Battery Life Prediction with Scarce Data using Physics-Informed Data Generation and Adaptive Autoencoder

Song Zhang, *Student Member, IEEE*, Mengru Liu, Ruohan Guo, *Member, IEEE*, Jinpeng Tian, *Member, IEEE*, Zhihong Man, *Member, IEEE*, and Weixiang Shen, *Senior Member, IEEE*

***Abstract*—Reliable prediction of battery remaining useful life (RUL) is essential for ensuring the safety and operational efficiency of electric vehicles (EVs). However, current data-driven RUL prediction methods often face significant limitations due to the costly and time-intensive process of collecting complete high-quality degradation data across the battery lifecycle. To address this challenge, a novel framework is proposed, in which physics-informed synthetic data generation is integrated with an adaptive autoencoder-based neural network for efficient RUL prediction. Unlabeled synthetic data reflecting battery degradation behaviors are used to augment the limited training samples. RUL prediction and input reconstruction are jointly performed by the autoencoder to enable structure-aware latent representation learning. Moreover, a dynamically adaptive loss weighting mechanism is introduced to balance predictive accuracy and structural consistency during training. Ablation studies demonstrate that incorporating input reconstruction and synthetic data generation reduces prediction error by over 60% compared to a conventional supervised-only baseline. Furthermore, under data-scarce conditions with only 150 labeled data, the proposed method further outperforms the best semi-supervised baseline by over 15%. These results highlight the robustness and generalization capability of the proposed framework under limited data scenarios.**

***Index Terms*—Battery management, Deep learning, Remaining useful life, Synthetic data generation.**

NOMENCLATURE

A. Abbreviations

EV	Electric vehicle
RUL	Remaining useful life
ECM	Equivalent circuit model
EN	Elastic net
SVR	Support vector regression
MLP	Multilayer perceptron
CNN	Convolutional neural network
RNN	Recurrent neural network
LSTM	Long short-term memory network
GAN	Generative adversarial network
VAE	Variational autoencoder
EOL	End of life
IC	Incremental capacity
PCA	Principal component analysis
GMM	Gaussian mixture model
NCM	nickel cobalt manganese oxide batteries
LFP	Lithium iron phosphate batteries
CC-CV	Constant current constant voltage
RMSE	Root mean squared error
PINN	Physics-informed neural networks
STD	Standard deviation
SOC	State of charge

B. Variables and Symbols

$x,\ X$	Input
y	Output
θ	parameter
H	Latent representation
Q	Capacity
V	Voltage
β	Loss weighting coefficient

I. INTRODUCTION

THE rapid global adoption of electric vehicles (EVs) has intensified the deployment of lithium-ion batteries, placing battery health management a key focus for both research and transportation electrification [1]. However, battery aging results from a complex interplay of electrochemical, thermal, and mechanical processes, including side reactions that are strongly influenced by external loads and ambient conditions. Accurate prediction of battery remaining useful life (RUL) is therefore critical for ensuring operational reliability, optimizing maintenance schedules, and enabling predictive diagnostics in EVs [2].

Existing battery RUL prediction methods can be broadly categorized into model-based and data-driven approaches. Model-based approaches range from simple equivalent circuit models (ECMs) to more elaborate physics-based models grounded in electrochemical principles. For instance, Tran et al. [3] incorporated battery health features into an ECM by characterizing its influence on ohmic resistance and polarization components, improving model accuracy for BMS integration. Guha and Patra [4] combined capacity fade and in-

The work of Song Zhang was supported in part by Australian Government Research Training Program Scholarship. *(Corresponding author: Jinpeng Tian.)*

Song Zhang, Zhihong Man, and Weixiang Shen are with the School of Engineering, Swinburne University of Technology, Hawthorn, VIC 3122, Australia (e-mail: songzhang@swin.edu.au; zman@swin.edu.au; wshen@swin.edu.au).

Mengru Liu, Ruohan Guo, and Jinpeng Tian are with the Department of Electrical and Electronic Engineering, Research Centre for Grid Modernisation, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China (e-mail: maura.liu@connect.polyu.hk; ruohan.guo@polyu.edu.hk; jinpeng.tian@polyu.edu.hk).

ternal resistance growth models within a particle filtering framework, demonstrating enhanced RUL prediction stability. Lashway and Mohammed [5] further developed an adaptive physics-based monitoring method that dynamically updates ECM parameters and usable capacity through pulsed-load tests, enabling chemistry detection and health tracking. While ECMs are widely used in practice due to their simplicity and computational efficiency, they provide limited insight into internal degradation mechanisms. Compared to ECMs, electrochemical models offer greater physical interpretability in simulating aging behaviors, but they often require extensive calibration, suffer from incomplete physics, and are typically constrained to specific chemistries and operating conditions, which limits their scalability in diverse applications [6].

In contrast, data-driven approaches have been gaining attention due to their ability to learn complex and nonlinear degradation patterns directly from historical data or extracted features to predict RUL [7]. These methods are particularly advantageous when mechanistic understanding is limited or when diverse real-world usage patterns must be accommodated. Early efforts were made to explore classical machine learning techniques such as elastic net (EN) [8], support vector regression (SVR) [9], and multilayer perceptrons (MLPs) [10], which typically rely on feature engineering and have shown competitive performance under sufficient data scenarios. For example, Zhao et al. [11] extracted statistical features from charging-discharging profiles and applied SVR to predict RUL, demonstrating the effectiveness of handcrafted features under constrained settings.

In parallel, other strategies have been proposed to enhance battery lifetime prediction. Fan et al. [12] introduced a feature-fusion framework leveraging voltage and impedance signals, while Li et al. [13] developed a sparse-sensor system for electrothermal monitoring in large-format battery packs. Although insightful, these approaches require either handcrafted feature engineering or extensive instrumentation. More recently, deep learning architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have achieved superior accuracy by directly learning from low-preprocessed data. Park et al. [14], for instance, utilized a long short-term memory (LSTM) network to capture temporal dependencies in multi-channel charging profiles, significantly improving RUL prediction performance.

Despite progress, most data-driven methods are constrained by their reliance on supervised learning, requiring large volumes of labeled degradation data [15], [16]. Acquiring such high-quality run-to-failure data remains challenging due to the long testing durations, complex environmental conditions, substantial financial, and energy costs associated with battery aging. This reliance on scarce data fundamentally limits the scalability and practical deployment of existing models, particularly in low-data environments [17].

In addition, battery RUL prediction is intrinsically complex due to the nonlinear nature of degradation trajectories across different operating conditions [18]. Capturing this nonlinearity typically demands large and diverse training datasets to ensure generalization across battery chemistries and usage patterns, but such resources are rarely available in practice.

To address the challenge of RUL prediction under data scarcity, existing research generally involves incorporating domain knowledge through physical modeling or physics-informed constraints, enabling the model to extract degradation-relevant features from limited data [19]. These approaches aim to improve generalization by leveraging aging-related mechanisms such as voltage plateaus, phase transitions, or impedance shifts. Among the available electrochemical signals, charge-voltage (Q-V) curves and their incremental capacity (IC) counterparts have emerged as particularly valuable indicators of degradation. Q-V curves directly reflect capacity-voltage relationships that compress and shift over time, while IC curves accentuate phase-transition signatures through the evolution of characteristic peaks. These consistent and interpretable patterns have been validated across various chemistries and cycling protocols, establishing IC analysis as a mature and widely adopted tool for lifetime estimation.

Nevertheless, purely physics-based models often struggle scale due to incomplete mechanistic understanding and the high cost of parameterization. Conversely, collecting sufficient labeled run-to-failure data across diverse usage profiles remains costly and time-consuming. In many real-world cases, available records are only partially observed or lack ground-truth RUL labels. As a result, accurate and robust RUL prediction continues to be a highly challenging task, requiring approaches that balance modeling flexibility, data efficiency, and resilience to incomplete supervision.

To mitigate dependency on the large demand of labeled data, semi-supervised learning methods have emerged [20], which utilizes both labeled and unlabeled data to improve prediction performance under partial supervision. For example, Jiang et al. [21] proposed an adaptive self-learning framework where a teacher-student model generates pseudo-labels for unlabeled samples. Lin et al. [22] introduced a co-training strategy using dual regressors trained on distinct feature views. Zhang et al. [23] applied label propagation to assign pseudo-RUL values based on neighborhood similarity. While effective, these approaches still require substantial real-world unlabeled degradation data covering diverse usage patterns, which is also expensive to obtain.

