Article:
Li, Y., Liu, K., Wang, W. et al. (2026) Auxiliary-Label Enhanced Semi Supervised Learning With Selective Pseudolabeling for Battery Capacity Estimation. IEEE Transactions on Industrial Informatics. ISSN: 1551-3203

https://doi.org/10.1109/tii.2025.3645950

Auxiliary-Label Enhanced Semi-Supervised Learning with Selective Pseudo-Labeling for Battery Capacity Estimation

Yihuan Li, Kaituo Liu, Wei Wang, Fang Fang, *Senior Member, IEEE*, and Kang Li, *Senior Member, IEEE*

***Abstract*— Recent advances in data-driven methods have significantly improved battery capacity estimation, yet most existing approaches remain constrained by their reliance on supervised learning, requiring substantial amounts of labeled cycling data that are often costly to obtain. To address this challenge, this study proposes a dual-branch network-based semi-supervised framework that integrates self-supervised learning and transfer learning mechanisms. First, the framework derives meaningful degradation-aware auxiliary labels from both labeled and unlabeled samples, creating reliable self-supervised signal for model training. Second, the designed dual-branch neural network architecture employs a shared feature extractor that processes input data for both the primary capacity estimation task and the auxiliary label prediction task, enabling effective knowledge transfer between labeled and unlabeled domains through their common representation space. Third, a pseudo-label filtering strategy is proposed to dynamically select high-confidence samples from the unlabeled dataset for self-training, thereby effectively expanding the training set with high-quality pseudo-labels and enhancing the capacity estimation accuracy. Finally, extensive experiments validate the framework's superior performance, achieving a worst-case RMSE of only 0.0143 Ah with merely 5% labeled data, representing a 27.04% reduction compared with the best-performing semi-supervised baseline (Co-training) under the same conditions.**

***Index Terms*— Semi-supervised learning, Self-training, Capacity estimation, Lithium-ion battery**

I. Introduction

LITHIUM-ION batteries play a pivotal role in promoting the low-carbon transition of transportation and energy systems [1]. However, during repeated charge–discharge cycles, the consumption of active materials and the loss of lithium inventory lead to irreversible performance degradation [2], [3]. Among various degradation indicators, the remaining capacity is one of the most critical metrics for assessing battery aging. Accurate capacity estimation is therefore crucial for health monitoring and operational safety.

This work was supported by the National Natural Science Foundation of China (52307242), in part by the Fundamental Research Funds for the Central Universities (2025JC003). (Corresponding author: Wei Wang.)

Yihuan Li, Kaituo Liu, Wei Wang and Fang Fang are with the School of Control and Computer Engineering, North China Electric Power University, Beijing, China (e-mail: yihuan@ncepu.edu.cn, kaituo@ncepu.edu.cn, wwang@ncepu.edu.cn, ffang@ncepu.edu.cn)

Kang Li is with the School of Electronic and Electrical Engineering, University of Leeds, UK (e-mail: K.Li1@leeds.ac.uk).

Recent advances in artificial intelligence have significantly propelled data-driven methods for battery capacity estimation [4]. Representative methods, such as Physics-Informed Neural Networks [5], Backpropagation Neural Networks [6], Gaussian Process Regression [7], [8], and Long Short-Term Memory (LSTM) networks [9], have shown strong capability in learning mappings from sensor data to capacity. Building upon these foundations, recent studies further refine these approaches by integrating advanced feature engineering [10], degradation-stage-aware modeling [11], and hybrid physics-data fusion [12] to tackle specific challenges like knee-point prediction. A common challenge underlying these data-driven methods is the substantial cost and time required to acquire the large labeled datasets needed for training, due to the long-term, progressive nature of battery aging. This poses a significant limitation for real-world deployment.

To alleviate the reliance on large amounts of labeled data, existing studies have explored strategies such as transfer learning and generative adversarial networks (GANs) as feasible solutions to the challenge of label scarcity. Prior studies [13]–[15] have demonstrated the potential of transfer learning in battery capacity estimation. However, with extremely limited labels, the knowledge learned from the source domain tends to shift during target task training, leading to suboptimal performance of transfer learning. Meanwhile, GAN-based approaches [16]–[18] can generate new samples that align with the distribution of original data, thereby enhancing the model's robustness. However, GANs' training process is inherently complex and prone to instability, particularly with limited training data. More critically, the lack of physical constraints in standard GAN architectures risks generating electrochemically implausible data, fundamentally limiting their reliability in capacity estimation tasks.

Beyond the above approaches, semi-supervised learning (SSL) has emerged as a more viable approach for label-efficient battery capacity estimation. This approach leverages both a few labeled samples and a large volume of unlabeled data, offering new research directions and technical pathways for enhancing battery capacity estimation [19], [20]. In existing SSL approaches, self-supervised learning has been widely adopted. By constructing pretext tasks, this method enables models to learn generalizable and structured representations from large volumes of unlabeled data, thereby providing transferable knowledge for downstream tasks [21], [22]. However, when labeled data are extremely scarce, the fine-tuning stage

may overly rely on the limited labeled samples, causing the model to overwrite or even forget the useful representations acquired during self-supervised pretraining. This degradation of the transfer learning process can ultimately lead to significant prediction bias, highlighting the need for robust pretraining frameworks that preserve learned representations during data-efficient fine-tuning [23]. Pseudo-labeling is another common SSL strategy, where predicted labels for unlabeled samples are generated using the current model and added to the supervised training process to expand the effective training set [24], [25]. However, the effectiveness of pseudo-labeling largely depends on the quality assessment mechanism. Existing methods typically rely on model confidence, which can be unreliable in early training stages and may lead to the inclusion of noisy labels and subsequent error propagation.

To further reduce the reliance on labeled data and better exploit the degradation information in unlabeled samples, this study proposes a novel SSL framework that integrates the core concepts of transfer learning and self-training. First, instead of following the conventional paradigm, we adopt the concept of transfer learning in a way that allows knowledge transfer to occur continuously throughout the training process, rather than as a pre-training stage. By providing ongoing guidance during model training, this approach enables dynamic and real-time knowledge transfer from unlabeled data to the target task, effectively guiding the model to learn degradation patterns from abundant unlabeled battery data. Subsequently, a pseudo-label quality evaluation and filtering mechanism is developed to ensure that only high-quality pseudo labels are introduced during self-training, thereby improving model performance through more effective utilization of unlabeled data. The proposed framework overcomes the inherent limitations of conventional transfer learning and self-training methods, achieving reliable predictions even when labeled data are available only from the early and middle degradation stages of a single battery. The main contributions of this work are summarized as follows:

1) To enable real-time knowledge transfer, we designed an auxiliary-label extraction mechanism capable of automatically deriving auxiliary labels from both labeled and unlabeled samples. These auxiliary labels closely reflect the battery aging process and continuously provide guidance throughout the training procedure.
2) A dual-branch neural network architecture was developed, in which the capacity serves as the training signal for the primary branch and the auxiliary label serves as the training signal for the auxiliary branch. Both branches share a common feature extraction backbone and are jointly optimized in a synchronized manner, thereby enabling dynamic and real-time knowledge transfer from unlabeled data to the target task.
3) A self-training mechanism with K-nearest neighbors (KNN) -based pseudo-label filtering is developed, where high-confidence pseudo-labels are selectively incorporated into the training process. In contrast to existing pseudo-labeling strategies that rely on model confidence, the proposed method selects pseudo-labels based on neighborhood consistency, effectively leveraging the continuity and manifold structure of battery degradation.

II. Dataset and data processing

This study employs three datasets. Dataset 1 is the primary research subject for main analyses, ablation studies and comparisons, while Datasets 2 and 3 serve as supplementary data to further evaluate the generalization across diverse battery lifespans and chemistries.

Dataset 1 was obtained from the MIT-Stanford-Toyota dataset [26], which comprises cycling data from lithium iron phosphate (LFP) batteries. These cells were charged to 80% state of charge (SOC) using multi-step constant current (CC) protocol, followed by a 1C constant current–constant voltage (CC–CV) charge to full capacity. The batteries have a nominal capacity of 1.1 Ah and a nominal voltage of 3.3 V. All tests maintained voltage limits between 2.0 V and 3.6 V, with CC discharge at 4C. In this work, 20 cells with capacity fade beyond 20% at test termination were selected from the first batch for analysis and experimental validation.

Datasets 2 and 3 were derived from the nickel cobalt manganese (NCM) and nickel cobalt aluminum (NCA) cells, as reported in [27]. All cells were cycled under three temperature conditions of 25 °C, 35 °C, and 45 °C. The cycling protocol involved CC charging to 4.2 V with current rates ranging from 0.25 C to 1 C, followed by CV charging at 4.2 V until the current dropped to 0.05 C. Discharge was performed under CC conditions to 2.65 V for NCA and 2.5 V for NCM cells. Dataset 2 comprised 12 NCM cells covering all three operating conditions, while Dataset 3 consisted of 21 NCA cells encompassing all five operating conditions.

The capacity degradation trajectories of the three datasets are illustrated in Fig. 1.

