

Article

Not peer-reviewed version

A Multi-Step RUL Prediction Method for Lithium-Ion Batteries Based on Multi-Scale Temporal Features and Frequency-Domain Spectral Interaction

[Ye Tu](#)^{*}, Shixiong Xu, Jie Wang, Mengting Jin

Posted Date: 20 March 2026

doi: 10.20944/preprints202603.1634.v1

Keywords: lithium-ion battery; remaining useful life; multi-step prediction; time-frequency fusion; spectral interaction; cross-gating



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A Multi-Step RUL Prediction Method for Lithium-Ion Batteries Based on Multi-Scale Temporal Features and Frequency-Domain Spectral Interaction

Ye Tu ^{1,2,3,*}, Shixiong Xu ², Jie Wang ^{1,4} and Mengting Jin ^{1,4}

¹ College of Engineering, Yuxi Normal University, Yuxi 653100, Yunnan, China

² School of Business, The University of New South Wales, Sydney, NSW, Australia

³ Yunnan Key Laboratory of Smart City in Cyberspace Security, Yuxi Normal University, Yuxi 653100, Yunnan, China

⁴ College of Mathematics and Physics, Yuxi Normal University, Yuxi 653100, Yunnan, China

* Correspondence: tuye@yxnu.edu.cn

Abstract

With the rapid development of new energy vehicles and energy storage systems, accurate prediction of the remaining useful life (RUL) of lithium-ion batteries is of great importance for predictive maintenance and operational safety. However, battery degradation during cycling usually exhibits multi-scale characteristics, including long-term degradation trends, stage-wise drifts, and stochastic disturbances, which makes existing methods still face significant challenges in multi-step forecasting and cross-domain generalization. To address this issue, this paper proposes a time–frequency fusion model for multi-step RUL prediction, termed TF-RULNet (Time-Frequency RUL Network). The model takes cycle-level feature sequences as input and consists of three components: a multi-scale temporal convolution encoder (MSTC) for parallel extraction of degradation cues at different temporal scales; a multi-head spectral interaction module (MHSI), which performs 1D-FFT along the temporal dimension for each head and further applies adaptive band-wise mask refinement to capture local spectral structures and hierarchical band patterns with a computational complexity of $O(L \log L)$; and a cross-gated fusion module (CGF), which generates gating signals from the summary of one domain to modulate the features of the other domain, thereby enabling dynamic balancing and complementary enhancement of time–frequency information. Experiments are conducted on the NASA dataset (B005/B007) for in-domain evaluation, and further cross-dataset tests from NASA to the Maryland dataset (CS-35/CS-37) are carried out to verify the robustness of the proposed model under distribution shifts. The results show that, compared with the strongest baseline PatchTST, TF-RULNet reduces RMSE and MAE by more than 38.23% and 50.51%, respectively, in cross-dataset generalization, while achieving an additional RMSE reduction of about 24% in in-domain prediction. In summary, TF-RULNet can effectively characterize the multi-scale time–frequency degradation patterns of batteries and improve cross-domain generalization, providing a high-accuracy and scalable modeling solution for practical battery health management and life prognostics.

Keywords: lithium-ion battery; remaining useful life; multi-step prediction; time–frequency fusion; spectral interaction; cross-gating

1. Introduction

With the rapid growth of new energy vehicles and grid-scale energy storage systems, lithium-ion batteries have become a critical energy unit whose operational safety and reliability are central to system design and maintenance [1–3]. Remaining useful life (RUL) describes the number of cycles or the remaining time from the current operating state until a predefined failure threshold is reached, and it serves as a key basis for predictive maintenance, operational optimization, and safety warning in battery health management systems [4,5]. However, battery degradation is jointly influenced by

charge/discharge conditions, environmental factors, and the evolution of internal electrochemical mechanisms [6]. As a result, degradation trajectories often exhibit strong nonlinearity and a complex mixture of stage-wise evolution and random fluctuations. Beyond the long-term trends of capacity fade and internal resistance increase, batteries may undergo regime shifts caused by changes in temperature, C-rate, and charging protocols, and may even exhibit short-term recovery (i.e., capacity regeneration) due to reversible processes such as relaxation and polarization decay [7]. Such a coupled multi-scale pattern of “trend–stage–disturbance” requires RUL prediction models to simultaneously capture long-range dependencies, remain sensitive to phase transitions, and be robust to noise, thereby posing higher demands on accuracy and generalization [8,9].

Existing RUL prediction approaches can be broadly categorized into model-driven methods and data-driven methods [8–10]. Model-driven approaches typically build electrochemical models, equivalent circuit models, or statistical degradation processes, offering interpretability advantages [10]. For example, Peng et al. combined an observer with an improved EKF to enhance SOC estimation under diverse operating conditions [11], while Zhang et al. introduced physics-informed constraints into degradation modeling to describe complex evolutionary behaviors [12]. Nevertheless, model-driven methods often rely on accurate parameters and strong prior assumptions, and their flexibility and generalization under highly variable operating conditions remain limited [13]. In contrast, data-driven methods learn degradation patterns from historical operating data [14], such as LSTM-based RUL prediction with feature selection [15], interpretable autoencoders for health indicator construction with similarity-based RUL inference [16], and federated learning frameworks for distributed scenarios [17]. However, many data-driven methods still emphasize local temporal relationships or learning within a single representation space, which can lead to instability and degraded generalization for long-window prediction and cross-condition transfer [18].

Recently, attention-based deep models—particularly Transformer architectures—have gained popularity for battery RUL prediction due to their capability of modeling global dependencies [19–23]. For instance, Wang et al. incorporated self-attention into an LSTM to strengthen the modeling of salient degradation information [20]; Rengarajan et al. proposed hybrid architectures combining Transformers with recurrent networks for RUL prediction [21]; and Aizpurua et al. introduced probabilistic forecasting into a Transformer framework for uncertainty-aware RUL estimation [22]. Despite these advances, several key issues remain in complex degradation scenarios: (i) *insufficient multi-scale awareness*—most models operate on a single temporal scale and struggle to capture coexisting long-term trends and short-term fluctuations [24]; (ii) *under-exploitation of frequency-domain information*—degradation signals carry informative spectral structures (low-frequency trends, mid-frequency regime shifts, and high-frequency disturbances), yet most methods remain predominantly time-domain [25]; (iii) *lack of dynamic time–frequency fusion*—time- and frequency-domain features are often fused via concatenation or static weighting, without adaptive adjustment across degradation stages [26]; and (iv) *high computational cost for long windows*—the quadratic complexity of self-attention, $\mathcal{O}(L^2)$, limits online deployment and resource-constrained applications.