As an alternative, synthetic data generation has attracted attention, particularly through generative adversarial networks (GANs) or variational autoencoders (VAEs), which model data variation patterns from observed data without relying on explicit physical models [24]. For instance, Jiang et al. [25] proposed a refined conditional variational autoencoder to generate high-resolution charging data conditioned on battery lifespan parameters, facilitating synthetic electrochemical datasets for state prediction. Zhang et al. [26] used a convolutional recurrent GAN to synthesize realistic run-to-failure degradation sequences. While these synthetic samples typically reflect degradation trends, they often lack ground-truth labels [27].

In parallel, physics-based simulation frameworks have also been explored for generating synthetic battery degradation

data. For example, Qasem et al. [28] proposed an electrochemical model-driven synthetic data strategy to support eVTOL battery modeling, using internal cell parameters and aging simulations to generate realistic degradation patterns. While such approaches benefit from strong physical interpretability, they often require extensive parameter calibration and prior knowledge of battery-specific electrochemical characteristics. This dependence on detailed internal models limits their scalability and adaptability to unseen batteries or varying operational conditions.

Therefore, both data-driven and purely physics-based methods face notable limitations. The former may lack physical consistency and generalizability, while the latter can be computationally demanding and dependent on difficult-to-obtain internal battery parameters. These trade-offs motivate the development of hybrid approaches that balance physical interpretability with data adaptability.

To address these challenges, a physics-informed synthetic data generation strategy is proposed, in which degradation trajectories are synthesized based on parameterized aging patterns derived from real battery cycling data. These synthetic samples preserve realistic degradation behaviors and enrich the training distribution without requiring additional experimental cycling. To utilize this unlabeled synthetic data, an autoencoder-based learning framework is adopted, which jointly learns feature representations and performs RUL regression. The encoder extracts latent features from both real and synthetic data, while the decoder reconstructs the input to enforce structural consistency. To coordinate the dual learning objectives, input reconstruction and RUL prediction, a dynamic loss weighting mechanism is incorporated, which adaptively balances their contributions throughout training. This mechanism ensures stable convergence, prevents the auxiliary reconstruction task from dominating, and enables the model to adjust its learning focus based on the relative reliability of real and synthetic data.

The main contributions of this work are as follows:

1) A physics-informed synthetic data generation strategy is developed to augment training samples with physically consistent and diverse degradation trajectories, reducing dependence on large-scale experimental cycling.

2) An adaptive autoencoder-based learning framework is introduced to extract structure-aware latent features from both real and synthetic data. The framework jointly optimizes input reconstruction and RUL prediction objectives, enabling effective feature learning under unlabeled data conditions and distribution shifts relative to real degradation data.

3) A unified and data-efficient RUL prediction framework is established. The framework is validated across two battery datasets, consistently achieving over 60% error reduction compared to conventional baselines.

II. Methodology

A. Problem Formulation and Overview of Proposed Method

Battery RUL refers to the number of charge-discharge cycles remaining before reaching end-of-life (EOL), typically defined as 80% of nominal capacity in EV applications [29]. It quantifies the expected lifespan at any given time and is commonly formulated as a supervised regression task, where future degradation is estimated from historical measurements. The prediction model is expressed as

$$\hat{y}_t = f(X_{1:t}, \theta) \tag{1}$$

where $\hat{y}_t$ denotes the predicted RUL at cycle t, $X_{1:t}$ represents the historical data observed up to cycle t, and $f(\cdot)$ indicates the prediction model, whose parameter set is θ.

Training such models requires labeled degradation trajectories with EOL annotations. However, collecting high-quality labeled data remains costly and time-consuming, motivating the use of frameworks that leverage both limited labeled and abundant unlabeled data.

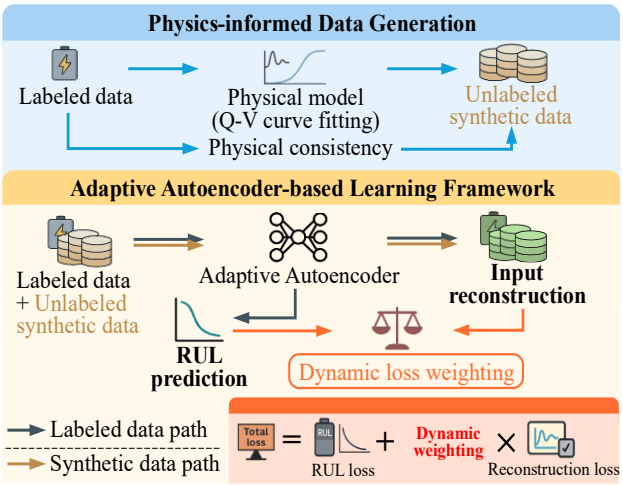

Fig. 1. Overview of the proposed battery RUL prediction framework.

The overall framework, illustrated in Fig. 1, consists of two components: a physics-informed synthetic data generation pipeline and an adaptive autoencoder-based prediction network. The first component generates physically consistent degradation samples to augment limited real datasets, enhancing training diversity without additional experimental cycling. The second component jointly performs input reconstruction and RUL prediction to extract informative features from both labeled and unlabeled data. A dynamic loss weighting mechanism coordinates the two learning objectives by prioritizing generalizable representations in early stages and gradually emphasizing prediction accuracy.

B. Physics-informed Data Generation

The performance of supervised battery RUL prediction is often constrained by limited access to degradation data. To mitigate this, a physics-informed synthetic data generation strategy is developed to produce unlabeled yet physically consistent synthetic data to augment training dataset.

Unlike data-driven generators such as GANs, which may produce unrealistic or physically inconsistent data, the proposed method synthesizes degradation data by learning and expanding the distribution of real battery behaviors. Instead of relying on complex electrochemical models, a compact

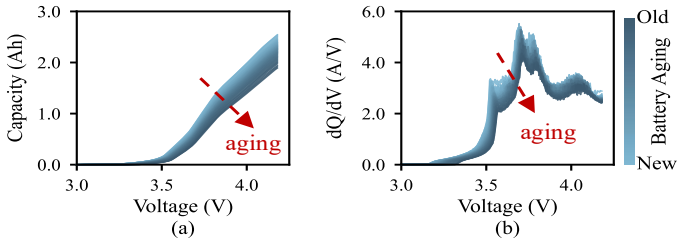

Fig. 2. Battery degradation curves: (a) Q-V curves; (b) IC curves.

parametric representation based on Q-V charging curves is adopted for its controllability and practical settings [30]. This modeling choice has also shown effectively in related studies which utilized Q-V-based features for battery health estimation [31]. Fitted parameters are reduced via principal component analysis (PCA), and synthetic latent vectors are sampled using a Gaussian mixture model (GMM). The inverse PCA transformation then maps these samples back to the parametric domain, yielding synthetic Q-V curves with retained diversity and physical meaning.

1) Physical-informed Q-V Curve Modeling

Battery Q-V curves and their derivatives (incremental capacity, IC or *dQ/dV*) curves are widely used for aging analysis due to their clear and consistent shifts during degradation [32]. The data used in this section are obtained from the Tongji battery dataset [33]. They serve as representative examples to illustrate the progressive degradation patterns and the proposed physics-informed data generation procedure. Fig. 2 illustrates the full lifetime evolution of this battery's Q-V and IC curves across hundreds of charging cycles.

As aging progresses, the Q-V curves exhibit plateau compression and horizontal shifts, reflecting increased internal resistance and reduced capacity. These gradual shape changes, although subtle in Q-V space, manifest as prominent evolving peaks in the IC domain (Fig. 2(b)), highlighting the sensitivity of the *dQ/dV* transform to underlying electrochemical variations [34]. Specifically, the positions, heights, and widths of IC peaks are known to correlate with phase transitions, lithium plating, and loss of active material in different stages of battery aging. Therefore, IC curves provide a compact and physically meaningful representation of the battery's internal degradation process.

These visualizations confirm that Q-V and IC profiles serve as degradation-sensitive features that are both physically grounded and well-suited for data-driven modeling. They form the basis of our physics-informed synthetic data generation strategy. Details of dataset descriptions are provided in *Supplementary Note 1*.

