no. 104646.

[46] Y. Liu, J. Ding, L. Yao, H. Su, Y. Chen, and Z. Wang, "A novel high-accuracy intelligent estimation method for battery state of health," *Measurement*, vol. 245, 2025, Art. no. 116620.

[47] F. Wang, Z. Zhai, B. Liu, S. Zheng, Z. Zhao, and X. Chen, "Open access dataset, code library and benchmarking deep learning approaches for state-of-health estimation of lithium-ion batteries," *J. Energy Storage*, vol. 77, 2024, Art. no. 109884.

[48] X. Lu, J. Qiu, G. Lei, and J. Zhu, "State of health estimation of lithium iron phosphate batteries based on degradation knowledge transfer learning," *IEEE Trans. Transp. Electrific.*, vol. 9, no. 3, pp. 4692–4703, Sep. 2023.

[49] A. Mondal, A. Routray, and S. Puravankara, "State-of-health estimation of li-ion batteries using semiparametric adaptive transfer learning," *IEEE Trans. Transp. Electrific.*, vol. 10, no. 1, pp. 1080–1088, Mar. 2024.

[50] S. Kim, Y. Y. Choi, K. J. Kim, and J.-I. Choi, "Forecasting state-of-health of lithium-ion batteries using variational long short-term memory with transfer learning," *J. Energy Storage*, vol. 41, 2021, Art. no. 102893.

[51] H. Sheng, B. Ray, S. Kayamboo, X. Xu, and S. Wang, "Battery health estimation based on multidomain transfer learning," *IEEE Trans. Power Electron.*, vol. 39, no. 4, pp. 4758–4770, Apr. 2024.

[52] Y. Chen et al., "Exploring life warning solution of lithium-ion batteries in real-world scenarios: TCN-transformer fusion model for battery pack SOH estimation," *Energy*, 2025, Art. no. 138053.

[53] Z. Liu, Y. Liu, Y. Zhang, C. Wu, S. Zhang, and C. Sun, "Data-driven lithium-ion battery SOH prediction: A novel SHMM-transformer-BIGRU hybrid neural network method," *Measurement*, vol. 257, 2025, Art. no. 118579.

[54] R. Zhu, J. Hu, and W. Peng, "Bayesian calibrated physics-informed neural networks for second-life battery SOH estimation," *Rel. Eng. Syst. Saf.*, vol. 264, 2025, Art. no. 111432.

[55] S. Wang, R. Zhou, Y. Ren, H. Liu, Y. Lin, and C. Lian, "A generalizable physics-informed neural network for lithium-ion battery SOH estimation utilizing partial charging segments," *J. Energy Chem.*, vol. 112, pp. 977–986, 2025.

Xuanang Gui received the B.Eng. degree in computer science and technology from the Beijing University of Technology, Beijing, China, in 2021, and the M.Eng. degree in electronic and information engineering from the University of Chinese Academy of Sciences, Beijing, 2024. He is currently working toward the Ph.D. degree with the Chinese University of Hong Kong, Shenzhen, China.

His current research interests include artificial intelligence, deep learning, reinforcement learning, remaining useful life prediction, and energy management.

Shu Zhang received the B.E. degree in intelligent science and technology from Central South University, Changsha, China, in 2017 and the M.Sc. degree in information technology from the Hong Kong University of Science and Technology, Hong Kong, in 2018.

Her current research interests include large language models, time-series modeling, and state-of-health (SOH) estimation for lithium-ion batteries.

Yuheng Cheng received the B.S. degree in electronic engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2017, and the M.S. degree in computer science from the Hong Kong University of Science and Technology, Hong Kong, in 2018. He is currently working toward the Ph.D. degree in computer science with the Chinese University of Hong Kong (Shenzhen), Shenzhen, China.

His research focuses on natural language processing, large language models, power systems, and energy.

Qianlong Wang received the B.S. degree in logistics management from Nankai University, Tianjin, China, in 2021, and the M.S. degree in management science and engineering from the University of Chinese Academy of Sciences, Beijing, China, in 2024.

His research interests include combinatorial optimization, metaheuristic algorithms, path planning, and intelligent autonomous agents for automated code generation.

Tong Zhao received the B.E. degree in accounting and the M.E. degree in management science and engineering from China University of Petroleum (East China), Qingdao, China, in 2021 and 2024, respectively. She is currently working toward the Ph.D. degree in energy science and engineering with the Chinese University of Hong Kong, Shenzhen, Shenzhen, China.

Her research interests include electricity-carbon market, electric vehicles, and virtual power plant.

Huan Zhao (Member, IEEE) received the B.E. and M.E. degrees in software engineering from Chongqing University, Chongqing, China, in 2012 and 2015, respectively. He received the Ph.D. degree in computer and information engineering from the Chinese University of Hong Kong, Shenzhen, in 2020.

He is currently working as the Research Assistant Professor with the Hong Kong Polytechnic University. His research interests include smart grid, reinforcement learning, and energy management.

Junhua Zhao (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Queensland, Brisbane, QLD, Australia, in 2007.

He was a Senior Lecturer with the University of Newcastle and also with the Center for Intelligent Electricity Networks, University of Newcastle, Australia. He is currently the Assistant Dean and an Associate Professor with the School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen, China. He is also the Director of the Energy Market and Finance Lab, Shenzhen Finance Institute, The Chinese University of Hong Kong. His research interests include power system analysis and computation, smart grid, electricity market, data mining, and artificial intelligence.