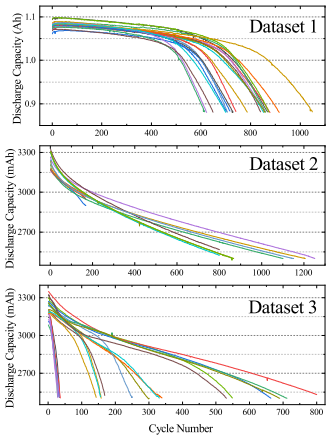

Fig. 1: Capacity decay trend of datasets 1-3.

Considering that the charging process exhibits higher stability compared to discharging, and that data from the terminal phase of charging is more accessible under practical operating conditions [28], this study selected the last 240 sampling points (including current, voltage, temperature, and timestamp) of each charging cycle in dataset 1 as the input features. For Datasets 2 and 3, 380 sampling points were selected.

Although all samples have corresponding capacity labels, to simulate real-world label scarcity, only a portion was treated as labeled, with the remainder as unlabeled. Specifically, in Dataset 1, 17 cells were used for training (split 8:2 into training/validation sets) and 3 for testing. Within the training

set, varying proportions of samples were designated as labeled data for the SSL framework. For Dataset 2, 9 cells were used for training and 3 for testing, whereas Dataset 3 comprised 16 cells for training and 5 for testing.

In practical applications, limited labeled data typically manifests as complete degradation trajectories from a few batteries, rather than sparse annotations distributed across a large battery population [17], [20]. To realistically simulate this, this study adopted a more practical data partitioning strategy, as illustrated in Fig. 2, in which labeled data were sequentially extracted from as few cells as possible instead of being randomly sampled from all available cells. This approach intentionally induces a realistic distributional shift between labeled and unlabeled datasets, resulting in labeled datasets with weaker cross-sample generalization capabilities, thereby rigorously evaluating model's capability to overcome both inter-battery variability and practical labeling constraints.

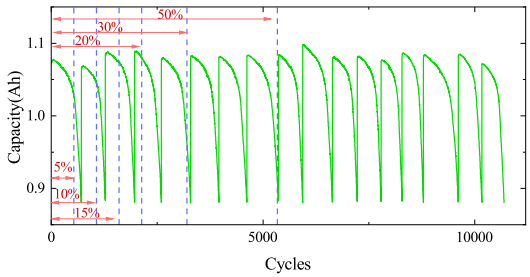

Fig. 2: Labeled Data Coverage over Degradation Cycles.

III. METHODOLOGY

To enhance capacity prediction accuracy under label scarcity and extract latent degradation information from unlabeled samples, this study proposes an SSL framework integrating the principles of self-supervised learning and transfer learning. As shown in Fig. 3, the framework first derives features highly correlated with capacity, then constructs an auxiliary task using these features as auxiliary labels. The primary capacity prediction branch and auxiliary feature learning branch are jointly optimized via a modified self-training strategy, where iterative pseudo-label refinement and feature-space alignment progressively enhance model accuracy while transferring degradation knowledge from unlabeled to labeled domains.

The following sections provide a detailed description of the proposed method from three key aspects: auxiliary labels extraction, model architecture design, and pseudo-label filtering based self-training mechanism design.

A. Auxiliary Labels Extraction

The core idea of self-supervised learning is to extract supervisory signals inherently embedded within large-scale unlabeled data [29]. By training networks using these constructed supervisory signals, the model learns representations highly valuable for downstream tasks. To enable the auxiliary task to effectively capture degradation-related information from unlabeled data, this study draws on the concept of constructing supervisory signals in self-supervised learning and designs a set of auxiliary labels. These auxiliary labels introduce constraints at the objective optimization level, guiding the training direction of the model and compensating for the lack of supervisory signals caused by the scarcity of labeled data.

These auxiliary labels are expected to possess the following characteristics: First, they should be automatically derived from the intrinsic features of each sample without the need for manual annotation. Second, the auxiliary labels should exhibit a stable and significant correlation with capacity, maintaining an approximately one-to-one mapping relationship. This ensures that the model focuses on degradation-related features rather than noise during the training process.

CC–CV charging is a widely used protocol [30]. In this study, the experimental samples are derived from partial charging data collected during the final CC-CV phase. Among them, the time proportion of the CC stage within the entire final CC-CV charging process, denoted as R_t, serves as an effective indicator of the battery's health condition.

From an electrochemical perspective, R_t is strongly linked to the key degradation mechanisms of lithium-ion batteries. As cycling progresses, loss of active materials reduces the number of sites available for lithium intercalation, while loss of lithium inventory caused by side reactions (e.g., SEI growth and lithium plating) decreases the amount of cyclable lithium. Both mechanisms shorten the CC charging stage, leading to a reduction in R_t. In addition, increased polarization and internal resistance in aged cells accelerate the voltage rise during charging, further reducing the duration of the CC phase. Consequently, R_t provides a comprehensive macroscopic indicator that implicitly reflects the underlying electrochemical degradation processes, making it physically interpretable and robustly correlated with capacity fade [30], [31].

More importantly, this parameter can be directly extracted from raw charging data with minimal computational cost, making it particularly suitable for self-supervised learning on large-scale unlabeled datasets. Therefore, in this work, R_t is designed as the auxiliary label for the auxiliary task branch to guide the network in learning degradation trends from unlabeled data, thereby providing implicit support for capacity estimation. The calculation formula for R_t is as follows:

$$R_t = \frac{t_{CC}}{t_{CC} + t_{CV}} \tag{1}$$

where t_{CV} and t_{CC} denote the durations of the CV and CC phases, respectively. Specifically, for Dataset 1, t_{CC} and t_{CV} are obtained from the final CC–CV charging stage, whereas for Datasets 2 and 3, t_{CC} and t_{CV} are extracted directly from the corresponding segments within the selected data window.

Fig. 4 visualizes the evolution of capacity and the auxiliary label R_t throughout the aging process of a representative cell, demonstrating a consistent downward trend.

Taking Dataset 1 as an example, the Pearson Correlation Coefficient (PCC) between the auxiliary label $R_t = \{R_{ti}\}_{i=1}^{n}$ and the actual battery capacity $C = \{C_i\}_{i=1}^{n}$ was computed. The PCC between them is calculated as follows:

$$r = \frac{\sum_{i=1}^{n}(R_{ti} - \bar{R}_t)(C_i - \bar{C})}{\sqrt{\sum_{i=1}^{n}(R_{ti} - \bar{R}_t)^2} \cdot \sqrt{\sum_{i=1}^{n}(C_i - \bar{C})^2}} \tag{2}$$

where R_{ti} and C_i represent the auxiliary label and capacity of the i-th sample, with $\bar{R}_t$ and $\bar{C}$ being their respective means.

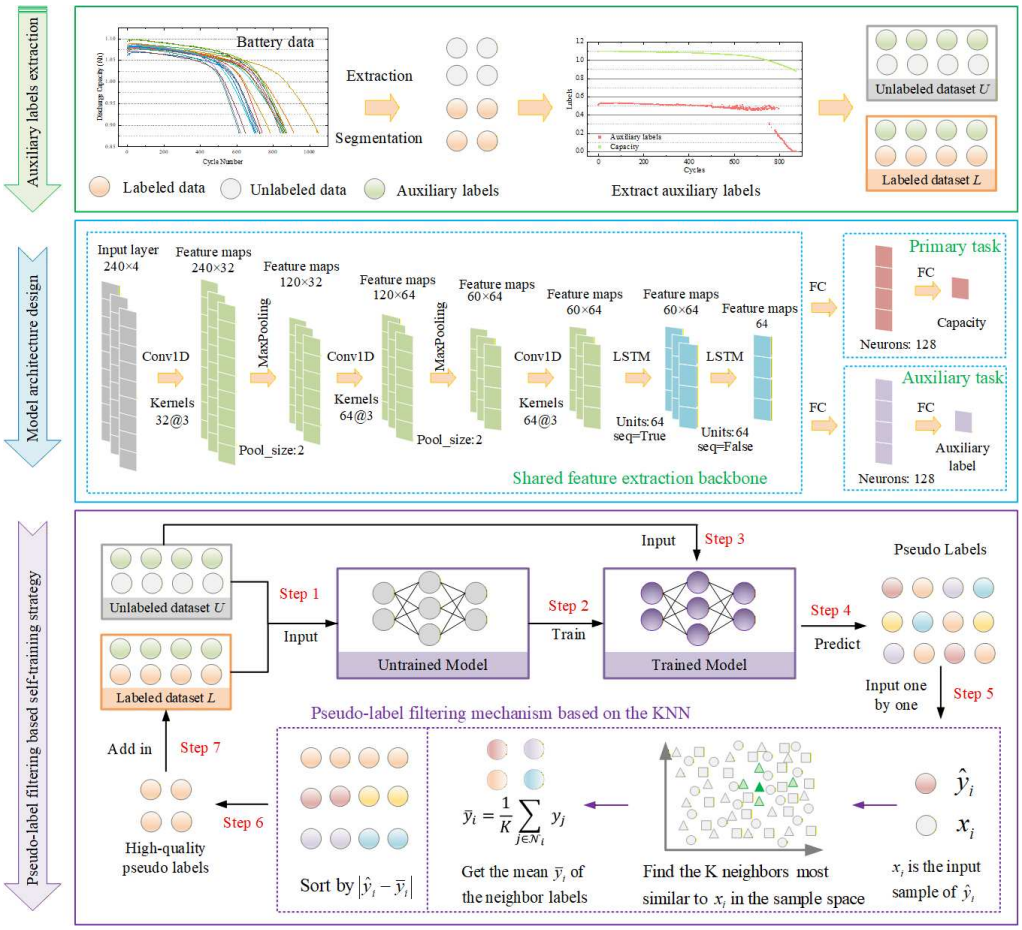

Fig. 3: The semi-supervised learning framework.