Notably, the peaks observed in IC curves closely resemble Lorentzian-shaped functions, suggesting that battery IC curves can be effectively approximated by a superposition of Lorentzian functions [35]:

$$\frac{dQ}{dV} = \sum_{i=1}^{n} \frac{2A_i}{\pi} \frac{\omega_i}{\omega_i^2 + 4\left(V - V_{0,i}\right)^2} \tag{2}$$

where n represents the number of Lorentzian peaks, and each Lorentzian peak is characterized by three physically interpretable parameters: amplitude A_i, half-width ω_i, and center voltage $V_{0,i}$. These parameters succinctly capture battery charging/discharging dynamics and degradation characteristics.

However, directly fitting IC curves is challenging due to noise amplification introduced by numerical differentiation of Q-V data. While smoothing or filtering can reduce noise, they may also distort critical degradation patterns. To circumvent this, we instead model Q-V charging curves analytically by integrating the Lorentzian function from (2), yielding a smooth and physically interpretable representation [35]:

$$Q = \sum_{i=1}^{n} \frac{A_i}{\pi} \tan^{-1}\left(2\frac{V - V_{0,i}}{\omega_i}\right) + C \tag{3}$$

In this integrated form, C represents an integration constant, while the remaining parameters, A_i, ω_i, and $V_{0,i}$ retain the same interpretations in (2). This formulation avoids derivative-induced noise and enables analytical derivation of IC curves from fitted Q-V data, preserving critical physical insights and information.

To construct a compact and physically interpretable representation of each Q-V curve, we begin by analyzing the structure of the corresponding IC curve. These IC curves typically exhibit multiple distinct peaks, each associated with specific electrochemical processes. To determine the optimal number of components, model fitting is performed on real Q-V data modeled as a superposition of Lorentzian-integrated functions. Any suitable fitting method can be adopted in this step, provided that it ensures faithful reconstruction of the Q-V profile. In this study, parameter fitting consistently identified four peaks as the most appropriate, aligning with the dominant four-peak pattern observed empirically. Details of the parameter fitting process are provided in *Supplementary Note 2*.

Each Lorentzian peak is characterized by three parameters: amplitude, width, and center voltage, yielding $4 \times 3 = 12$ values per cycle. Together with an integration constant C, each fitted Q-V curve is thus represented by a 13-dimensional parameter vector. Thus, the parameter space is organized as a matrix $X \in \mathbb{R}^{n_s \times 13}$, where n_s denotes the number of battery cycles. Each row $x^{(j)} \in \mathbb{R}^{13}$ corresponds to the fitted parameters for the j-th battery cycle:

$$x^{(j)} = \left[A_1^{(j)}, \omega_1^{(j)}, V_{0,1}^{(j)}, \ldots, A_4^{(j)}, \omega_4^{(j)}, V_{0,4}^{(j)}, C^{(j)}\right] \tag{4}$$

These representations transform raw degradation data into a structured parametric form, offering a physically interpretable basis for subsequent dimensionality reduction and synthetic data generation. Examples of fitted parameter evolution and fitting quality are provided in *Supplementary Note 2*.

The Lorentzian-fitted parameters extracted from individual Q-V curves encode intra-cycle structural characteristics and nonlinear electrochemical phenomena such as polarization and intercalation phase transitions [36]. These features manifest as cycle-wise changes in the amplitude, position, and width of the fitted peaks, providing physically meaningful indicators of battery aging.

2) Dimensionality Reduction and Distribution Modeling

The Lorentzian-based parameterization provides a compact, physically meaningful 13-dimensional representation of each Q-V curve. However, directly modeling this parameter space remains challenging due to strong correlations and redundancy among the peak parameters, further complicated by nonlinear degradation trends across cycles. To reduce complexity, PCA is applied, which efficiently compresses the parameter matrix $X \in \mathbb{R}^{n_s \times 13}$ (containing fitted vectors from n_s cycles) into a lower-dimensional representation while preserving essential variance [37]. Each parameter vector $x^{(j)}$ into a lower-dimensional latent space via:

$$z^{(j)} = \left(x^{(j)} - \mu\right)W, \quad j = 1,2,\dots,n_s \tag{5}$$

where $\mu \in \mathbb{R}^{13}$ denotes the empirical mean vector computed across all fitted parameter vectors $x^{(j)}$, i.e., $\frac{1}{n_s}\sum_{j=1}^{n_s} x^{(j)}$, and $W \in \mathbb{R}^{13\times m}$ consists of the eigenvectors corresponding to the top m largest eigenvalues of the covariance matrix of the parameter matrix X. Here, a value of $m = 3$ is selected to balance dimensionality reduction and information retention. PCA applied to the 13-dimensional parameter vectors extracted from all cycles of the representative Tongji battery (as discussed in Section II-C.1) shows that the top three principal components account for over 99.96% of the total explained variance. This confirms that most of the structural variability in the Q-V degradation behavior can be efficiently represented in a compact latent space.

Rather than filtering out aging-related features, PCA serves to reduce redundancy while retaining the dominant variance associated with degradation. The resulting low-dimensional representation captures the gradual evolution of physically meaningful signals, provided the degradation follows a relatively smooth trajectory under controlled charging protocols. Furthermore, the linear nature of PCA enables straightforward inverse transformations from the latent space back to the original parameter space for reconstructing Q-V curves when the underlying parameter evolution follows a relatively smooth trajectory, as is typically observed under controlled charging protocols.

To facilitate synthetic data generation, a GMM is employed to model the distribution of degradation features within the physics-informed latent space. This latent space is constructed via PCA from Lorentzian-fitted Q-V parameters, which encode electrochemical degradation dynamics. As these parameters are fitted from real curves, key physical characteristics, such as capacity fade and voltage plateau shifts, are already embedded. Within this structured space, the GMM provides a statistically principled and physically constrained method for oversampling. The use of PCA not only reduces dimensionality and computational complexity but also helps regularize the latent space, making GMM fitting more robust and interpretable across batteries with varying chemistries. While GMM does not aim to approximate every degradation mechanism, it densifies the existing distribution without distorting physical consistency.

In this study, the number of components was empirically set to $K = 4$, based on cluster visualization and validation performance, striking a balance between sample diversity and overfitting risk. The latent space is modeled as a weighted mixture of K Gaussian components:

$$p(z) = \sum_{k=1}^{K} \pi_k \mathcal{N}(z|\mu_k, \sum_k) \tag{6}$$

where π_k, μ_k , and $\sum_k$ denote the weight, mean, and covariance of the k-th Gaussian component, respectively, with $\sum_k \pi_k = 1$. This probabilistic formulation follows the standard GMM representation [37]. After training the GMM, synthetic latent vectors $z_{syn}^{(j)}$ can be oversampled by:

$$z_{syn}^{(j)} \sim \sum_{k=1}^{K} \pi_k \mathcal{N}(\cdot|\mu_k, \sum_k), \quad j = 1,2,\dots,N_{syn} \tag{7}$$

where N_{syn} represents the desired number of synthetic samples, is typically chosen significantly greater than the original sample number n_s, effectively performing oversampling."

3) Inverse Mapping and Synthetic Q-V Curve Construction

After generating synthetic latent vectors ($z_{syn}^{(j)}$) from the GMM, they are transformed back into the original Lorentzian parameter space using the inverse PCA transformation:

$$x_{syn}^{(j)} = z_{syn}^{(j)} W^T + \mu, \quad j = 1,2,\dots,N_{syn} \tag{8}$$

where$x_{syn}^{(j)} \in \mathbb{R}^{13}$ represents the reconstructed synthetic parameter vector for the j-th synthetic sample, W^T is the transpose of the PCA projection matrix, and μ is the mean vector defined in (5).

Each reconstructed parameter vector defines a physically interpretable Q-V curve based on the Lorentzian formulation introduced in (3). By substituting the synthetic parameters into the model, we obtain smooth and physically consistent Q-V data that resemble realistic battery aging patterns.