To address these limitations, recent studies have explored multi-scale modeling, time–frequency representations, and dynamic fusion strategies. Li et al. proposed MC-CNN-TimesNet to enhance multi-scale dependency modeling [6]. Related multi-scale learning ideas have also been developed in other forecasting tasks, such as EnergyFormer [27], dual-branch multi-scale Transformers [28], and multi-scale spatiotemporal/ensemble frameworks [29,30]. For time–frequency modeling, Qian et al. integrated EMD decomposition and 2D signal representations into Transformers [31]; Duan et al. proposed a cross-domain time–frequency Mamba model [32]; Hao et al. enhanced frequency-domain representations via joint decomposition [33]; and Fan et al. and Wu et al. improved frequency-aware learning via loss-level constraints and hybrid time–frequency architectures, respectively [34,35]. For adaptive fusion, attention mechanisms and dynamic relational modeling have been used to mitigate the limitations of static fusion [36–38], and similar dynamic fusion designs have been validated in spatiotemporal prediction settings [39,40]. Nevertheless, for battery RUL prediction, there is still a

lack of a unified framework that simultaneously achieves (a) *lightweight yet effective spectral modeling* for long windows and (b) *cross-domain dynamic modulation* between time- and frequency-domain representations, thereby balancing accuracy, robustness, and deployment efficiency.

To this end, we propose a lightweight time–frequency fusion model, **TF-RULNet** (Time-Frequency Cross-Gated Network), for battery RUL prediction. TF-RULNet consists of a multi-scale time-domain convolutional encoder (MSTC), a multi-head spectral interaction module (MHSI), and a cross-gated fusion module (CGF). Specifically, MSTC extracts degradation cues at multiple temporal scales in parallel. MHSI performs 1D FFT only along the temporal dimension to capture global spectral structures with linear-logarithmic complexity, and employs learnable band-wise masks to obtain hierarchical representations for trend/stage/disturbance components. Finally, CGF adaptively modulates the contribution of time- and frequency-domain features via cross-gating, enabling stage-aware time–frequency collaboration.

The main contributions of this work are summarized as follows:

- **A lightweight time–frequency RUL prediction framework.** We develop TF-RULNet, a unified pipeline of multi-scale time-domain convolution (MSTC), spectral interaction (MHSI), and cross-gated fusion (CGF) to characterize multi-scale and non-stationary battery degradation.
- **Efficient multi-head spectral interaction with band-wise refinement.** Without incurring the $\mathcal{O}(L^2)$ cost of self-attention, MHSI leverages temporal 1D FFT to capture global spectral structures and employs learnable band masks to refine low-/mid-/high-frequency components in a hierarchical manner, achieving a favorable trade-off between efficiency and representation power.
- **Cross-domain, stage-adaptive time–frequency fusion.** CGF enables dynamic reweighting and modulation between time- and frequency-domain representations across degradation stages, improving robustness and cross-condition generalization beyond static concatenation or fixed-weight fusion.

2. Materials and Methods

2.1. RUL Definition and Dataset Description

2.1.1. Definition of Remaining Useful Life (RUL)

Remaining Useful Life (RUL) quantifies the number of remaining charge–discharge cycles (or the remaining time) that a battery can continue to operate while meeting performance requirements. In this study, the end-of-life (EOL) criterion is defined by a capacity-fade threshold: a battery reaches EOL when its measured capacity drops to 80% of the nominal capacity, i.e., $0.8 C_{\text{nominal}}$. Let $C(n)$ denote the capacity measured at cycle n , and let n_{current} be the current cycle index. We define the EOL cycle as

$$n_{\text{EOL}} = \min\{n \mid C(n) \leq \theta C_{\text{nominal}}\}, \quad \theta = 0.8, \quad (1)$$

and the ground-truth RUL (in cycles) at cycle n_{current} as

$$\text{RUL}(n_{\text{current}}) = \max(0, n_{\text{EOL}} - n_{\text{current}}). \quad (2)$$

Therefore, RUL is determined by the capacity degradation trajectory and is essentially a forward-looking estimate of “how many cycles remain until EOL”.

2.1.2. Dataset Description

To validate the applicability of the proposed model under different battery chemistries and operating conditions, this study selects two publicly available lithium-ion battery degradation datasets, namely the NASA dataset [41] and the CALCE dataset [42] from the University of Maryland. These two datasets differ in terms of cell type, material system, experimental protocol, and end-of-life (EOL) criterion, thereby providing strong support for evaluating both in-domain prediction and cross-domain generalization.

Table 1. Basic information of the selected battery datasets.

Dataset	Selected Batteries	Cell Type	Rated Capacity	Chemistry
NASA	B0005, B0006, B0007, B0018	Cylindrical	2 Ah	NCA/Graphite
CALCE	CS2-35, CS2-36, CS2-37, CS2-38	Prismatic	1.1 Ah	LCO/Graphite

The NASA dataset uses LG Chem 18650 cylindrical cells with a rated capacity of 2 Ah and an NCA/graphite chemistry. In this study, four cells, namely B0005, B0006, B0007, and B0018, are selected for investigation. The experiments were conducted at room temperature. During charging, a constant-current (CC) mode of 1.5 A was first applied until the terminal voltage reached 4.2 V, followed by a constant-voltage (CV) stage until the current dropped below 20 mA. During discharging, a constant current of 2 A was applied, and the cutoff voltage was set to 2.7 V, 2.5 V, or 2.2 V depending on the specific battery. The EOL criterion for this dataset is defined as the capacity fading to 70% of the rated capacity.