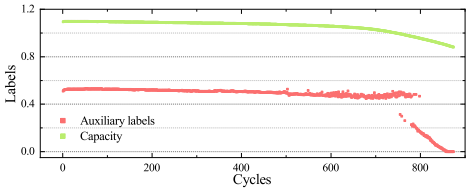

Fig. 4: Evolution of Capacity and Auxiliary Label.

TABLE I: Proportion of Intervals Exhibiting Low R_t Divergence.

h	0.03	0.05	0.07	0.09
p	55.17%	72.82%	87.45%	93.73%

The PCC between R_t and C is 0.9016, indicating a strong positive correlation.

To further verify the effectiveness of the auxiliary labels, the capacities of all cells in Dataset 1 were ranked in descending order and evenly divided into 4,000 intervals, with each interval approximately corresponding to a specific capacity value. For each interval, the pairwise differences among the corresponding R_t values were computed. A threshold h was set, and if the absolute differences within a given interval were all below h, the set of auxiliary labels corresponding to that capacity value was considered stable rather than divergent. As shown in Table I, 72.82% of these intervals exhibited extremely low R_t divergence ($h \leq 0.05$), and 93.73% were within 0.09, confirming that most capacity values map to a highly stable auxiliary label.

In summary, the auxiliary labels evolve in close alignment with the capacity degradation trend and exhibit strong positive correlation, while also maintaining stable values at each corresponding capacity level. This demonstrates a near one-to-one correspondence between capacity and the auxiliary labels, confirming the effectiveness of the auxiliary label design.

To ensure robust model training across diverse operating conditions and aging behaviors, the auxiliary labels were normalized independently within each condition. This preserves the strong correlation with capacity that exists under a given condition while eliminating scale discrepancies caused by different operating conditions.

B. Model Architecture Design

This study proposes a transfer-enhanced dual-branch network with a shared feature extraction backbone (Fig. 3, blue box), where the primary branch performs capacity regression while the auxiliary branch predicts the constructed degradation-related auxiliary label through self-supervision. We adopt the concept of transfer learning rather than its conventional methodology, enabling knowledge transfer to occur throughout the entire training process rather than prior to it. By maintaining continuous guidance from the auxiliary task during training, the proposed approach achieves dynamic and real-time knowledge transfer.

To fully exploit the local and temporal characteristics embedded in the battery charging data and capture critical

degradation patterns, a 1D CNN-LSTM model is adopted as the shared backbone of the SSL framework. The CNN module is responsible for extracting key variation patterns within local time windows, capturing local signal trends and enhancing feature representation. Building on these, the LSTM module models long-term temporal dependencies, thereby capturing the dynamic evolution of the degradation process. This backbone effectively integrates local features with global degradation trends, providing rich and robust representations to support the subsequent prediction tasks in both branches.

The auxiliary branch acts as the "source task" in the context of transfer learning, leveraging broader unlabeled data to drive the model to learn degradation-related representations. By predicting the auxiliary labels, it extract transferable knowledge from unlabeled samples that benefits the target capacity estimation task. During training, the auxiliary task is performed on all samples, and the inclusion of unlabeled data enriches the learning space, thus enhancing model generalization. From a transfer learning perspective, degradation-related representations learned by the auxiliary branch are accumulated and transferred within the shared backbone, ultimately contributing to improved performance on the primary task. Notably, the auxiliary branch is designed as a temporary training aid to support the primary task. It is removed after training and is not part of the final deployment model.

The primary task branch corresponds to the "target task" in transfer learning, which utilizes the true capacity values from a limited number of labeled samples as supervisory signals to perform accurate capacity estimation. This branch shares the high-level feature representations extracted by the CNN-LSTM backbone with the auxiliary task branch, thereby indirectly benefiting from the degradation-relevant representations learned through the auxiliary task. This collaborative structure helps alleviate the challenges associated with learning from small sample sizes. During training, the primary task branch is optimized exclusively on labeled samples.

In the proposed framework, the auxiliary branch processes both labeled and unlabeled data, while the primary branch uses only labeled data. During training, the auxiliary branch guides the shared backbone to extract generic degradation representations from unlabeled data. The backbone thus acts as a shared repository for these representations, which the primary branch continuously accesses in real time to guide and stabilize its learning. This design, centered on the auxiliary task's physically meaningful target, ensures the retention of critical aging-related information throughout training, which is fundamental for robust estimation under extreme label scarcity.

C. Pseudo-label Filtering Based Self-training Strategy

In the SSL framework, the effective utilization of unlabeled data is critical to enhance model generalization and robustness. To further exploit the latent degradation-related information embedded in the unlabeled data, this study adopts a self-training strategy to progressively expand the effective training set, thereby alleviating the issue of limited labeled data.

The core principle of self-training is to leverage the current model to generate predictions on unlabeled data and incorporate samples with predictions—along with their pseudo-labels—into the training set [32]. In doing so, the model iteratively refines itself through "self-guidance" on unlabeled samples, building upon the original labeled dataset. However, predictions on unlabeled data may exhibit bias, resulting in pseudo-labels of varying quality. Directly incorporating these pseudo-labels without proper filtering may introduce erroneous supervision and misguide the model [32]–[34]. Therefore, identifying and selecting high-quality pseudo-labels is critical to ensuring the effectiveness of the self-training process.

To address this issue, this study proposes a pseudo-label filtering mechanism based on the KNN algorithm. During the pseudo-label incorporation phase, this mechanism evaluates and filters the reliability of pseudo-labels to ensure that only trustworthy samples are included in subsequent training. The core idea of the filtering mechanism is as follows: for an unlabeled sample, if its predicted pseudo-label is close to the mean of the labels of its K most similar neighbors in the feature space, then it is considered to be of high confidence. In this work, the similarity between samples is measured using Euclidean distance. Let the feature representation of the i-th sample be $\mathbf{x}_i \in \mathbb{R}^d$, then, the Euclidean distance between sample $\mathbf{x}_i$ and sample $\mathbf{x}_j$ is defined as:

$$D(i,j) = \| \mathbf{x}_i - \mathbf{x}_j \|_2 \tag{3}$$

When labeled samples are extremely limited, the neighborhood selection becomes overly narrow, making it difficult to comprehensively assess the feature space location of unlabeled samples. To improve the coverage of neighborhood estimation, this study expands the selection domain to the entire dataset, including both labeled and unlabeled samples. By leveraging all available sample information, this strategy enhances the stability and expressiveness of the neighborhood relationships. Moreover, battery samples exhibit strong temporal continuity, with adjacent samples tending to have highly consistent capacity values. Based on this, a dual-neighbor selection strategy combining "global + local" perspectives is proposed. Specifically, global selection identifies K_g nearest neighbors by Euclidean distances across the entire candidate set, forming a broad reference group to improve the stability of pseudo-label filtering. In contrast, local selection identifies K_l nearest neighbors within the temporal neighborhood of the unlabeled sample, emphasizing local degradation consistency. The final neighborhood of each unlabeled sample thus consists of $K = K_g + K_l$ samples selected through this combined strategy.

For each unlabeled sample $\mathbf{x}_i$, let its predicted pseudo label for capacity be denoted as $\hat{y}_i$, and let the set of true or pseudo labels of its K nearest neighbors be represented by $\{y_j\}_{j\in\mathcal{N}_i}$, where $\mathcal{N}_i$ denotes the index set of the neighbors of sample $\mathbf{x}_i$. The mean value of the neighbor labels is calculated as follows:

$$\bar{y}_i = \frac{1}{K} \sum_{j\in\mathcal{N}_i} y_j \tag{4}$$

The pseudo labels $\hat{y}_i$ whose absolute differences from the neighborhood mean $\bar{y}_i$ are among the smallest P_r proportion are regarded as high-quality pseudo labels:

$$\hat{y}^{(P_r)} = \{\hat{y}_i \mid \operatorname{rank}(|\hat{y}_i - \bar{y}_i|) \le k\}, k = [P_r \cdot n] \tag{5}$$

Here, $\hat{y}^{(P_r)}$ denotes the set of high-quality pseudo labels obtained, $rank(\cdot)$ indicates sorting the absolute differences in ascending order, n represents the total number of pseudo labels, and P_r denotes the proportion of pseudo labels selected. The $\hat{y}^{(P_r)}$ are then incorporated into the labeled dataset for subsequent model updates.

This mechanism effectively improves the accuracy of pseudo labels, prevents noisy supervision from adversely affecting the model, and thereby enhances the robustness of the model in learning from unlabeled data. K_g, K_l, and P_r are chosen to be 3, 3, and 0.2, respectively, as determined by empirical sensitivity studies.

The proposed self-training strategy with pseudo label selection and incorporation is not executed in a single step. As illustrated by the purple box in Fig. 3, the process operates iteratively to progressively improve model performance.