C. Adaptive Autoencoder Learning with Dynamic Loss Weighting

To effectively leverage both labeled and unlabeled synthetic data, an adaptive autoencoder-based learning framework is developed to jointly handle feature extraction, input reconstruction, and RUL prediction. As illustrated in Fig. 3, the model employs a dual-head architecture comprising a shared encoder and two task-specific heads. The overall formulation follows standard autoencoder design principles [38], with task-specific adaptation as follows:

Encoder: it maps each input sample $x \in \mathbb{R}^{d_{in}}$, where d_{in} denotes the dimensionality of the input feature vector, into a latent representation $H \in \mathbb{R}^m$, where $m < d_{in}$,

$$H = f_{enc}(x, \theta_{enc}) \tag{9}$$

Decoder: it reconstructs the input from the latent representation, generating the reconstructed data $\hat{x} \in \mathbb{R}^{d_{in}}$,

$$\hat{x} = f_{dec}(H, \theta_{dec}) \tag{10}$$

Predictor: it predicts battery RUL based on the latent representation, using only labeled samples,

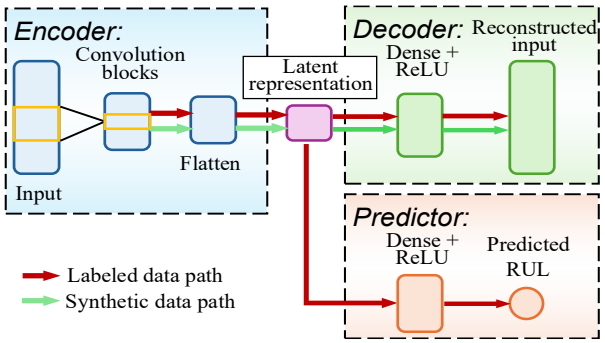

Fig. 3. Adaptive autoencoder framework with a shared encoder and dual heads for reconstruction and RUL prediction. The outputs are optimized using a dynamically weighted loss (see Fig. 1), which is not explicitly shown here for clarity.

$$\hat{y} = f_{pred}(H, \theta_{pred}) \tag{11}$$

Here, θ_{enc}, θ_{dec}, and θ_{pred} denote the trainable parameters of the encoder, decoder, and predictor, respectively.

This architecture enables the encoder to learn shared latent features from both real and synthetic inputs. Unlike conventional decoupled designs, reconstruction and regression are optimized jointly. The reconstruction loss guides the encoder to capture structural information, while the prediction loss provides task-specific supervision:

$$L_{recon}(x, \hat{x}) = \frac{1}{N}\sum_{i=1}^{N}\|x_i - \hat{x}_i\|_2^2 \tag{12}$$

$$L_{RUL}(y, \hat{y}) = \frac{1}{N}\sum_{i=1}^{N}\|y_i - \hat{y}_i\|_2^2 \tag{13}$$

where N denotes the mini-batch size.

A fixed weighting between reconstruction and prediction losses is often suboptimal, as their relative importance evolves during training. In early stages, emphasizing reconstruction encourages robust and generalizable latent features, especially when using unlabeled synthetic data. Later, prediction accuracy becomes more critical. To adaptively balance these objectives, a dynamic loss weighting mechanism is introduced. The total loss at training epoch e is defined as:

$$L_{total,e} = L_{RUL} + \beta_e \cdot L_{recon} \tag{14}$$

where β_e is a dynamically adjusted weighting coefficient balances the contribution of reconstruction and prediction losses at epoch e. It is updated based on the relative magnitude of the two losses:

$$\Delta L_e = \frac{L_{recon} - L_{RUL}}{L_{RUL} + \epsilon}, \quad \epsilon = 1 \times 10^{-8} \tag{15}$$

This relative loss difference (ΔL_e) reflects the model's current training emphasis: a positive ΔL indicates that the reconstruction loss dominates, prompting increased reconstruction supervision, and a negative value suggests a greater focus on prediction accuracy. The loss weighting coefficient is then adaptively updated at each training epoch using an exponential strategy:

$$\beta_{e+1} = \beta_e \cdot exp(\eta \cdot \Delta L_e) \tag{16}$$

where η controls the rate of adjustment. A larger η leads to more aggressive adjustments. To maintain training stability, the updated loss weight (β_{e+1}) in the next epoch is further bounded within a predefined range:

$$\beta_{e+1} = \max(\beta_{min}, \min(\beta_{max}, \beta_{e+1})) \tag{17}$$

In this formulation, β_{min} and β_{max} act as soft bounds to prevent the loss weight from becoming excessively large or small during training. Their values (0 and 10, respectively) are empirically selected to ensure that neither loss dominates training for extended periods.

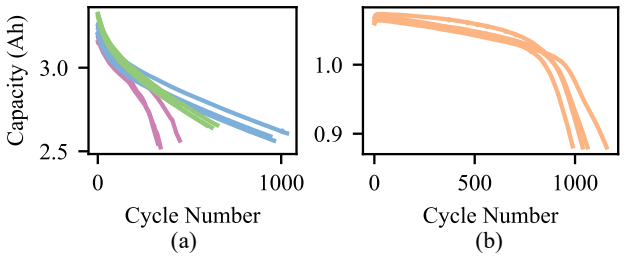

Fig. 4. Representative capacity degradation curves: (a) Tongji dataset, CC-CV charging at 25 °C (purple), 35 °C (blue), and 45 °C (green); (b) MIT dataset, two-stage fast-charging protocol (5.3 C-4 C) at 30 °C.

III. Battery Datasets Description

The proposed method is evaluated on two public datasets with distinct characteristics: Tongji dataset [33] and MIT dataset [8]. The Tongji dataset includes commercial 18650-format lithium-ion batteries with $Li(Ni_{0.83}Co_{0.11}Mn_{0.07})O_2$ (NCM) cathodes and graphite-silicon (Si) composite anodes, whose nominal capacity is 3.5 Ah. These batteries were charged under constant-current-constant-voltage (CC-CV) protocols at 25 °C, 35 °C, and 45 °C, serving as the main benchmark for evaluating performance under temperature variations. Specifically, a constant-current (CC) charging was first applied until the cell voltage reached 4.2 V, followed by a constant-voltage (CV) phase where the voltage was held until the current dropped below C/20. Discharging was conducted under a CC mode, terminating when the voltage decreased to 2.5 V.

In contrast, the MIT dataset consists of APR18650M1A lithium iron phosphate (LFP) batteries from A123 Systems with $LiFePO_4$ cathodes and graphite anodes, whose nominal capacity is 1.1 Ah. The batteries were cycled with a two-stage 5.3 C-4 C fast-charging protocol at 30 °C, followed by a standard 1 C CC-CV phase, where C-rate refers to a battery's charge/discharge rate of a battery relative to its nominal capacity, which is calculated as charge or discharge current (A)/battery nominal capacity (Ah). For the example of a battery with the nominal capacity of 10 Ah, 1 C corresponds to the current of 10 A at which a battery is fully charged or discharged in one hour. This dataset represents a markedly different chemistry and degradation pattern. The corresponding capacity degradation curves are shown in Fig. 4. The Tongji batteries exhibit temperature-sensitive degradation behaviors, while the MIT batteries

demonstrate slower capacity decay due to the chemical stability of LFP (see details in *Supplementary Note 1*).

In this work, the Tongji dataset serves as the primary benchmark for evaluating model robustness under diverse operating conditions, while the MIT dataset is used as an auxiliary case study to examine cross-chemistry applicability under a distinct fast-charging protocol with different temperature.

IV. Results and Discussions

A. Model Setups and Training Configurations

The model takes normalized Q-V curves interpolated to a fixed length of 1000 as input. Charge values are linearly interpolated over evenly spaced voltage points within a predefined range during the CC charging phase, ensuring consistent input dimensionality across samples.

An adaptive autoencoder is used to extract latent degradation features from both real and synthetic data. The encoder includes three 1D convolutional blocks with ReLU and max pooling, progressively reducing resolution while capturing structural features. CNNs have been widely adopted in battery health modeling due to their ability to extract regularities from electrochemical signals such as Q-V and IC curves [39]. Detailed layer configuration and shape transitions are summarized in Table I.

All models are implemented in PyTorch 2.2.0 (Python 3.9) and trained using the Adam optimizer (learning rate = 1×10^{-4}, batch size = 256) for 1500 epochs on an RTX 4090 GPU.

Model accuracy is evaluated using root mean squared error (RMSE) on the test set:

$$RMSE = \sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}(y_i - \hat{y}_i)^2} \tag{18}$$

where N_t represents the total number of test samples. To reduce randomness, each experiment is repeated three times, and results are reported as mean ± standard deviation (STD).