The CALCE dataset consists of 1.1 Ah prismatic lithium-ion cells with an LCO/graphite chemistry. In this study, four cells, namely CS2-35, CS2-36, CS2-37, and CS2-38, are selected as the research objects. The experiments were also carried out at room temperature. The charging protocol uses a 0.5C constant-current process up to 4.2 V, followed by a constant-voltage stage until the current decreases to 0.02 A. The discharge process is performed under a constant current of 2 A down to 2.7 V. The EOL criterion for this dataset is defined as the capacity fading to 80% of the rated capacity.

Overall, the NASA and CALCE datasets exhibit clear differences in cell form factor, capacity level, cathode chemistry, and EOL criterion. Based on these characteristics, this study conducts in-domain prediction experiments on the NASA dataset and further constructs a NASA→CALCE cross-dataset testing scenario to evaluate the robustness and cross-domain generalization ability of the proposed model under distribution shifts.

2.1.3. Construction and Selection of Health Indicators

To ensure that the input features can effectively characterize battery degradation states, the raw charge/discharge signals are first preprocessed, and cycle-level health indicators (HIs) are then constructed. The overall procedure consists of three steps: raw signal analysis, candidate HI construction, and correlation-based screening.

First, the variation patterns of the charge/discharge curves under different cycles are analyzed. The typical voltage–current curves of representative cells during the constant-current/constant-voltage (CC–CV) charging process and the discharge process are shown in Figure 1. As the cycle number increases, the duration of the constant-current stage changes gradually, the discharge voltage plateau shifts downward, and the current decay process in the constant-voltage stage also changes noticeably. These phenomena indicate typical aging characteristics, including capacity fading, aggravated polarization, and increased internal resistance. Furthermore, by examining the voltage–time curves under different cycles, it can be observed that, as aging progresses, the charging curves shift upward, the discharge curves shift downward, and the discharge plateau gradually shortens. These observations suggest that the raw charge/discharge signals contain rich degradation-related information and can therefore serve as an important basis for subsequent HI construction.

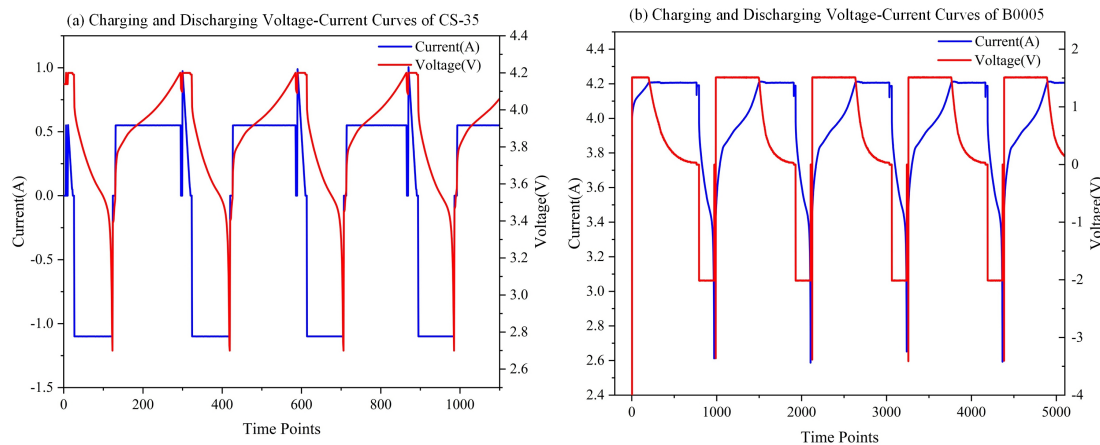


Figure 1. Typical voltage–current profiles of representative cells during the CC and CV phases.

Based on the above analysis, a set of cycle-level candidate HIs is extracted from the charge/discharge process to reflect degradation evolution. These features mainly include the constant-current charge time (CCCT), CC time percentage, CC area, discharge-stage duration (disCC Time), as well as other statistics that describe the temporal characteristics and stage-wise variations of the charge/discharge process. Compared with directly using raw curves, such cycle-level HIs have the advantages of lower dimensionality, reduced noise sensitivity, and clearer physical interpretability, making them more suitable for subsequent sliding-window modeling and multi-step RUL prediction.

To quantitatively evaluate the association between each candidate HI and the degradation state, the Spearman rank correlation coefficient (SCC) is adopted for correlation analysis. Unlike the Pearson correlation coefficient, which mainly characterizes linear relationships, the Spearman coefficient measures monotonic relationships and is therefore more suitable for the nonlinear degradation patterns commonly observed in battery aging. Let the candidate feature sequence be $X = \{x_1, x_2, \dots, x_n\}$ and the target degradation sequence be $Y = \{y_1, y_2, \dots, y_n\}$. The Spearman rank correlation coefficient is defined as

$$\rho_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \quad (3)$$

where d_i denotes the rank difference between the two sequences for the i -th sample, and n is the number of samples. The coefficient satisfies $\rho_s \in [-1, 1]$. The closer $|\rho_s|$ is to 1, the stronger the monotonic correlation between the feature and the degradation state.

According to the Spearman correlation analysis, the HIs that exhibit stronger correlations with the target degradation variable are preferentially retained as model inputs. The correlation results for NASA B005 are illustrated in Figure 2. It can be observed that temporal features, such as CCCT, CC time percentage, and disCC Time, generally show strong monotonic correlations with the degradation state, while area-related features (e.g., CC area), which reflect cumulative changes during the charging stage, can also effectively characterize the lifetime evolution trend. Therefore, the final input features of this study consist of cycle-level HIs with high correlation and clear physical meaning, providing a reliable data basis for the subsequent multi-scale temporal modeling, spectral interaction learning, and cross-gated fusion in TF-RULNet.