D. Model Configurations

The detailed architecture and parameter configuration of the proposed SSL model are illustrated in the blue box of Fig. 3. The batch size is set to 256, and the number of training epochs is 200. Model parameters are optimized using the Adam optimizer with a learning rate of 10^{-3}. The model performs self-training until no further improvement in accuracy is observed.

Furthermore, to independently control the relative contributions of the main task loss, auxiliary task loss, and pseudo labels during training, a tailored loss function is designed for the proposed SSL framework.

Let the training dataset be denoted as

$$\mathcal{D} = \{(\mathbf{x}_i, y_i, a_i, m_i)\}_{i=1}^{N} \tag{6}$$

For each sample, y_i represents the label for the main task (which may be a ground-truth label or a pseudo label), a_i denotes the auxiliary task label, and $m_i \in \{0, \gamma, 1\}$ is the label weight mask that indicates the type of supervision, corresponding to unlabeled, pseudo-labeled, and labeled samples, respectively. Here, $\gamma \in (0, 1)$ is a tunable scalar representing the relative weight of pseudo labels. And N refers to the total number of samples.

The overall loss function is defined as:

$$\mathcal{L}_{\text{total}} = \alpha \cdot \mathcal{L}_{\text{real}} + \beta \cdot \mathcal{L}_{\text{aux}} \tag{7}$$

where $\mathcal{L}_{\text{real}}$ is the loss associated with the main task, and $\mathcal{L}_{\text{aux}}$ corresponds to the auxiliary task. The coefficients α and β are hyperparameters that balance the influence of the main and auxiliary losses during optimization.

For the main task,

$$\mathcal{L}_{\text{real}} = \frac{1}{\sum_{i=1}^{N} m_i} \sum_{i=1}^{N} m_i \cdot (y_i - \hat{y}_i)^2 \tag{8}$$

where $\hat{y}_i$ denote the predicted value. The weighting coefficient γ is introduced to mitigate the impact of pseudo label uncertainty on the main task learning process.

For the auxiliary task,

$$\mathcal{L}_{\text{aux}} = \frac{1}{N} \sum_{i=1}^{N} \ell_{\text{Huber}}(a_i, \hat{a}_i) \tag{9}$$

where $\hat{a}_i$ denote the predicted auxiliary label. The Huber loss$\ell_{\text{Huber}}(a, \widehat{a})$ is formulated as:

$$\ell_{\text{Huber}}(a, \widehat{a}) = \begin{cases} \frac{1}{2}(a - \widehat{a})^2, & \text{if } |a - \widehat{a}| \le \delta \\ \delta \cdot \left(|a - \widehat{a}| - \frac{1}{2}\delta\right), & \text{otherwise} \end{cases} \tag{10}$$

where δ is a smoothing parameter that controls the error tolerance range and enhances the model's robustness to outliers.

The hyperparameters α, β, and γ regulate the contributions of ground-truth, auxiliary, and pseudo-labels, respectively. By default, the weights are set as α:β:γ = 1:0.3:0.3. The SSL model training process is shown in Table II.

TABLE II: The complete training process of the SSL model.

Step	Description
1	Initialize the labeled dataset L and unlabeled dataset U.
2	Extract the auxiliary labels R_t from L and U.
3	Train the SSL model using L and U, guided by capacity C and auxiliary labels R_t.
4	For each unlabeled sample $x_i \in U$, obtain its pseudo label $\hat{y}_i$ from the trained SSL model.
5	Compute the Euclidean distance between any two samples as defined in (3).
6	For each unlabeled sample x_i, find K_g nearest neighbors from $L \cup U$ and K_l temporal neighbors in its local domain, forming the neighbor set N_i with $K = K_g + K_l$. Compute the mean label $\bar{y}_i$ of its neighbors according to (4).
7	For all unlabeled samples, rank the absolute difference $\lvert\hat{y}_i - \bar{y}_i\rvert$ in ascending order, and select the smallest P_r proportion of pseudo labels according to (5).
8	Add the selected high-quality pseudo labels $\hat{y}^{(P_r)}$ into L and assign them a confidence weight γ.
9	Repeat Steps 3–8 until the model performance no longer improves.

IV. Experimental Results and Analysis

Model performance was evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Maximum Absolute Error (MAX). All experiments were implemented in Python with TensorFlow and conducted on a computer equipped with an i5-14400F CPU and an NVIDIA GeForce RTX 4060 Ti GPU. Unless otherwise specified, Dataset 1 was used by default for all evaluation and analysis.

A. Experimental Results

To evaluate the effectiveness of the proposed method, experiments were conducted on Dataset 1 under different proportions of labeled data: 100%, 50%, 30%, 20%, 15%, 10%, and 5%. When the proportion of labeled data is 100%, the auxiliary branch and self-training strategy within the SSL model are deactivated, reducing the model to a conventional supervised learning framework based on CNN-LSTM.

The three test cells are denoted as Cell1-Cell3. The experimental results are illustrated in Fig. 5, with the corresponding evaluation metrics summarized in Table III. As the proportion of labeled data decreases, the proposed SSL model consistently tracks the true capacity with high accuracy. At a 20% labeling ratio, the 90th percentile absolute error (P90) remains low at 0.00979, 0.00627, and 0.01649 Ah for Cells1-3, respectively. Even under an extreme 5% labeling ratio—covering only the early- to mid-stage degradation of a single cell—the model is still capable of accurately fitting the actual capacity curves,

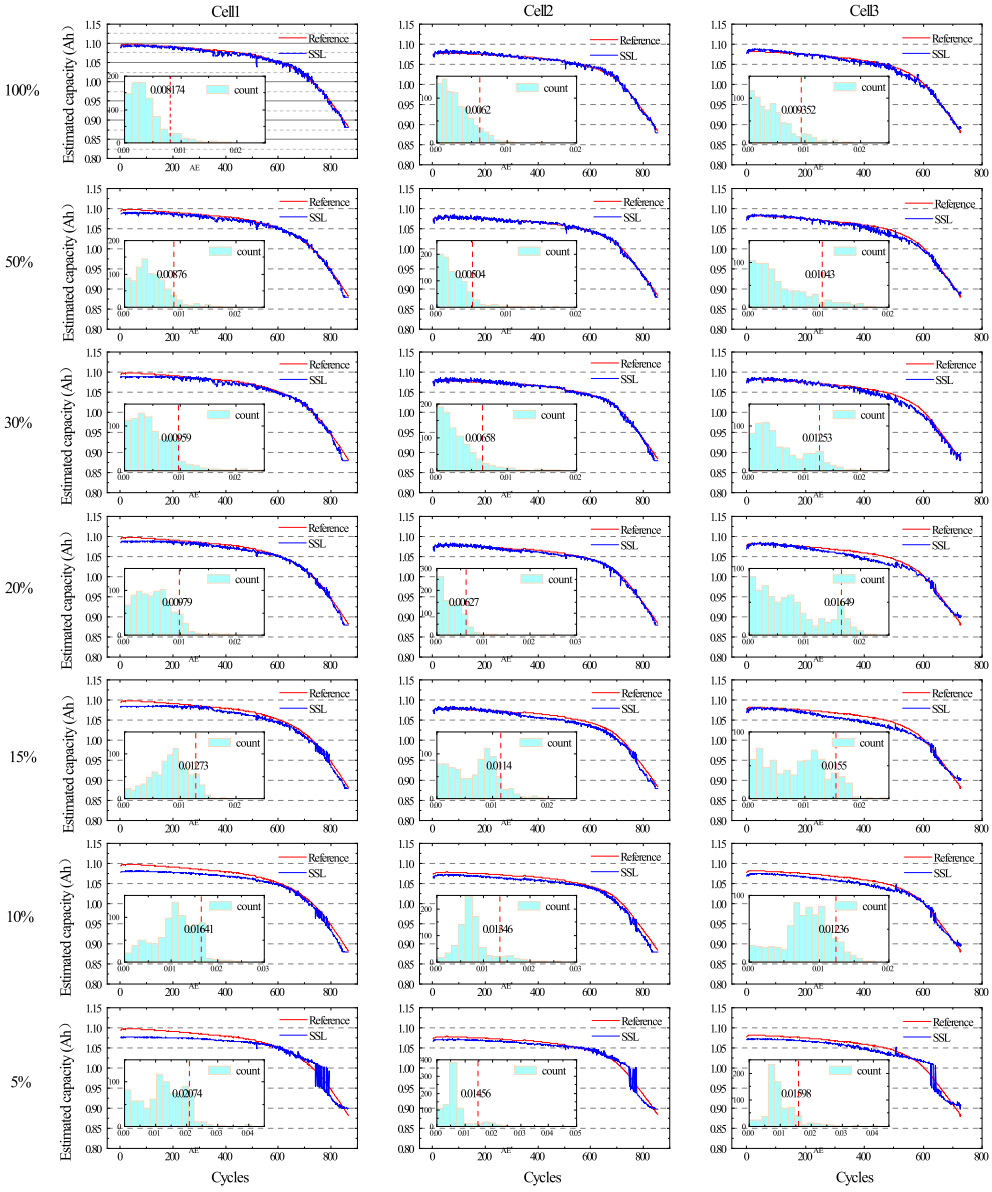

Fig. 5: Capacity prediction results of Cell1-Cell3 under varying labeled data proportions.

with MAE values for Cells1-3 remaining at 0.0124 Ah, 0.0069 Ah, and 0.0101 Ah, respectively. These results confirm that the proposed SSL model can maintain reasonable prediction accuracy even with fewer than one fully labeled battery by effectively leveraging unlabeled samples.