TABLE I
Network Architecture and Data Shape Transitions

Stage	Operation Description	Output Shape
Input	Q-V curve (length = 1000)	[B, 1000]
Encoder (3 Blocks)	Three blocks of 1D convolution (kernel size = 3), ReLU activation, and max pooling (kernel size = 5, stride = 5); output channels: 16 → 32 → 64	[B, 64, 8]
Flatten	Flatten the feature map	[B, 512]
Latent Projection	Fully connected layer, followed by ReLU	[B, 64]
RUL Regression	Fully connected layers: 64 → 32 → 1 with ReLU activation	[B, 1]
Q-V Reconstruction	Fully connected layers: 64 → 512 → 1000 with ReLU activation	[B, 1000]

B. Performance Comparison and Ablation Study

The proposed framework is evaluated through three stages: (1) comparison with classical supervised and semi-supervised methods; (2) comparison of different synthetic data generation strategies; and (3) ablation of key components. To ensure fairness, all methods adopt identical encoder architectures, input formats, and sample sizes where applicable.

1) Comparison with Supervised and Semi-Supervised Methods

This subsection compares the proposed method with supervised and semi-supervised baselines under limited labeled data. The baselines are categorized by how they utilize available data. Supervised models, including EN [8], SVR [9], MLP [10], LSTM [14], and CNN, rely solely on labeled samples and directly perform regression. Semi-supervised methods, such as self-learning [21], co-training [22], and label propagation [23], leverage unlabeled data via pseudo-labeling: self-learning iteratively refines predictions; co-training exchanges pseudo-labels between two models; label propagation diffuses label information over a similarity graph.

In contrast, the proposed method addresses the scarcity of real unlabeled data by generating physically consistent synthetic samples, enhancing training diversity without requiring additional experimental collection. It further integrates structure-aware feature learning and predictive supervision into a unified framework.

All methods are evaluated on the same dataset of 702 cycles, with 150, 300, 450, or 600 labeled samples, and the rest treated as unlabeled. CNN-based baselines (including the proposed method) share an identical encoder, while traditional models (EN, SVR, MLP, LSTM) operate directly on normalized Q-V inputs. Experimental results are summarized in Table II. Further implementation details provided in *Supplementary Note 3*.

As shown in Table II, the proposed method consistently achieves the lowest test RMSE across all levels of labeled data availability. Under highly limited supervision (150 labels), it outperforms the best semi-supervised baseline (label propagation) by 16%. At 600 labels, the RMSE is reduced by nearly 40%, demonstrating strong scalability with increased supervision.

Across supervised baselines, RMSEs remain relatively high and exhibit only moderate improvement as labeled data increases. The LSTM shows the highest RMSE under all conditions, indicating significant training instability due to high computational costs and the mismatch between sequential modeling assumptions and the weak temporal dependency in Q-V curve inputs. The SVR performs poorly as well, reflecting challenges in applying kernel-based methods to high-dimensional, nonlinearly distributed degradation features. Although the EN outperforms both the SVR and the LSTM, its purely linear modeling capacity remains insufficient to capture the complexity of battery degradation. The CNN baseline converges faster and is relatively insensitive to labeled data quantity, yet its standalone use still yields substantial prediction errors, suggesting that extracting local structure alone is inadequate without additional regularization or unlabeled data exploitation. The MLP baseline shows moder-

TABLE II
AVERAGE TEST RMSE (± STD CYCLES) ACROSS METHODS UNDER VARYING LABELED DATA QUANTITIES

Available labeled sample		150	300	450	600
Supervised learning methods	EN [8]	37.05 ± 2.55	36.61 ± 2.43	36.70 ± 2.49	36.86 ± 2.44
	SVR [9]	116.22 ± 7.96	111.13 ± 8.4	107.44 ± 11.21	103.97 ± 11.21
	MLP [10]	31.35 ± 1.62	26.97 ± 2.14	23.35 ± 0.93	23.29 ± 1.40
	LSTM [12]	150.84 ± 7.96	122.52 ± 8.49	122.44 ± 8.49	118.17 ± 9.59
	CNN	34.22 ± 0.46	33.40 ± 0.92	33.36 ± 0.96	33.30 ± 0.72
Semi-supervised learning methods	Self-learning [19]	31.34 ± 0.74	25.61 ± 0.82	24.24 ± 1.18	22.60 ± 0.26
	Co-training [20]	34.18 ± 1.62	30.54 ± 2.51	30.42 ± 5.86	28.67 ± 7.61
	Label propagation [21]	29.68 ± 3.21	24.82 ± 3.81	22.69 ± 2.94	24.99 ± 2.55
The proposed methods		**24.97** ± 7.34	**17.83** ± 8.62	**14.31** ± 8.11	**13.69** ± 8.69

ate improvement over linear models such as EN, it still lacks the hierarchical feature abstraction capacity offered by deeper neural architectures.

Among semi-supervised baselines, label propagation shows competitive performance by exploiting geometric consistency in the latent space. However, it lacks task-specific supervision and is prone to error accumulation when feature distributions are complex. Self-learning and co-training introduce task-awareness through pseudo-labeling but remain vulnerable to noise propagation, particularly under low-label scenarios, where pseudo-label quality is poor.

By contrast, the proposed method generates 4000 synthetic data, explicitly integrates structure-aware reconstruction and physics-informed synthetic data within a unified framework, enhancing training stability, predictive accuracy, and feature consistency across varying available data. While the proposed method consistently achieves the lowest RMSE across all label levels, it is worth noting that slightly higher variance is observed under the most limited supervision (150 labels). This is primarily attributed to the small amount of labeled data and the inclusion of synthetic data, which may introduce minor stochastic effects due to distributional sampling while still being beneficial for improving average accuracy. Nevertheless, even with this variability, the proposed method outperforms all baseline methods in terms of mean performance, highlighting its robustness under data-scarce conditions.

2) Evaluating Synthetic Data Quality and Generation Strategy

To validate the effectiveness of the proposed physics-informed synthetic data generation strategy, we systematically compare it against a conventional GAN-based data-driven method and an autoencoder framework without synthetic data. The objective is to assess whether incorporating physics prior enables improved downstream RUL prediction performance and better preservation of battery degradation characteristics.

All the models are trained using 400 available real samples. When synthetic data is incorporated, 4000 additional samples are generated under identical configurations. The GAN baseline adopts a standard three-layer fully connected generator and discriminator with ReLU and Sigmoid activations, respectively. No extensive hyperparameter tuning is applied to ensure fairness. As shown in Fig. 5, models using physics-informed synthetic data achieve the lowest RMSE and exhibit the most compact prediction error distributions, indicating superior generalization and stronger alignment with degradation patterns.

Distribution Similarity Analysis: A high-quality synthetic dataset should closely mirror the statistical distribution of real data. Distribution similarity is qualitatively assessed by applying t-distributed stochastic neighbor embedding (t-SNE) to project real, physics-informed synthetic, and GAN-generated data into a two-dimensional space, as shown in Fig. 6(a). Distinct differences can be observed among the datasets. The GAN-generated data points scatter and diverge significantly from the distribution of real data, forming fragmented clusters that indicate statistical misalignment. In contrast, physics-informed synthetic samples are tightly intermixed with real data, demonstrating superior capability to capture intrinsic battery degradation patterns.

Physical Consistency Analysis: Beyond statistical similarity, maintaining physical characteristics is critical for synthetic data. IC curves derived from synthetic data and real data are compared to assess physical fidelity. As illustrated in Fig. 6(b), IC curves derived from GAN-generated data show noise and artificial peaks. In comparison, those from physics-informed synthetic data are smooth, with consistent peak positions and trend continuity, closely matching real data and reflecting stronger physical fidelity and better alignment with those from real data.

While the proposed method uses a PCA-GMM sampling strategy to generate synthetic data, we acknowledge that this approach assumes a Gaussian latent structure, which may not fully capture all possible degradation behaviors. However, in our case, GMM operates on a compressed and physically interpretable latent space derived from Lorentzian-fitted Q-V parameters. Within this structured space, GMM serves as a physics-informed oversampling tool that expands existing degradation patterns without distorting the underlying data distribution, making it highly suitable for limited-data scenarios.