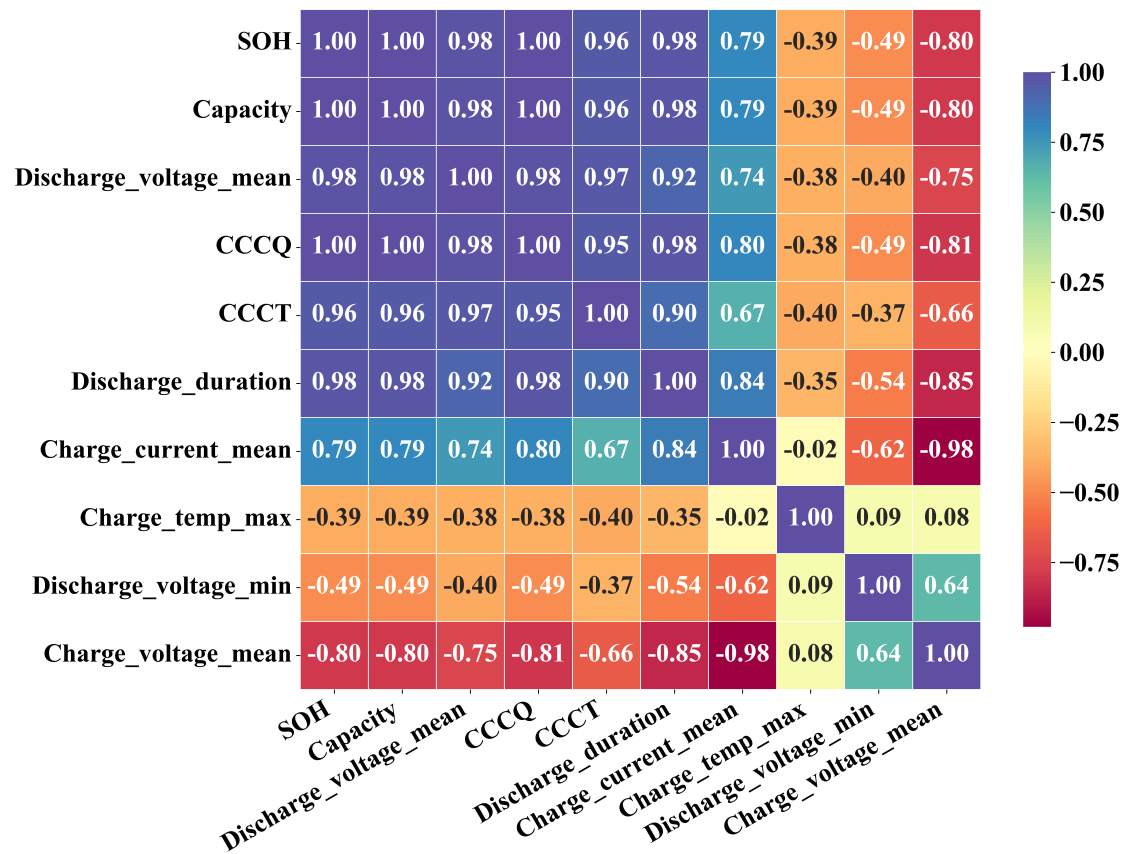


Figure 2. Spearman correlation heatmap of candidate health indicators for NASA B005.

2.2. Methodology

To capture the multi-scale degradation dynamics of lithium-ion batteries during cycling and to fuse complementary information across multiple signals (e.g., voltage, current, temperature, internal resistance, and capacity-related features), we propose a lightweight time–frequency RUL prediction model, **TF-RULNet** (Time-Frequency Cross-Gated Network). As illustrated in Figure 3, TF-RULNet consists of three components: (i) a multi-scale time-domain convolutional encoder (Multi-Scale Temporal Convolution, MSTC), (ii) a frequency-domain multi-head spectral interaction module (Multi-Head Spectral Interaction, MHSI), and (iii) a cross-gated time–frequency fusion module (Cross-Gated Fusion, CGF). The core idea is to first extract degradation cues at multiple temporal scales in the time domain, then capture global spectral structures and band-wise patterns with linear-logarithmic complexity in the frequency domain, and finally adaptively fuse both domains via cross-gating to obtain an optimal representation for RUL prediction.