B. Ablation Experiment

To validate the effectiveness of each component in the proposed SSL framework, comparative experiments were conducted under varying proportions of labeled data using three model configurations: CNN-LSTM (denoted as CL), CL with the auxiliary task (CL-Aux), and CL with both the auxiliary task and self-training strategy (CL-Aux-ST).

The experimental results on Cell1-Cell3 are presented in Table IV. It can be observed that the incorporation of the auxiliary task and self-training strategy significantly improves model performance. For example, results on Cell1 demonstrate that across all labeling ratios, CL-Aux model consistently outperforms the baseline CL model in terms of MAE and RMSE. Specifically, under the 5% labeled data condition, CL-Aux achieves significant reductions of 46.13% in MAE and 51.18% in RMSE compared to CL model. In most cases, the proposed CL-Aux-ST model achieves additional performance gains over CL-Aux, exhibiting stronger generalization capabilities. Under the extreme label scarcity condition (5% labeled data), CL-Aux-ST model achieves additional reductions of 25.75% in MAE and 23.12% in RMSE compared to CL-Aux. And when compared to the original CL model, the CL-Aux-ST model achieves comprehensive error reductions of 60.00% in MAE, 62.47% in RMSE, and 67.44% in MAX. These results validate the effectiveness of incorporating the auxiliary task and self-training strategy in enhancing model performance, particularly in scenarios with limited labeled data.

To further validate the model's performance, a visual analysis of the prediction results for Cell2 was conducted. As shown in Fig. 6(a)–(f), when the proportion of labeled data is relatively high, the differences in prediction accuracy among the

TABLE III: Evaluation metrics (Ah) of the proposed SSL model on three cells under varying proportions of labeled data

Proportions of labeled data (%)	Cell1			Cell2			Cell3		
	MAE	RMSE	MAX	MAE	RMSE	MAX	MAE	RMSE	MAX
100	0.0040	0.0051	0.0252	0.0029	0.0037	0.0148	0.0042	0.0055	0.0186
50	0.0048	0.0057	0.0206	0.0025	0.0033	0.0174	0.0046	0.0061	0.0181
30	0.0052	0.0066	0.0244	0.0031	0.0042	0.0187	0.0059	0.0074	0.0239
20	0.0054	0.0064	0.0209	0.0032	0.0041	0.0264	0.0073	0.0092	0.0227
15	0.0083	0.0089	0.0191	0.0067	0.0078	0.0212	0.0084	0.0099	0.0246
10	0.0108	0.0118	0.0290	0.0079	0.0089	0.0297	0.0081	0.0088	0.0181
5	0.0124	0.0143	0.0409	0.0069	0.0087	0.0467	0.0101	0.0116	0.0415

TABLE IV: Ablation experiment: model performance under varying proportions of labeled data (unit: Ah).

	Proportions of labeled data (%)	CL			CL-Aux			CL-Aux-ST		
		MAE	RMSE	MAX	MAE	RMSE	MAX	MAE	RMSE	MAX
Cell1	50	0.0080	0.0090	**0.0171**	0.0058	0.0069	0.0206	**0.0048**	**0.0057**	0.0206
	30	0.0092	0.0101	**0.0232**	0.0069	0.0082	**0.0232**	**0.0052**	**0.0066**	0.0244
	20	0.0052	0.0064	0.0185	**0.0048**	**0.0055**	0.0198	0.0054	0.0064	0.0209
	15	0.0095	0.00105	0.0269	0.0090	0.0102	0.0296	**0.0083**	**0.0089**	**0.0191**
	10	0.0160	0.0165	**0.0271**	0.0123	0.0131	0.0275	**0.0108**	**0.0118**	0.0290
	5	0.0310	0.0381	0.1256	0.0167	0.0186	0.0447	**0.0124**	**0.0143**	**0.0409**
Cell2	50	0.0039	0.0047	0.0182	0.0030	0.0037	**0.0173**	**0.0025**	**0.0033**	0.0174
	30	0.0036	0.0046	0.0216	0.0037	0.0049	0.0188	**0.0031**	**0.0042**	**0.0187**
	20	0.0053	0.0067	0.0205	0.0047	0.0055	**0.0184**	**0.0032**	**0.0041**	0.0264
	15	0.0086	0.0113	0.0357	0.0073	0.0099	0.0373	**0.0067**	**0.0078**	**0.0212**
	10	0.0141	0.0147	**0.0280**	0.0101	0.0108	0.0319	**0.0079**	**0.0089**	0.0297
	5	0.0227	0.0319	0.1176	0.0108	0.0141	0.0552	**0.0069**	**0.0087**	**0.0467**
Cell3	50	0.0079	0.0092	0.0253	0.0064	0.0078	0.0216	**0.0046**	**0.0061**	**0.0181**
	30	0.0070	0.0089	0.0655	0.0059	0.0079	**0.0236**	**0.0059**	**0.0074**	0.0239
	20	0.0081	0.0103	0.0339	0.0085	0.0109	0.0272	**0.0073**	**0.0092**	0.0227
	15	0.0091	0.0106	0.0326	**0.0066**	**0.0077**	**0.0220**	0.0084	0.0099	0.0246
	10	0.0165	0.0176	0.0498	0.0093	0.0102	**0.0177**	**0.0081**	**0.0088**	0.0181
	5	0.0270	0.0373	0.1309	**0.0095**	0.0116	0.0441	0.0101	**0.0116**	**0.0415**

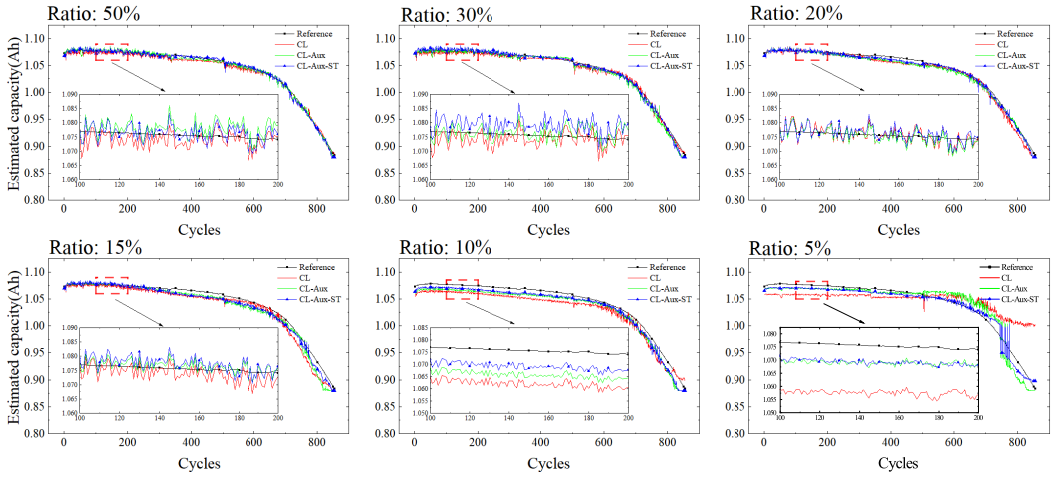

Fig. 6: Visualization of capacity predictions for the models on Cell2 in the ablation experiment.

three models are minor. However, as the proportion of labeled data decreases, both CL-Aux-ST and CL-Aux exhibit superior prediction capability compared to the original CL model, and the performance gap between them becomes increasingly apparent. Particularly, when the labeled data proportion drops to 5%, the CL model fails to effectively track the actual capacity curve, whereas CL-Aux and CL-Aux-ST still closely approximate the ground truth.

Furthermore, taking Cell2 as an example, the introduction of the auxiliary task and self-training strategy leads to a more concentrated distribution of absolute errors (AE), evidenced by a leftward shift in the P90 metric—indicating an overall reduction in estimation errors, as illustrated in Fig. 7(a)–(f). These results further validate the effectiveness of the auxiliary task and self-training strategy in improving capacity estimation accuracy and highlight the efficacy of the proposed SSL framework under limited labeled data scenarios.

C. Comparative Experiment

To validate the effectiveness of the SSL method proposed in this study, we compared it with several traditional supervised learning models, including CNN, LSTM, GRU, and ResNet, under varying proportions of labeled data.

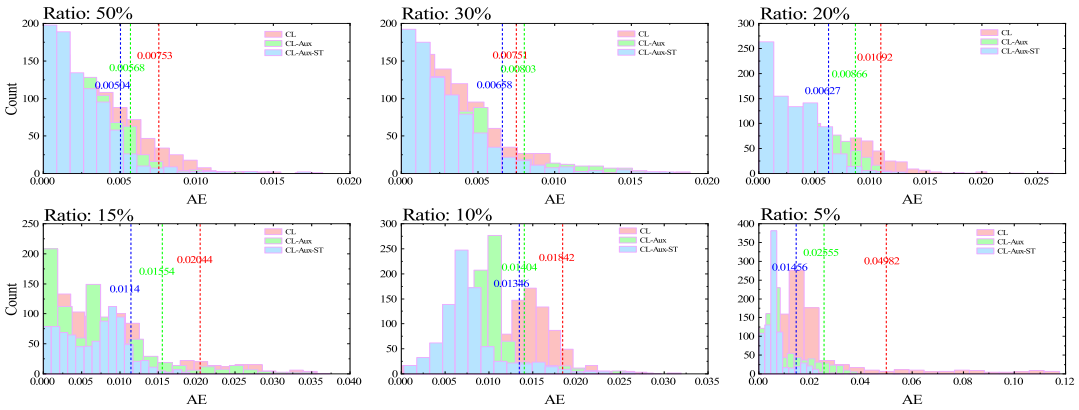

Fig. 7: Visualization of AE statistics for the models on Cell2 in the ablation experiment.