It is worth noting that GMM is not intended to generate entirely new degradation modes that are absent from the training data. Rather, it densifies the existing parameter distribution, assuming that diverse degradation patterns are already implicitly represented in the input samples. For unseen degradation behaviors, additional strategies such as transfer

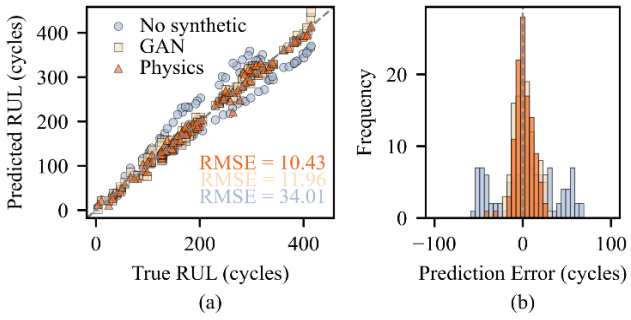

Fig. 5. Results of RUL predictions under different synthetic data strategies: (a) prediction distributions; (b) histogram of prediction errors illustrating reduced error variance.

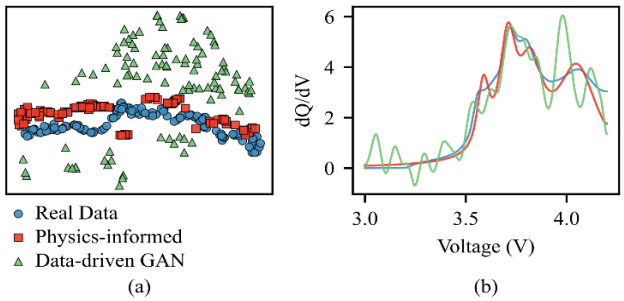

Fig. 6. Examples of data generation performance: (a) t-SNE visualization of real and synthetic Q-V samples; (b) Derived IC curve comparisons among real, GAN-generated, and physics-informed generated data.

learning or domain adaptation would be needed.

From a computational standpoint, the physics-informed method offers high efficiency and stability. Parameter identification takes approximately 0.093 seconds per cycle on CPU, and once fitted, synthetic generation (e.g., 4000 samples) can be completed within two seconds. In contrast, GAN-based data-driven generation, while capable of modeling more complex distributions, requires adversarial training that is often unstable, sensitive to hyperparameter tuning, and computationally demanding. Although GANs produce samples quickly once trained, the overall pipeline tends to involve a much higher cost in terms of tuning effort, reproducibility, and resource requirements. In addition, although more expressive generative models such as VAEs could capture more complex distributions, they are conceptually similar to GANs in being purely data-driven. These models often require large-scale training datasets and offer limited physical interpretability. As our study already includes a representative GAN-based baseline for comparison, we chose not to introduce VAEs as an additional baseline to avoid redundancy.

3) Ablation Study of Model Components

To quantify the contribution of each key design within the proposed framework, a systematic ablation study is conducted. Three key components are evaluated: (1) the dual-head decoder for joint reconstruction and prediction, (2) the dynamic loss weighting mechanism for balancing training objectives; and (3) the use of physics-informed synthetic data for distribution enhancement. Each component is individually disabled while keeping the others intact. The following variants are tested:

TABLE III
ABLATION STUDY OF KEY DESIGN COMPONENTS WITHIN THE PROPOSED FRAMEWORK

Model	M1	M2	M3	Test RMSE (cycles)
Full Proposed Method	✓	✓	✓	**13.06 ± 7.41**
w/o Reconstruction				35.19 ± 11.91
w/o Dynamic β	✓		✓	16.52 ± 9.05
w/o Physics-informed Data	✓	✓		31.84 ± 12.07

Notes: M1-Reconstruction Module (Dual-head); M2-Dynamic Loss Weight Module (β); and M3- Physics-informed Synthetic Data Module.

- **w/o Reconstruction**: The reconstruction head is removed, disabling both structural regularization and the use of unlabeled synthetic data. The model degenerates into a purely supervised setting, relying solely on limited labeled data.
- **w/o Dynamic loss weighting**: The model uses a fixed loss weight ($\beta = 1$), preventing adaptive adjustment during training.
- **w/o Physics-informed data**: Synthetic data is excluded, restricting training distribution and model generalization

The ablation settings and corresponding test RMSE values are summarized in Table III, where each core module is indicated as enabled (✓) and highlighted in green.

Disabling any component leads to clear performance degradation. The absence of reconstruction causes the largest RMSE increase (from 13.06 cycles → 35.19 cycles), highlighting the importance of structure-aware supervision and access to synthetic samples. Without the dynamic β mechanism, RMSE increases moderately (16.52 cycles), suggesting that static weighting fails to adapt to shifting training dynamics. Excluding physics-informed data also severely harms performance (RMSE 31.84 cycles), indicating that synthetic augmentation plays a vital role in preventing overfitting and enriching feature learning.

Compared to the supervised-only baseline (w/o Reconstruction), the full proposed method achieves over 60% reduction in RMSE. While exact attribution is non-trivial due to interactions among modules, the results indicate that the reconstruction and physics-informed data modules contribute the most substantial gains (Δ+22.1 and Δ+18.8 cycles), while the dynamic weighting provides a smaller yet consistent benefit (Δ+3.5). These findings confirm that the overall improvement stems from the synergistic integration of structural learning, physics-informed data generation, and adaptive objective balancing, rather than from any single dominant factor.

C. Effect of Synthetic Data Quantity on RUL Prediction

While physics-informed synthetic data augmentation has demonstrated clear benefits for battery RUL prediction, the quantity of synthetic samples introduced during training can significantly influence model performance. Insufficient synthetic data may limit training diversity and representation, whereas excessive synthetic data may introduce distributional biases, redundancy, or unrealistic patterns, ultimately impair-

TABLE IV
TEST RMSE (± STD CYCLES) ACROSS SYNTHETIC DATA QUANTITIES AND LABELED DATA LEVELS

Labeled data	S1-0	S2-500	S2-1000	S2-2000	S2-4000	S2-6000	S2-8000
39	36.85 ± 8.53	36.41 ± 4.65	36.27 ± 4.82	35.68 ± 5.41	36.16 ± 4.58	36.41 ± 4.69	**34.93** ± 6.07
78	34.70 ± 5.17	34.32 ± 4.71	31.19 ± 7.50	32.39 ± 6.91	**30.53** ± 8.71	31.89 ± 6.38	33.23 ± 6.12
156	35.12 ± 5.54	32.17 ± 7.45	30.69 ± 8.32	27.12 ± 10.50	**24.87** ± 7.41	28.49 ± 9.44	32.84 ± 7.15
234	34.16 ± 5.59	31.65 ± 6.96	27.26 ± 12.19	20.04 ± 11.05	**19.28** ± 7.45	22.94 ± 14.09	29.09 ± 10.85
312	33.90 ± 5.14	27.75 ± 10.56	22.51 ± 11.79	18.35 ± 12.61	**17.76** ± 8.24	19.33 ± 10.67	24.74 ± 12.31
390	33.04 ± 5.92	29.82 ± 8.59	22.86 ± 11.84	19.37 ± 11.21	**14.44** ± 7.64	17.06 ± 12.19	20.64 ± 13.34
468	33.32 ± 5.91	23.42 ± 12.25	24.01 ± 13.35	15.59 ± 9.52	**14.59** ± 8.55	15.61 ± 9.84	20.34 ± 12.30
546	34.05 ± 5.39	23.16 ± 11.68	21.19 ± 12.11	15.48 ± 10.04	**13.53** ± 8.59	15.43 ± 10.67	21.54 ± 13.27
624	33.33 ± 5.93	23.85 ± 11.11	17.27 ± 9.96	13.50 ± 5.79	13.74 ± 9.40	**13.32** ± 6.66	17.19 ± 11.02
702	30.15 ± 8.48	24.25 ± 9.99	18.44 ± 11.15	14.12 ± 7.74	**13.73** ± 8.77	14.72 ± 9.29	16.83 ± 11.44

Note: S1-0 denotes supervised learning without synthetic data (N_{syn} = 0), while S2-*x* indicates the proposed method with *x* synthetic training samples.