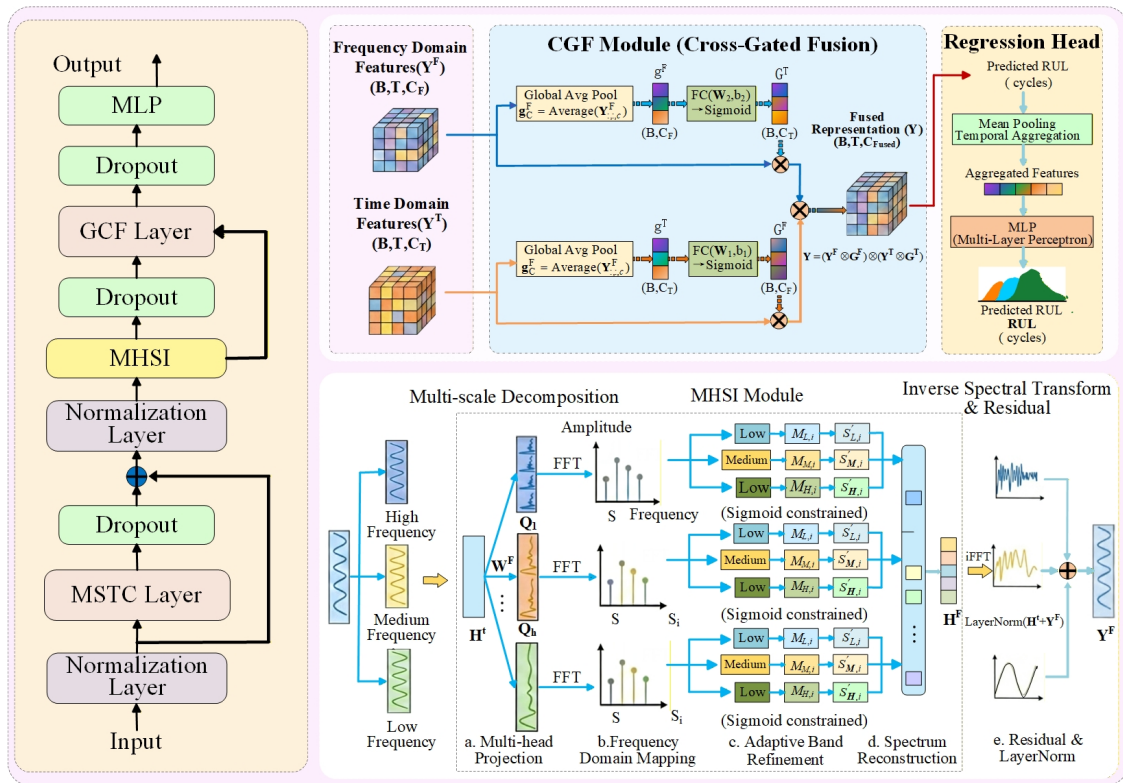


Figure 3. Overall architecture of the proposed TF-RULNet (Time-Frequency RUL Network). The model consists of a multi-scale time-domain convolution encoder (MSTC), a multi-head spectral interaction module (MHSI) with adaptive band refinement, a cross-gated fusion (CGF) module for dynamic time–frequency weighting, and a regression head for multi-step RUL prediction.

2.2.1. Task Definition and Notation

Consider the cycle-level feature sequence extracted from battery i during cycling:

$$\mathbf{x}_n^{(i)} \in \mathbb{R}^D, \quad n = 1, 2, \dots, N_i, \quad (4)$$

where n is the cycle index and D is the feature dimension per cycle. In this study, each cycle-level feature vector includes the following *process-aware* statistics:

$$\mathbf{x}_n^{(i)} = [t_{cc,n}^{(i)}, p_{cc,n}^{(i)}, a_{cc,n}^{(i)}, t_{disc,n}^{(i)}, \dots]^\top \in \mathbb{R}^D, \quad (5)$$

where t_{cc} is the constant-current (CC) charging time (s), p_{cc} is the CC time percentage (%), a_{cc} is the CC charge area (A·s, i.e., charge throughput during CC), and t_{disc} is the discharge CC time (s). The ellipsis “...” indicates that additional degradation-relevant features (e.g., capacity, internal resistance, temperature statistics, or IC/DV peak features) can be appended when available.

Using a sliding window of length L (set to $L = 10$ in this work), we construct the model input sample as

$$\mathbf{x}_n^{(i)} = [\mathbf{x}_{n-L+1}^{(i)}, \dots, \mathbf{x}_n^{(i)}] \in \mathbb{R}^{L \times D}, \quad n \geq L. \quad (6)$$

Under mini-batch training, the input tensor is

$$\mathbf{X} \in \mathbb{R}^{B \times L \times D}, \quad (7)$$

where B is the batch size (number of window samples), $L = 10$ is the window length (number of historical cycles), and $D = 4$ if only the four features in Eq. (5) are used.

Alignment principle.

The window end corresponds to the *current* cycle n , and the model predicts the RUL label $y_n^{(i)}$ at that cycle. Namely, TF-RULNet uses the most recent L cycles to estimate the remaining cycles from the current cycle until EOL.

Let $C^{(i)}(n)$ denote the capacity trajectory of battery i and C_{nom} its nominal capacity. With the EOL criterion defined by capacity reaching θC_{nom} (typically $\theta = 0.8$), the EOL cycle is

$$n_{\text{EOL}}^{(i)} = \min \left\{ n \mid C^{(i)}(n) \leq \theta C_{\text{nom}} \right\}, \quad \theta \in (0, 1). \quad (8)$$

Then the ground-truth RUL (in cycles) at cycle n is

$$y_n^{(i)} = \max \left(0, n_{\text{EOL}}^{(i)} - n \right). \quad (9)$$

Single-step vs. multi-horizon prediction.

By default, TF-RULNet performs single-step scalar regression:

$$\hat{y}_n^{(i)} = f_{\Theta} \left(\mathbf{X}_n^{(i)} \right), \quad \hat{y} \in \mathbb{R}^{B \times 1}. \quad (10)$$

If multi-horizon outputs are required (here we set the horizon to $H = 4$), we predict the RUL sequence of the next H cycles:

$$\hat{\mathbf{y}}_n^{(i)} = \left[\hat{y}_n^{(i)}, \hat{y}_{n+1}^{(i)}, \dots, \hat{y}_{n+H-1}^{(i)} \right] = f_{\Theta} \left(\mathbf{X}_n^{(i)} \right), \quad \hat{\mathbf{y}} \in \mathbb{R}^{B \times H}. \quad (11)$$

2.2.2. Multi-Scale Time-Domain Degradation Modeling (MSTC)

The implementation of multi-scale time-domain degradation modeling (MSTC) is illustrated in Figure 4. Battery degradation exhibits pronounced multi-scale characteristics. At a long-term scale, capacity gradually fades while internal resistance increases, reflecting irreversible aging. At a medium-term scale, degradation rates may drift across operating regimes (e.g., changes in temperature, C-rate, or charging strategy), and short-term recovery phenomena (capacity regeneration) may occur due to reversible processes such as relaxation and polarization decay. At a short-term scale, measurement noise and transient fluctuations induce local perturbations.