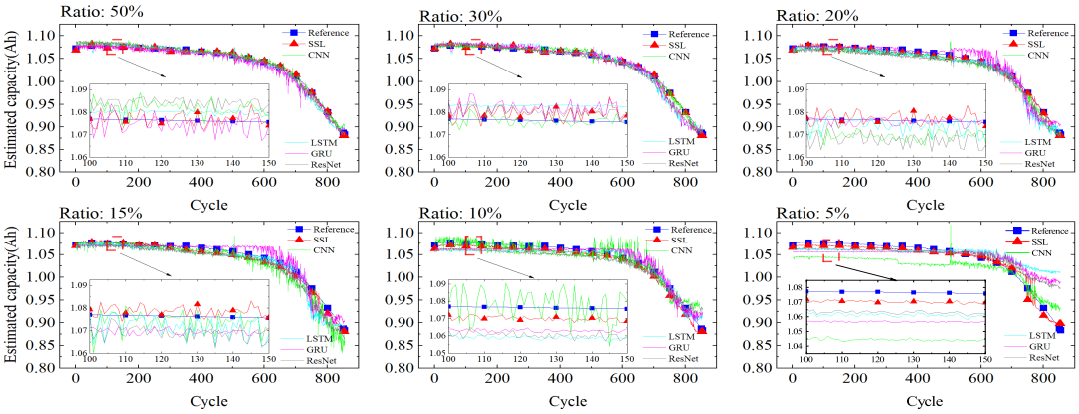

Fig. 8: Capacity prediction results of each model on Cell2 under varying proportions of labeled data.

Fig. 8 presents a visual analysis of the capacity prediction results of each model on Cell2. The results indicate that when the proportion of labeled data is relatively high, the SSL model exhibits similar prediction performance to the other models. However, SSL demonstrates superior capability in tracking the trend of capacity variation. As the proportion of labeled data decreases, the performance gap between SSL and the other models gradually widens, with the advantages of SSL becoming increasingly evident. Notably, when the labeled data proportion drops to 5%, traditional supervised models struggle to effectively track the actual capacity variations, whereas SSL maintains high prediction accuracy. These experimental results clearly demonstrate that the proposed SSL method retains strong modeling capability and reliability, even in scenarios with extremely limited labeled data.

In addition, Table V summarizes the MAE, RMSE, and MAX performance metrics for each model on Cell2. The results indicate that the proposed SSL model consistently outperforms the other models in these all three metrics, achieving the lowest error levels at every labeled data proportions. Under the extreme case of only 5% labeled data, the RMSE of the SSL model is 0.0087 Ah, which represents a 64.63% reduction compared to the best-performing supervised model (ResNet) under the same conditions.

To further validate the effectiveness of the proposed SSL method, additional comparisons were conducted with several representative semi-supervised learning approaches, including Self-training (ST), Consistency Regularization (CR), Transfer Learning (TL), and Co-training (CT) [35], [36]. To ensure a fair comparison, all baselines share the CNN-LSTM backbone and differ primarily in their training strategies. ST uses an iterative pseudo-labeling scheme to progressively augment the labeled data, CR combines supervised and consistency losses with a weighting ratio of 1:0.1, TL pre-trains on Datasets 2 and 3 and is subsequently fine-tuned on Dataset 1 with the convolutional layers frozen, and CT employs two models, a CNN-LSTM and a ResNet, that mutually exchange predictions on unlabeled data to enhance learning.

TABLE V: Model evaluation metrics (Ah) on Cell2

Performance metric	Proportions (%)	Model				
		SSL	CNN	LSTM	GRU	ResNet
MAE	50	**0.0025**	0.0046	0.0061	0.0058	0.0055
	30	**0.0031**	0.0036	0.0087	0.0069	0.0064
	20	**0.0032**	0.0080	0.0081	0.0082	0.0109
	15	**0.0067**	0.0117	0.0111	0.0099	0.0127
	10	**0.0079**	0.0129	0.0121	0.0103	0.0136
	5	**0.0069**	0.0257	0.0221	0.0213	0.0171
RMSE	50	**0.0033**	0.0059	0.0081	0.0080	0.0069
	30	**0.0042**	0.0050	0.0120	0.0099	0.0089
	20	**0.0041**	0.0089	0.0125	0.0128	0.0119
	15	**0.0078**	0.0170	0.0162	0.0138	0.0147
	10	**0.0089**	0.0162	0.0135	0.0122	0.0147
	5	**0.0087**	0.0280	0.0341	0.0291	0.0246
MAX	50	**0.0174**	0.0239	0.0277	0.0621	0.0200
	30	**0.0187**	0.0256	0.0382	0.0448	0.0364
	20	**0.0264**	0.0347	0.0658	0.0625	0.0330
	15	**0.0212**	0.0846	0.0692	0.0635	0.0432
	10	**0.0297**	0.0605	0.0428	0.0424	0.0479
	5	**0.0467**	0.0629	0.1279	0.1062	0.0939

Table VI summarizes the RMSE values of all models across three cells. The results indicate that the proposed SSL model

consistently outperforms all compared methods, achieving the lowest error levels under every labeling ratio. Notably, in the most challenging scenario with only 5% labeled data, the SSL model achieved a maximum RMSE of 0.0143 Ah, representing a 27.04% reduction compared with the best-performing semi-supervised baseline (CT) under the same condition.

TABLE VI: RMSE (Ah) of Semi-supervised models on Cells1-3.

Proportions of labeled data (%)		Model				
		SSL	ST	CR	TL	CT
Cell1	50	**0.0057**	0.0061	0.0067	0.0062	0.0066
	30	**0.0066**	0.0091	0.0080	0.0080	0.0085
	20	**0.0064**	0.0112	0.0106	0.0077	0.0099
	15	**0.0089**	0.0096	0.0086	0.0102	0.0125
	10	**0.0118**	0.0222	0.0131	0.0178	0.0153
	5	**0.0143**	0.0366	0.0370	0.0292	0.0196
Cell2	50	**0.0033**	0.0038	0.0047	0.0042	0.0038
	30	**0.0042**	0.0055	0.0047	0.0049	0.0054
	20	**0.0041**	0.0075	0.0064	0.0049	0.0064
	15	**0.0078**	0.0084	0.0131	0.0083	0.0130
	10	**0.0089**	0.0128	0.0122	0.0144	0.0141
	5	**0.0087**	0.0340	0.0344	0.0216	0.0145
Cell3	50	**0.0061**	0.0084	0.0092	0.0087	0.0069
	30	**0.0074**	0.0073	0.0102	0.0096	0.0075
	20	**0.0092**	0.0108	0.0104	0.0110	0.0105
	15	**0.0099**	0.0125	0.0110	0.0096	0.0093
	10	**0.0088**	0.0182	0.0178	0.0177	0.0121
	5	**0.0116**	0.0363	0.0365	0.0217	0.0189

In addition, Table VII summarizes the floating-point operations (FLOPs), inference time, memory usage, and training time of the semi-supervised models. Although the proposed SSL model requires a longer training time, it achieves superior predictive performance while maintaining computational complexity, inference efficiency, and memory consumption comparable to other semi-supervised approaches.

TABLE VII: Computational Efficiency of Semi-Supervised Models.

Model	SSL	ST	CR	TL	CT
FLOPs	3201538	3184769	3184769	3184769	3562754
Inference Time (ms)	7.837	7.280	7.401	7.251	8.656
Memory Usage (MB)	0.389	0.357	0.357	0.357	1.079
Training Time (s)	2006.33	896	812	198.7	717.6

These results further validate the superior reliability of the proposed method in the capacity prediction task, particularly in data-scarce scenarios.

D. Parameter sensitivity experiments

To clarify the rationale for the default configuration α:β:γ = 1:0.3:0.3, a sensitivity analysis was performed with α fixed at 1, as it weights the supervised loss from the true labels. The study examined β (with γ =0.3), γ (with β =0.3), and the pseudo-label selection ratio P_r by systematically varying them over the range [0, 0.1, ..., 1.0]. Using Cell2 from Dataset 1 as an example, the results are presented in Fig. 9.

When the labeling ratio was 50%, the model's performance remained stable across parameter variations. This is expected, as an abundance of labeled data diminishes the reliance on unlabeled samples, leading to minimal performance fluctuations with respect to changes in the semi-supervised learning parameters.