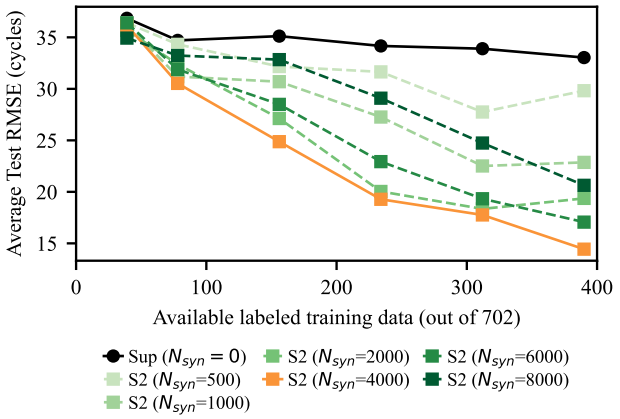

Fig. 7. Test RMSE comparison under varying synthetic data quantities (N_{syn}) across different labeled data settings

ing generalization.

In this study, a systematic evaluation is conducted to assess the impact of synthetic sample quantity on model performance under varying availability of labeled data. The number of synthetic samples, denoted as N_{syn}, is varied across 500, 1000, 2000, 4000, 6000, and 8000. For each setting, the number of labeled real cycles is fixed based on predefined label ratios relative to the full dataset (702 cycles), covering scenarios from highly limited (e.g., 39 cycles) to fully available conditions. For each combination of labeled and synthetic data quantities, the models are trained over three random seeds, and the average test RMSE is computed to ensure statistical robustness. A subset of the results (for labeled samples ≤ 400) is visualized in Fig. 7, while the complete numerical results are summarized in Table IV.

As shown in Fig. 7 and Table IV, introducing synthetic data significantly improves model performance compared to supervised learning without augmentation. For instance, with only 78 labeled cycles, the baseline supervised model yields an RMSE of 34.70 ± 5.17 cycles, while adding 4000 synthetic samples reduces the RMSE to 30.53 ± 8.71 cycles, representing approximately a 12% relative improvement.

The performance gain becomes even more pronounced with larger labeled datasets. For example, with 390 labeled cycles, adding 4000 synthetic samples reduces the RMSE from 33.04 ± 5.92 (baseline) to 14.44 ± 7.64, corresponding

TABLE V
TEST RMSE UNDER DIFFERENT GENERALIZATION SCENARIOS

Scenario	Dataset	Test temperature (°C)	Model	Tet RMSE (cycles)
Case I	Tongji	35	Baseline	20.52 ± 6.57
		35	Proposed	**11.96** ± 7.59
		45	Baseline	17.04 ± 4.23
		45	Proposed	**12.72** ± 5.72
Case II	Tongji	25	Baseline	33.40 ± 8.12
		25	Proposed	**15.16** ± 7.64
	MIT	30	Baseline	9.64 ± 1.03
		30	Proposed	**8.84** ± 2.46

to more than a 56% reduction in prediction error. This highlights the strong synergistic effect between physics-informed synthetic data augmentation and increased available data.

Model performance generally improves as the number of synthetic samples increases, with the lowest RMSE achieved when N_{syn} = 4000. For instance, with 390 labeled samples, N_{syn} = 4000 achieves an RMSE of 14.44 ± 7.64, while N_{syn} = 8000 yields a higher RMSE of 20.64 ± 13.34. Beyond the optimal point of approximately 4000 samples, further synthetic data augmentation leads to diminishing or even negative gains in RUL prediction performance. This trend suggests that introducing excessive synthetic samples may dilute the signal-to-noise ratio, introduce redundancy, or induce distributional shifts that degrade model generalization.

Across most available data conditions, the optimal synthetic data quantity is consistently around 4000 samples. This finding suggests that moderate augmentation enhances the training distribution and feature learning capacity, whereas excessive synthetic data may destabilize model training.

D. Robustness Validation under Varying Temperatures and Datasets

To evaluate the practical applicability of the proposed framework, we conduct robustness tests under two generalization settings. The first (temperature generalization, Case I) assesses performance across different ambient temperatures (35 °C and 45 °C) within the Tongji dataset [33]. The second (dataset generalization, Case II) examines transferability across datasets with differing chemistries and charging protocols, training is performed on the Tongji dataset at 25 °C, and

testing on the MIT dataset [8], which involves LFP batteries under fast-charging protocols at 30 °C.

Two model variants are compared: (1) a supervised baseline trained solely on 400 labeled real samples, and (2) the proposed physics-informed generation framework, incorporating a joint reconstruction–prediction architecture trained on 400 labeled and 4,000 synthetic samples. Test RMSEs for both settings are summarized in Table V.

As shown in Table IV, the proposed method consistently outperforms the supervised baseline across all conditions. Under temperature variations (35 °C and 45 °C), it achieves approximately 30-40% lower RMSE, demonstrating strong resilience to thermal shifts. This improvement is attributed to the regularization and enhanced feature diversity introduced by reconstruction learning and synthetic data. In the cross-dataset setting, the proposed model achieves a lower RMSE (8.84 cycles) compared to the baseline (9.64 cycles), confirming its ability to generalize across different battery chemistries and operating protocols. This confirms that the physics-informed synthetic data and structure-aware training strategy facilitate transferable feature learning across different cycling protocols and battery chemistries.

E. Future Extensions

While the proposed method focuses on data-scarce RUL prediction using complete Q-V profiles, several directions remain open for future research. In practical applications such as electric vehicles, only partial Q-V curves may be available due to constrained state of charge (SOC) ranges or limited operating conditions. In principle, the proposed framework is inherently extensible to such scenarios; however, future work is needed to investigate localized representation learning from partial Q-V segments and adaptive modeling techniques under incomplete information.

Another important direction is the explicit integration of temperature effects. Although the current work demonstrates robustness across different temperatures, temperature is not directly encoded in the generative or predictive process. Since temperature plays a critical role in influencing degradation patterns, incorporating it as a conditioning variable in the synthetic data generation or learning process may enable more purposeful data augmentation and enhance cross-condition generalization. Such efforts could benefit from physics-guided constraints or temperature-aware latent representations.

In addition, cross-domain generalization to batteries with distinct chemistries and degradation patterns remains a critical challenge. This may be addressed by incorporating transfer learning, meta-learning, or other domain adaptation strategies to enhance model robustness when faced with unseen or heterogeneous datasets.

Further refinements to the dynamic loss weighting mechanism may also be explored. Specifically, the exponential adjustment rule introduced in this work could be systematically evaluated through sensitivity analysis or replaced by more principled balancing strategies to improve training interpretability and stability.

Lastly, recent advances in predictive modeling, such as Transformer-based regressors and physics-informed neural networks (PINNs), may be integrated into the current framework as modular components. These architectures offer potential benefits in handling sequential inputs or enforcing physics-based constraints, and could be considered in future work to further enhance model performance in specific contexts, such as sequence modeling or physics-grounded generalization.

V. Conclusion

This paper has proposed a framework that combines physics-informed synthetic data generation with an adaptive autoencoder-based neural network for battery RUL prediction under limited labeled data. By integrating synthetic data augmentation, a dual-head autoencoder, and dynamic loss weighting, the proposed method effectively leverages both labeled and unlabeled synthetic data to improve feature learning and predictive accuracy. Simulation results demonstrate over 60% error reduction compared to supervised-only baselines, and 15% improvement under low-label scenarios, confirming the effectiveness of synthetic augmentation and adaptive training. Robustness evaluations across different temperatures and datasets confirm strong generalization of the proposed framework.