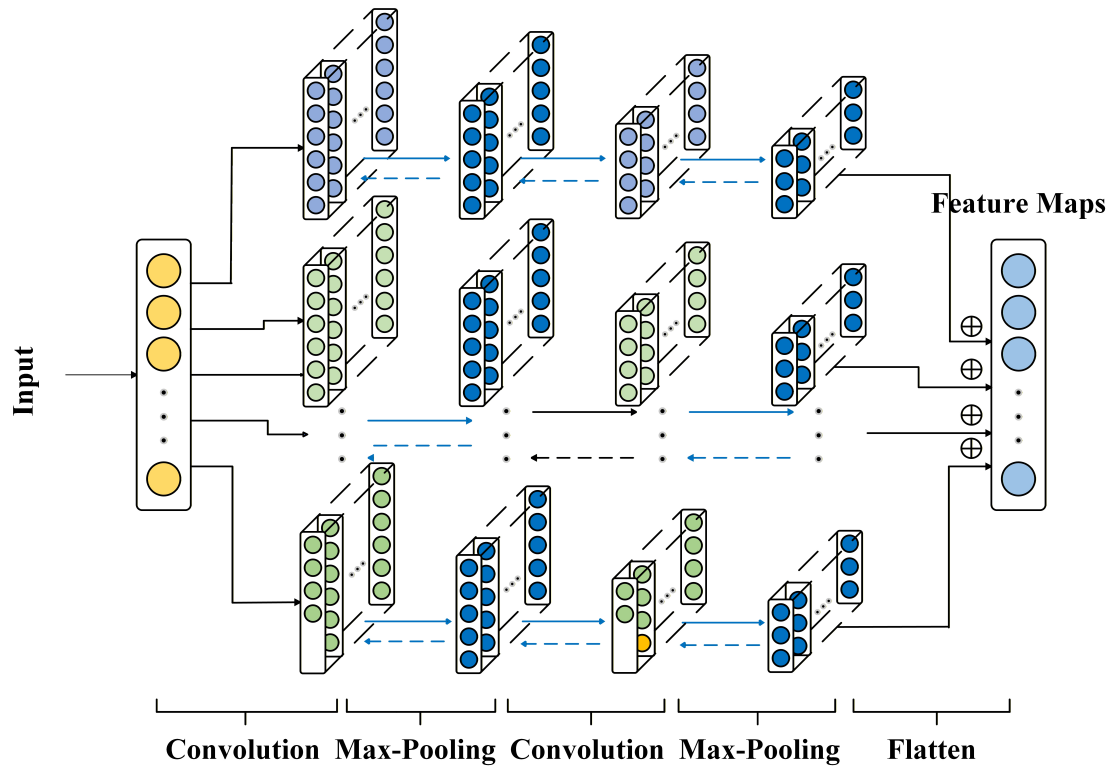


Figure 4. Overall architecture of the MSTC.

Given the input tensor $\mathbf{X} \in \mathbb{R}^{B \times L \times D}$ in Eq. (7), to apply 1D convolution along the cycle axis (length L), we first permute dimensions:

$$\mathbf{X}' \in \mathbb{R}^{B \times D \times L}. \quad (12)$$

MSTC employs K parallel convolution branches, where branch k uses a kernel size s_k and dilation d_k to cover different receptive fields:

$$\mathbf{F}^{(k)} = \phi\left(\text{Conv1D}\left(\mathbf{X}'; \mathbf{W}^{(k)}, s_k, d_k\right) + \mathbf{b}^{(k)}\right), \quad k = 1, \dots, K, \quad (13)$$

where $\phi(\cdot)$ is a nonlinear activation function. The branch outputs are concatenated along the channel dimension:

$$\mathbf{F} = \text{Concat}\left(\mathbf{F}^{(1)}, \dots, \mathbf{F}^{(K)}\right). \quad (14)$$

To suppress channel expansion and enable cross-scale interaction, we apply a 1×1 convolution (channel-wise linear projection) to compress channels back to D :

$$\mathbf{Z}_t = \text{Conv}_{1 \times 1}(\mathbf{F}) \in \mathbb{R}^{B \times D \times L}. \quad (15)$$

This operation mixes channels within the same cycle without mixing information across cycles, thereby preserving temporal degradation structures while re-organizing multi-scale features.

2.2.3. Multi-Head Spectral Interaction and Band-Wise Refinement (MHSI)

Battery degradation signals also carry interpretable spectral structures: low-frequency components often reflect long-term trends, mid-frequency components reflect regime shifts, and high-frequency components capture short-term disturbances and noise. Standard self-attention models global dependencies but incurs $\mathcal{O}(L^2)$ complexity, which is costly for long windows. Moreover, homogeneous modeling may lead to band mixing and weaken hierarchical structure representation.

To address these issues, we propose the MHSI module (see Figure 5), which projects the encoded time-domain features into the frequency domain under a multi-head scheme, enables spectral interaction with a global receptive field via FFT, and performs adaptive band-wise refinement with linear-logarithmic complexity.

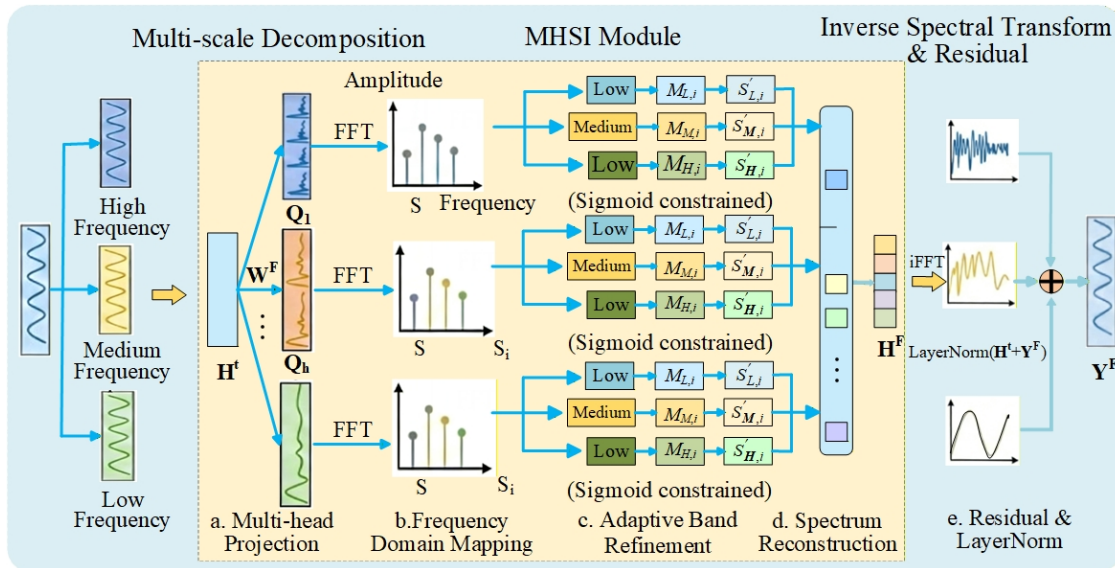


Figure 5. Overall architecture of the MHSI.