Under the 5% labeling condition, however, the parameters proved crucial. The absence of auxiliary label guidance (β=0) severely degraded model performance, with RMSE reaching 0.0295 Ah. As β increased to 0.3, the RMSE rapidly decreased to 0.0089 Ah, demonstrating the substantial benefit of incorporating auxiliary labels. When β continued to increase up to 1.0, the model performance remained stable. Similarly, excluding pseudo-label supervision (γ=0) raised RMSE to 0.0141 Ah. Increasing γ to 0.3 significantly enhanced accuracy, with the RMSE falling to 0.0078 Ah. However, a further increase to γ=1.0 caused a gradual performance decline, suggesting that excessive weight on pseudo-labels can be detrimental, likely due to the accumulation of noisy labels.

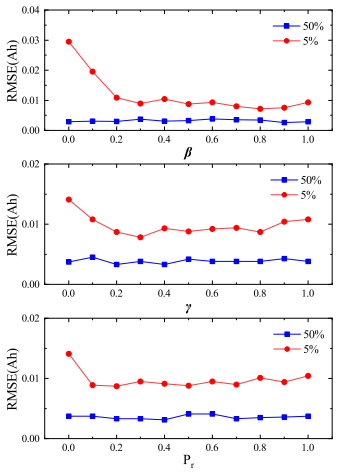

Fig. 9: Parameter Sensitivity of the SSL Model.

Similarly, when P_r = 0, the scenario corresponds to the absence of pseudo-label utilization, consistent with γ = 0. As P_r increased from 0 to 0.2, the RMSE markedly decreased to 0.0087 Ah, demonstrating the benefit of selective pseudo-label incorporation. When P_r further increased to 0.5, the model performance remained stable. However, beyond this point, the RMSE showed a slight increase, ultimately reaching 0.0104 Ah—an increase of 19.54% compared with P_r = 0.2. These findings indicate that the selection of high-quality pseudo-labels effectively improves model performance, whereas an excessive inclusion ratio may introduce noise and reduce overall accuracy.

E. Supplementary experiments

To further evaluate the performance of the proposed SSL model, additional experiments were conducted on Datasets 2 and 3 with limited labeled data (50%, 20%, 10%). The CNN-LSTM model was employed as the baseline (BL) for comparison. Table VIII summarizes the performance on three cells from Dataset 2 that tested at 25°C, 35°C, and 45°C, with charge/discharge rates fixed at 0.5C/1C. It shows that the SSL model consistently surpasses the BL across all temperatures and labeling ratios. The performance gain is especially pronounced under the most challenging 10% labeling condition. For instance, on the cell with the fewest cycles (CY25-05_1, ∼160 cycles), the SSL model reduced MAE and RMSE by 39.55% and 44.52% compared with the BL model, respectively. Similar substantial improvements were observed for cells CY35-05_1 (68.63% in MAE, 67.58% in RMSE) and CY45-05_1 (47.16% in MAE, 43% in RMSE), demonstrating its robustness even with severely limited labeled data.

TABLE VIII: Performance Metrics (Ah) of SSL vs. BL on Dataset 2.

Proportions of labeled data (%)		BL			SSL		
		MAE	RMSE	MAX	MAE	RMSE	MAX
CY25-05_1	10	0.0751	0.0867	0.2053	0.0454	0.0481	0.0780
	20	0.0404	0.0461	0.0900	0.0385	0.0432	0.0788
	50	0.0259	0.0294	0.0618	0.0224	0.0261	0.0580
CY35-05_1	10	0.0816	0.0978	0.2093	0.0256	0.0317	0.1720
	20	0.0222	0.0295	0.1635	0.0184	0.0236	0.1375
	50	0.0157	0.0220	0.1500	0.0142	0.0216	0.1522
CY45-05_1	10	0.0424	0.0500	0.3670	0.0224	0.0285	0.1904
	20	0.0204	0.0293	0.1654	0.0116	0.0178	0.1609
	50	0.0158	0.0207	0.1660	0.0095	0.0148	0.1642

Table IX shows the evaluation metrics on Dataset 3. In most cases, the proposed SSL model generally achieves higher accuracy than the baseline. This advantage is most evident under extreme data scarcity. With only 10% labeled data, the BL model fails to track degradation trends effectively, whereas the SSL model maintains reliable estimation accuracy, with the MAE less than 0.074 Ah and the RMSE less than 0.085 Ah across the five cells. When the labeled ratio increases to 20%, the SSL model's accuracy on cell CY25-1_1 is compromised by the extremely limited cycles ($\sim$30 cycles). However, it consistently outperforms the BL on the other four cells, achieving MAE/RMSE reductions of 51.65%/57.55% (CY25-05_1), 42.05%/45.52% (CY25-025_1), 67.74%/71.76% (CY35-05_1), and 59.51%/66.97% (CY45-05_1). The consistent superiority of SSL across all batteries at 50% labeled data further confirms its robust generalization capability.

TABLE IX: Performance Metrics (Ah) of SSL vs. BL on Dataset 3.

Proportions of labeled data (%)		BL			SSL		
		MAE	RMSE	MAX	MAE	RMSE	MAX
CY25-1_1	10	0.1934	0.2400	0.4896	0.0534	0.0612	0.1309
	20	0.0878	0.0984	0.1438	0.1038	0.1221	0.1870
	50	0.0585	0.0691	0.1512	0.0391	0.0531	0.0991
CY25-05_1	10	0.1582	0.2014	0.4866	0.0650	0.0751	0.1667
	20	0.1731	0.2127	0.5199	0.0837	0.0903	0.1425
	50	0.0304	0.0369	0.0837	0.0188	0.0236	0.0826
CY25-025_1	10	0.1962	0.2476	0.5294	0.0281	0.0354	0.1266
	20	0.0478	0.0681	0.2276	0.0277	0.0371	0.1916
	50	0.0211	0.0298	0.1428	0.0179	0.0241	0.1118
CY35-05_1	10	0.1842	0.2243	0.5076	0.0739	0.0851	0.1728
	20	0.1088	0.1558	0.4797	0.0351	0.0440	0.1729
	50	0.0793	0.1711	0.6424	0.0166	0.0275	0.1803
CY45-05_1	10	0.2130	0.2589	0.5136	0.0124	0.0191	0.1209
	20	0.0568	0.0893	0.3030	0.0230	0.0295	0.1081
	50	0.0280	0.0349	0.2171	0.0191	0.0218	0.1484

V. Conclusion

This study proposes a novel semi-supervised learning framework that achieves high-precision battery capacity estimation through synchronously and collaboratively optimization of a primary capacity prediction task and a self-supervised auxiliary task, coupled with a self-training strategy that dynamically expands the effective training set. The framework's label-efficient learning capability is systematically validated across varying labeled data proportions, and its performance advantages have been comprehensively evaluated through ablation studies and comparative experiments.

Under the condition of using only 5% labeled data, the CL-Aux model achieves a 51.18% reduction in RMSE compared to the baseline CL model, indicating that the joint training of the main and auxiliary tasks effectively facilitates the knowledge transfer of degradation-related information from unlabeled data to the main task. The CL-Aux-ST model further reduces RMSE by an additional 23.12% relative to the CL-Aux model, demonstrating the effectiveness of the self-training mechanism in enhancing the model's generalization ability and prediction accuracy. Moreover, under the 5% labeled data scenario, the proposed SSL model achieves a worst-case RMSE of 0.0143 Ah, representing a 62.47% reduction compared with the unmodified model and a 27.04% improvement over the best-performing semi-supervised baseline (CT) under the same conditions. Furthermore, the experimental results on Datasets 2 and 3 confirm the robustness and generalization capability of the proposed framework across different operating conditions. The framework demonstrates a practical generalization capability, achieving accurate estimation in new conditions with limited labeled data for calibration.

Despite its promising results, this study has certain limitations. The proposed auxiliary-label extraction method is specifically designed for CC–CV charging. To extend it to various charging protocols, more generalizable auxiliary-label design will be explored in the future. Nevertheless, the underlying rationale and construction principles of the auxiliary labels established in this work provide a foundation for further generalization. Although the present study focuses on single-cell analysis, the proposed approach offers theoretical support for extending degradation estimation to battery pack applications. In future research, we plan to enhance the generality of the auxiliary-label design and adapt the proposed SSL framework to pack-level implementations.

References

[1] K. Liu, Z. Wei, C. Zhang, Y. Shang, R. Teodorescu, and Q.-L. Han, "Towards long lifetime battery: Ai-based manufacturing and management," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 7, pp. 1139–1165, 2022.

[2] M. Dubarry and D. Beck, "Perspective on mechanistic modeling of li-ion batteries," *Accounts of Materials Research*, vol. 3, no. 8, pp. 843–853, 2022.

[3] N. E. Galushkin, N. N. Yazvinskaya, and D. N. Galushkin, "Causes and mechanism of thermal runaway in lithium-ion batteries, contradictions in the generally accepted mechanism," *Journal of Energy Storage*, vol. 86, p. 111372, 2024.

[4] Y. Xie, S. Wang, G. Zhang, P. Takyi-Aninakwa, C. Fernandez, and F. Blaabjerg, "A review of data-driven whole-life state of health prediction for lithium-ion batteries: Data preprocessing, aging characteristics, algorithms, and future challenges," *Journal of Energy Chemistry*, vol. 97, pp. 630–649, 2024.