References

[1] X. Hu, L. Xu, X. Lin, and M. Pecht, ‘Battery Lifetime Prognostics’, *Joule*, vol. 4, no. 2, pp. 310–346, Feb. 2020

[2] S. B. Sarmah *et al.*, ‘A Review of State of Health Estimation of Energy Storage Systems: Challenges and Possible Solutions for Futuristic Applications of Li-Ion Battery Packs in Electric Vehicles’, *J. Electrochem. Energy Convers. Storage*, vol. 16, no. 4, p. 040801, Nov. 2019

[3] M.-K. Tran *et al.*, ‘A comprehensive equivalent circuit model for lithium-ion batteries, incorporating the effects of state of health, state of charge, and temperature on model parameters’, *J. Energy Storage*, vol. 43, p. 103252, Nov. 2021

[4] A. Guha and A. Patra, ‘State of Health Estimation of Lithium-Ion Batteries Using Capacity Fade and Internal Resistance Growth Models’, *IEEE Trans. Transp. Electrification*, vol. 4, no. 1, pp. 135–146, Mar. 2018

[5] C. R. Lashway and O. A. Mohammed, ‘Adaptive Battery Management and Parameter Estimation Through Physics-Based Modeling and Experimental Verification’, *IEEE Trans. Transp. Electrification*, vol. 2, no. 4, pp. 454–464, Dec. 2016

[6] R. Li, N. D. Kirkaldy, F. F. Oehler, M. Marinescu, G. J. Offer, and S. E. J. O’Kane, ‘The importance of degradation mode analysis in parameterising lifetime prediction models of lithium-ion battery degradation’, *Nat. Commun.*, vol. 16, no. 1, p. 2776, Mar. 2025

[7] H. Rauf, M. Khalid, and N. Arshad, ‘Machine learning in state of health and remaining useful life estimation: Theoretical and technological development in battery degradation modelling’, *Renew. Sustain. Energy Rev.*, vol. 156, p. 111903, Mar. 2022

[8] K. A. Severson *et al.*, ‘Data-driven prediction of battery cycle life before capacity degradation’, *Nat. Energy*, vol. 4, no. 5, pp. 383–391, Mar. 2019

[9] M. A. Patil *et al.*, ‘A novel multistage Support Vector Machine based approach for Li ion battery remaining useful life estimation’, *Appl. Energy*, vol. 159, pp. 285–297, Dec. 2015

[10] Y. Wu, W. Li, Y. Wang, and K. Zhang, ‘Remaining Useful Life Prediction of Lithium-Ion Batteries Using Neural Network and Bat-Based Particle Filter’, *IEEE Access*, vol. 7, pp. 54843–54854, 2019

[11] Q. Zhao, X. Qin, H. Zhao, and W. Feng, 'A novel prediction method based on the support vector regression for the remaining useful life of lithium-ion batteries', *Microelectron. Reliab.*, vol. 85, pp. 99–108, June 2018

[12] Y. Fan *et al.*, 'State of health estimation of lithium-ion batteries based on the fusion of aging feature extraction and SSA-ELM machine learning algorithms', *Ionics*, vol. 31, no. 8, pp. 7897–7915, Aug. 2025

[13] W. Li *et al.*, 'A Holistic Electrothermal Profiles Online Sensing Method With Sparse Sensor System in Large-Format Battery Pack', *IEEE Trans. Ind. Electron.*, pp. 1–10, 2025

[14] K. Park, Y. Choi, W. J. Choi, H.-Y. Ryu, and H. Kim, 'LSTM-Based Battery Remaining Useful Life Prediction With Multi-Channel Charging Profiles', *IEEE Access*, vol. 8, pp. 20786–20798, 2020

[15] X. Li, D. Yu, V. Søren Byg, and S. Daniel Ioan, 'The development of machine learning-based remaining useful life prediction for lithium-ion batteries', *J. Energy Chem.*, vol. 82, pp. 103–121, July 2023

[16] K. Das, R. Kumar, and A. Krishna, 'Analyzing electric vehicle battery health performance using supervised machine learning', *Renew. Sustain. Energy Rev.*, vol. 189, p. 113967, Jan. 2024

[17] Y. Che, D.-I. Stroe, X. Hu, and R. Teodorescu, 'Semi-Supervised Self-Learning-Based Lifetime Prediction for Batteries', *IEEE Trans. Ind. Inform.*, vol. 19, no. 5, pp. 6471–6481, May 2023

[18] S. S. Y. Ng, Y. Xing, and K. L. Tsui, 'A naive Bayes model for robust remaining useful life prediction of lithium-ion battery', *Appl. Energy*, vol. 118, pp. 114–123, Apr. 2014

[19] L. Ma, J. Tian, T. Zhang, Q. Guo, and C. Hu, 'Accurate and efficient remaining useful life prediction of batteries enabled by physics-informed machine learning', *J. Energy Chem.*, vol. 91, pp. 512–521, Apr. 2024

[20] L. Ma, J. Tian, T. Zhang, Q. Guo, and C. Yung Chung, 'Enhanced battery life prediction with reduced data demand via semi-supervised representation learning', *J. Energy Chem.*, p. S2095495624006910, Oct. 2024

[21] F. Jiang *et al.*, 'An adaptive semi-supervised self-learning method for online state of health estimation of lithium-ion batteries', *Energy*, vol. 305, p. 132213, Oct. 2024

[22] C. Lin, J. Xu, and X. Mei, 'Improving state-of-health estimation for lithium-ion batteries via unlabeled charging data', *Energy Storage Mater.*, vol. 54, pp. 85–97, Jan. 2023

[23] 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 Trans. Transp. Electrification*, vol. 11, no. 2, pp. 5938–5949, Apr. 2025

[24] Y. Ren *et al.*, 'A novel state of health estimation method for lithium-ion battery pack based on cross generative adversarial networks', *Appl. Energy*, vol. 377, p. 124385, Jan. 2025

[25] L. Jiang, C. Hu, S. Ji, H. Zhao, J. Chen, and G. He, 'Generating comprehensive lithium battery charging data with generative AI', *Appl. Energy*, vol. 377, p. 124604, Jan. 2025

[26] X. Zhang, Y. Qin, C. Yuen, L. Jayasinghe, and X. Liu, 'Time-Series Regeneration With Convolutional Recurrent Generative Adversarial Network for Remaining Useful Life Estimation', *IEEE Trans. Ind. Inform.*, vol. 17, no. 10, pp. 6820–6831, Oct. 2021

[27] W. Yao, R. Lai, Y. Tian, X. Li, and J. Tian, 'State of Health Estimation of Lithium-Ion Batteries Using Data Augmentation and Feature Mapping', *IEEE Trans. Transp. Electrification*, vol. 11, no. 1, pp. 4895–4905, Feb. 2025

[28] M. Qasem *et al.*, 'Synthetic Data-Integrated Li-Ion Battery Modeling for eVTOL Energy Systems', *IEEE Access*, vol. 12, pp. 76329–76343, 2024

[29] S. Zhang, Z. Liu, and H. Su, 'Battery Early Prognostics Based on Pseudo Meta-Learning', *IEEE Trans. Ind. Inform.*, vol. 20, no. 10, pp. 11655–11665, Oct. 2024

[30] B. Gou, Y. Xu, and X. Feng, 'State-of-Health Estimation and Remaining-Useful-Life Prediction for Lithium-Ion Battery Using a Hybrid Data-Driven Method', *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 10854–10867, Oct. 2020

[31] W. Li, N. Sengupta, P. Dechent, D. Howey, A. Annaswamy, and D. U. Sauer, 'Online capacity estimation of lithium-ion batteries with deep long short-term memory networks', *J. Power Sources*, vol. 482, p. 228863, Jan. 2021

[32] G. Dong, W. Han, and Y. Wang, 'Dynamic Bayesian Network-Based Lithium-Ion Battery Health Prognosis for Electric Vehicles', *IEEE Trans. Ind. Electron.*, vol. 68, no. 11, pp. 10949–10958, Nov. 2021

[33] J. Zhu *et al.*, 'Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation', *Nat. Commun.*, vol. 13, no. 1, p. 2261, Apr. 2022

[34] X. Li, 'State of health estimation for Li-Ion battery using incremental capacity analysis and Gaussian process regression', 2020

[35] X. Li, J. Jiang, L. Y. Wang, D. Chen, Y. Zhang, and C. Zhang, 'A capacity model based on charging process for state of health estimation of lithium ion batteries', *Appl. Energy*, vol. 177, pp. 537–543, Sept. 2016

[36] M. Dubarry and D. Anseán, 'Best practices for incremental capacity analysis', *Front. Energy Res.*, vol. 10, p. 1023555, Oct. 2022

[37] C. M. Bishop, *Pattern recognition and machine learning*. in Information science and statistics. New York: Springer, 2006.

[38] Z. Yang, B. Xu, W. Luo, and F. Chen, 'Autoencoder-based representation learning and its application in intelligent fault diagnosis: A review', *Measurement*, vol. 189, p. 110460, Feb. 2022

[39] L. Ren, J. Dong, X. Wang, Z. Meng, L. Zhao, and M. J. Deen, 'A Data-Driven Auto-CNN-LSTM Prediction Model for Lithium-Ion Battery Remaining Useful Life', *IEEE Trans. Ind. Inform.*, vol. 17, no. 5, pp. 3478–3487, May 2021