The time-domain output $\mathbf{Z}_t \in \mathbb{R}^{B \times D \times L}$ is first permuted to $\tilde{\mathbf{Z}}_t \in \mathbb{R}^{B \times L \times D}$ and linearly projected:

$$\mathbf{U} = \tilde{\mathbf{Z}}_t \mathbf{W}_p \in \mathbb{R}^{B \times L \times D}. \quad (16)$$

We then split \mathbf{U} into H heads along the feature dimension:

$$\mathbf{U} = \{\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(H)}\}, \quad \mathbf{U}^{(h)} \in \mathbb{R}^{B \times L \times d_h}, \quad d_h = D/H. \quad (17)$$

For each head, we perform a 1D FFT *only along the temporal dimension* (length L) for each channel:

$$\mathbf{s}^{(h)} = |\mathcal{F}_L(\mathbf{U}^{(h)})|, \quad (18)$$

where $\mathcal{F}_L(\cdot)$ denotes the 1D FFT along the length- L axis and $|\cdot|$ denotes magnitude spectrum (amplitude), which is used for subsequent band-wise refinement.

Using two frequency thresholds ω_1 and ω_2 (normalized w.r.t. the Nyquist frequency), the spectrum is partitioned into three bands:

$$\mathbf{s}^{(h)} = \{\mathbf{s}_L^{(h)}, \mathbf{s}_M^{(h)}, \mathbf{s}_H^{(h)}\}. \quad (19)$$

For each band $b \in \{L, M, H\}$, we introduce a learnable mask $\mathbf{M}_b^{(h)}$ and apply a sigmoid gate to enable soft selection:

$$\tilde{\mathbf{s}}_b^{(h)} = \mathbf{s}_b^{(h)} \odot \sigma(\mathbf{M}_b^{(h)}), \quad b \in \{L, M, H\}, \quad (20)$$

where \odot denotes the Hadamard product. This mechanism allows the model to adaptively adjust band contributions across batteries and degradation stages (e.g., emphasizing trend-related low-frequency components during accelerated aging while suppressing noise-related high-frequency components).

We then concatenate the refined bands to recover the full spectrum per head and fuse all heads:

$$\tilde{\mathbf{s}}^{(h)} = \text{Concat}(\tilde{\mathbf{s}}_L^{(h)}, \tilde{\mathbf{s}}_M^{(h)}, \tilde{\mathbf{s}}_H^{(h)}), \quad (21)$$

$$\tilde{\mathbf{s}} = \text{Concat}(\tilde{\mathbf{s}}^{(1)}, \dots, \tilde{\mathbf{s}}^{(H)}). \quad (22)$$

We map back to the time domain via inverse FFT:

$$\mathbf{Z}_f = \mathcal{F}_L^{-1}(\tilde{\mathbf{S}}), \quad (23)$$

and apply a residual connection with layer normalization:

$$\mathbf{Z}_f = \text{LN}(\mathbf{Z}_f + \mathbf{Z}_t). \quad (24)$$

2.2.4. Cross-Gated Time–Frequency Fusion and RUL Regression (CGF)

After MSTC and MHSL, we obtain two complementary representations: \mathbf{Z}_t emphasizes multi-scale time-domain patterns and local degradation trajectories, whereas \mathbf{Z}_f highlights global spectral structures and band-wise hierarchy information. Due to the non-stationary nature of battery degradation, their relative importance varies across stages and operating conditions; thus, naive summation or concatenation is insufficient for adaptive balancing. We design the CGF module to generate gating signals from one domain’s global summary to modulate the other domain, enabling dynamic complementary enhancement.

We first apply global average pooling (GAP) over the temporal dimension to obtain channel-wise summaries:

$$\mathbf{g}_f = \text{GAP}(\mathbf{Z}_f) \in \mathbb{R}^{B \times D}, \quad \mathbf{g}_t = \text{GAP}(\mathbf{Z}_t) \in \mathbb{R}^{B \times D}. \quad (25)$$

The gating coefficients are generated via sigmoid-activated linear projections:

$$\mathbf{m}_t = \sigma(\mathbf{g}_f \mathbf{W}_f + \mathbf{b}_f) \in \mathbb{R}^{B \times D}, \quad \mathbf{m}_f = \sigma(\mathbf{g}_t \mathbf{W}_t + \mathbf{b}_t) \in \mathbb{R}^{B \times D}. \quad (26)$$

With broadcasting along the temporal axis, the fused representation is

$$\mathbf{Z} = \mathbf{m}_t \odot \mathbf{Z}_t + \mathbf{m}_f \odot \mathbf{Z}_f \in \mathbb{R}^{B \times D \times L}. \quad (27)$$

Larger gate values indicate higher contribution from the corresponding domain, enabling the model to adaptively select more reliable information across batteries and degradation stages, which improves generalization and enhances the representation of critical phases near EOL.

Finally, we feed \mathbf{Z} into a regression head to output the RUL prediction. We first aggregate along the temporal axis (e.g., last-step selection, average pooling, or attention pooling):

$$\mathbf{h} = \text{Pool}(\mathbf{Z}) \in \mathbb{R}^{B \times D}, \quad (28)$$

then apply an MLP regressor:

$$\hat{\mathbf{y}} = \text{MLP}(\mathbf{h}) \in \mathbb{R}^{B \times 1} \quad (\text{or } \mathbb{R}^{B \times H} \text{ for multi-horizon}). \quad (29)$$

Here, $\hat{\mathbf{y}}$ denotes the predicted RUL (in cycles), which is consistent with the supervised labels defined in Eq. (9).