[5] P. Wen, Z.-S. Ye, Y. Li, S. Chen, P. Xie, and S. Zhao, "Physics-informed neural networks for prognostics and health management of lithium-ion batteries," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 1, pp. 2276–2289, 2023.

[6] J. Wen, X. Chen, X. Li, and Y. Li, "Soh prediction of lithium battery based on ic curve feature and bp neural network," *Energy*, vol. 261, p. 125234, 2022.

[7] Z. Deng, L. Xu, H. Liu, X. Hu, Z. Duan, and Y. Xu, "Prognostics of battery capacity based on charging data and data-driven methods for on-road vehicles," *Applied Energy*, vol. 339, p. 120954, 2023.

[8] K. Peng, Z. Deng, Z. Bao, and X. Hu, "Data-driven battery capacity estimation based on partial discharging capacity curve for lithium-ion batteries," *Journal of Energy Storage*, vol. 67, p. 107549, 2023.

[9] J. He, Y. Tian, and L. Wu, "A hybrid data-driven method for rapid prediction of lithium-ion battery capacity," *Reliability Engineering & System Safety*, vol. 226, p. 108674, 2022.

[10] P. Ding, T. Li, Y. Qiao, L. Zheng, H. Deng, and W. Wu, "Integrated framework for battery soh estimation using incremental capacity and image feature transformation," *Green Energy and Intelligent Transportation*, p. 100366, 2025.

[11] T. Wang, Y. Zhu, W. Zhao, Y. Gong, Z. Zhang, W. Gao, and Y. Shang, "Capacity degradation analysis and knee point prediction for lithium-ion batteries," *Green Energy and Intelligent Transportation*, vol. 3, no. 5, p. 100171, 2024.

[12] D. Lu, F. Ren, C. Li, H. Ming, and N. Cui, "Fusing coupled degradation mechanisms with machine learning: A multi-fidelity framework for lithium-ion battery lifespan prediction," *Green Energy and Intelligent Transportation*, p. 100367, 2025.

[13] G. Ma, S. Xu, B. Jiang, C. Cheng, X. Yang, Y. Shen, T. Yang, Y. Huang, H. Ding, and Y. Yuan, "Real-time personalized health status prediction of lithium-ion batteries using deep transfer learning," *Energy & Environmental Science*, vol. 15, no. 10, pp. 4083–4094, 2022.

[14] J. Yao and T. Han, "Data-driven lithium-ion batteries capacity estimation based on deep transfer learning using partial segment of charging/discharging data," *Energy*, vol. 271, p. 127033, 2023.

[15] Y. Li, K. Li, X. Liu, Y. Wang, and L. Zhang, "Lithium-ion battery capacity estimation—a pruned convolutional neural network approach assisted with transfer learning," *Applied Energy*, vol. 285, p. 116410, 2021.

[16] J. Sun, A. Gu, and J. Kainz, "A solution framework for the experimental data shortage problem of lithium-ion batteries: Generative adversarial network-based data augmentation for battery state estimation," *Journal of Energy Chemistry*, vol. 103, pp. 476–497, 2025.

[17] 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, p. 130882, 2024.

[18] Z. Ye, J. Chang, and J. Yu, "Prognosability regularized generative adversarial network for battery state of health estimation with limited samples," *Energy*, vol. 325, p. 135922, 2025.

[19] N. Guo, S. Chen, J. Tao, Y. Liu, J. Wan, and X. Li, "Semi-supervised learning for explainable few-shot battery lifetime prediction," *Joule*, vol. 8, no. 6, pp. 1820–1836, 2024.

[20] L. Ma, J. Tian, T. Zhang, Q. Guo, and C. Y. Chung, "Enhanced battery life prediction with reduced data demand via semi-supervised representation learning," *Journal of Energy Chemistry*, vol. 101, pp. 524–534, 2025.

[21] Q. Wang, M. Ye, S. Celik, Z. Deng, B. Li, D. U. Sauer, and W. Li, "Unlocking the potential of unlabeled data: Self-supervised machine learning for battery aging diagnosis with real-world field data," *Journal of Energy Chemistry*, vol. 99, pp. 681–691, 2024.

[22] Y. Che, Y. Zheng, X. Sui, and R. Teodorescu, "Boosting battery state of health estimation based on self-supervised learning," *Journal of Energy Chemistry*, vol. 84, pp. 335–346, 2023.

[23] A. Arunan, Y. Qin, X. Li, U.-X. Tan, H. V. Poor, and C. Yuen, "Learning more with less: A generalizable, self-supervised framework for privacy-preserving capacity estimation with ev charging data," *IEEE Transactions on Industrial Informatics*, 2025.

[24] S. Zhang, Y. Li, J. Tian, Z. Man, C. Y. Chung, and W. Shen, "Improving battery life prediction with unlabeled data: Confidence-weighted semi-supervised learning with label propagation," *IEEE Transactions on Transportation Electrification*, 2024.

[25] Y. Li, Z. He, M. Ye, Q. Wang, G. Lian, Y. Sun, and M. Wei, "A semi-supervised learning strategy for lithium-ion battery capacity estimation with limited impedance data," *Energy*, vol. 319, p. 135129, 2025.

[26] K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis *et al.*, "Data-driven prediction of battery cycle life before capacity degradation," *Nature Energy*, vol. 4, no. 5, pp. 383–391, 2019.

[27] J. Zhu, Y. Wang, Y. Huang, R. Bhushan Gopaluni, Y. Cao, M. Heere, M. J. Mühlbauer, L. Mereacre, H. Dai, X. Liu *et al.*, "Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation," *Nature communications*, vol. 13, no. 1, p. 2261, 2022.

[28] X. Li, J. Jiang, D. Chen, Y. Zhang, C. Zhang *et al.*, "A capacity model based on charging process for state of health estimation of lithium ion batteries," *Applied energy*, vol. 177, pp. 537–543, 2016.

[29] T. Wang, Z. Ma, S. Zou, Z. Chen, and P. Wang, "Lithium-ion battery state-of-health estimation: A self-supervised framework incorporating weak labels," *Applied Energy*, vol. 355, p. 122332, 2024.

[30] M. Wang, S. Wu, Y. Chen, and W. Luan, "The snowball effect in electrochemical degradation and safety evolution of lithium-ion batteries during long-term cycling," *Applied Energy*, vol. 378, p. 124909, 2025.

[31] P. M. Attia, A. Grover, N. Jin, K. A. Severson, T. M. Markov, Y.-H. Liao, M. H. Chen, B. Cheong, N. Perkins, Z. Yang *et al.*, "Closed-loop optimization of fast-charging protocols for batteries with machine learning," *Nature*, vol. 578, no. 7795, pp. 397–402, 2020.

[32] X. Fan, W. Zhang, H. Qi, and X. Zhou, "Accurate battery temperature prediction using self-training neural networks within embedded system," *Energy*, vol. 313, p. 134031, 2024.

[33] Y. Che, D.-I. Stroe, X. Hu, and R. Teodorescu, "Semi-supervised self-learning-based lifetime prediction for batteries," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 5, pp. 6471–6481, 2022.

[34] S. Zheng and J. Zhao, "A self-adaptive temporal-spatial self-training algorithm for semisupervised fault diagnosis of industrial processes," *IEEE transactions on industrial informatics*, vol. 18, no. 10, pp. 6700–6711, 2021.

[35] X. Yang, Z. Song, I. King, and Z. Xu, "A survey on deep semi-supervised learning," *IEEE transactions on knowledge and data engineering*, vol. 35, no. 9, pp. 8934–8954, 2022.

[36] J. Zhao, D. Li, Y. Li, D. Shi, J. Nan, and A. F. Burke, "Battery state of health estimation under fast charging via deep transfer learning," *iScience*, vol. 28, no. 5, 2025.

Yihuan Li received the Ph.D. degree in Electronic and Electrical Engineering from the University of Leeds, U.K., in 2021. She is currently a Lecturer at North China Electric Power University, Beijing, China. Her research interests include intelligent modeling, monitoring and control, and their applications to energy storage systems.

Kaituo Liu received the B.S. degree in robotic engineering from Nanjing University of Information Science and Technology, Nanjing, China, in 2023. He is currently pursuing the M.S. degree in control science and engineering at North China Electric Power University, Beijing, China. His research interests include battery state monitoring and machine learning.

Wei Wang received the Ph.D. degree from North China Electric Power University, Beijing, China, in 2011. He is currently a Professor with the State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University. His research interests mainly include the modeling, optimization and control of integrated energy systems.

Fang Fang (Senior Member, IEEE) received the Ph.D. degree from North China Electric Power University, Beijing, China, in 2005. His current research interests include the configuration and operation of integrated energy systems, and the optimal dispatching of virtual power plants. Prof. Fang is the Chair of China Energy Research Society Technical Committee on Smart Power Generation and a Council Member of the Technical Committee on Industrial Information Physical Systems, IEEE Industrial Electronics Society.

Kang Li (Senior Member, IEEE) received the Ph.D. degree in control theory and applications from Shanghai Jiao Tong University, Shanghai, China, in 1995. He is currently the Chair of Smart Energy Systems at the University of Leeds, Leeds, U.K. His current research interests include nonlinear system modeling, identification and controls, and artificial intelligence, with applications to energy and power systems, smart grids, electrification of railway systems, and energy management in energy-intensive manufacturing processes.