2.3. TF-RULNet (Time-Frequency RUL Network) Architecture

To comprehensively characterize the multi-scale temporal degradation patterns and the interpretable spectral structures of lithium-ion batteries during cycling aging, we propose **TF-RULNet** (Time-Frequency RUL Network). TF-RULNet takes a sequence of cycle-level features as input and learns a mapping from a historical health-indicator window to future RUL through a cascade of key components, including *multi-scale temporal encoding*, *spectral interaction enhancement*, *cross-gated time–frequency fusion*, and a *multi-step forecasting head*. The overall architecture is illustrated in Figure 3. The workflow is summarized as follows:

1. **Temporal encoding (MSTC):** Given the input tensor \mathbf{X} , a multi-scale temporal convolutional encoder is employed to extract degradation cues at different time scales, producing the temporal representation \mathbf{Z}_t .
2. **Spectral enhancement (MHSI):** The temporal features are mapped into the frequency domain, where multi-head spectral interaction and adaptive band-wise refinement are performed to obtain the frequency-enhanced representation \mathbf{Z}_f .
3. **Dynamic fusion (CGF):** A cross-gated fusion module adaptively re-weights the contributions of temporal and spectral branches, yielding the fused feature representation \mathbf{Z} .
4. **Multi-step regression (Forecast Head):** The fused representation \mathbf{Z} is aggregated and fed into a forecasting head to output the multi-step RUL prediction $\hat{\mathbf{y}} \in \mathbb{R}^{B \times P}$, where B denotes the batch size and P is the prediction horizon.

3. Experimental Settings and Model Training Procedure

3.1. Hardware Configuration

All experiments were conducted on a well-equipped workstation. The hardware platform consists of an Intel Core i9-10900K CPU and 32 GB of DDR4 RAM. To accelerate large-scale matrix computations and model training, an NVIDIA RTX 4060 GPU with 8 GB of dedicated VRAM was used. All deep learning models were implemented in PyTorch (v2.4.1), with GPU acceleration enabled via CUDA v12.4.

3.2. Model Hyperparameter Settings

In terms of model architecture, the proposed TF-RULNet consists of a multi-scale temporal convolution encoder (MSTC), a multi-head spectral interaction module (MHSI), and a cross-gated fusion module (CGF). The model takes a sliding-window sequence of length L as input and performs P -step-ahead forecasting. During training, the Adam optimizer is employed, and both learning-rate scheduling and early stopping are triggered when the validation loss no longer improves, which helps stabilize convergence while mitigating overfitting.

Specifically, the number of training epochs is set to `num_epochs = 100` with a batch size of 64. GELU is used as the activation function. The patience parameter of the learning-rate scheduler is set to `scheduler_patience = 0.25`, meaning that the learning rate will be reduced if the validation loss does not improve within the corresponding patience window. The early-stopping ratio is set to `early_patience = 0.2`. The remaining structural hyperparameters are summarized in Table 2, and a grid-search strategy is adopted to select the optimal configuration.

Table 2. Hyperparameter grid and the selected setting of the proposed TF-RULNet.

Hyperparameter	Search range (grid)	Selected
Model dimension (d_{model})	{32, 64, 128}	64
Number of heads (n_{heads})	{1, 8}	4
Feed-forward dimension (d_{ff})	{1, 128}	32
Dropout rate	{0.0, 0.5}	0.1
Initial learning rate (η)	$\{1 \times 10^{-4}, 1 \times 10^{-3}, 1 \times 10^{-2}\}$	1×10^{-3}

3.3. Model Training

For the i -th battery cell, a cycle-level feature vector extracted at the n -th cycle is denoted as $\mathbf{x}_n^{(i)} \in \mathbb{R}^D$. In this study, the core inputs are composed of discharge-stage statistical features, as defined in Eq. (5). Given a sliding-window length of L , the most recent L cycles are stacked to form one input sample, as shown in Eq. (6), and TF-RULNet performs multi-step prediction with horizon P .

To validate the effectiveness of the proposed model, we design two data partitioning and evaluation protocols. **(1) Within-dataset small-sample setting:** for a single cell, its lifetime sequence is split chronologically, where the first 30% cycles are used for training and the remaining 70% cycles are used

for validation/testing. This setting mimics practical scenarios where only early-life observations are available and the model must extrapolate the subsequent degradation trajectory. **(2) Cross-dataset generalization setting:** we select one cell (B007) from the NASA dataset as the source-domain training set, and directly evaluate on the University of Maryland dataset as the target domain. The source and target cells are strictly non-overlapping, enabling an assessment of TF-RULNet under distribution shifts caused by different operating conditions and/or protocols. All splits are performed at the *cell level* to prevent any sample overlap between domains and thus avoid information leakage.

The training objective is multi-step RUL regression using the mean squared error (MSE). Let $\mathbf{y} \in \mathbb{R}^{B \times P}$ be the ground-truth targets and $\hat{\mathbf{y}} \in \mathbb{R}^{B \times P}$ be the model outputs. The loss function is defined as

$$\mathcal{L}_{\text{MSE}} = \frac{1}{BP} \sum_{b=1}^B \sum_{p=1}^P (y_{b,p} - \hat{y}_{b,p})^2. \quad (30)$$

where B denotes the batch size, P is the multi-step forecasting horizon, $y_{b,p}$ and $\hat{y}_{b,p}$ represent the ground-truth and predicted RUL at the p -th step for the b -th sample, respectively. The term BP is the total number of scalar predictions in a mini-batch, thus \mathcal{L}_{MSE} computes the averaged squared error over all samples and all prediction steps.

3.4. Evaluation Metrics

To comprehensively assess the accuracy and stability of multi-step RUL forecasting, this study adopts three standard metrics, i.e., the root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2), which evaluate prediction performance from complementary perspectives.

Assume the test set contains N sliding-window samples and the forecasting horizon is P . Let $y_{i,p}$ and $\hat{y}_{i,p}$ denote the ground-truth and predicted RUL values of the i -th sample at the p -th forecasting step, respectively. For multi-step outputs, we aggregate the errors over all samples and all horizons to obtain the overall RMSE, MAE, and R^2 as follows:

$$\text{RMSE} = \sqrt{\frac{1}{NP} \sum_{i=1}^N \sum_{p=1}^P (y_{i,p} - \hat{y}_{i,p})^2}, \quad (31)$$

$$\text{MAE} = \frac{1}{NP} \sum_{i=1}^N \sum_{p=1}^P |y_{i,p} - \hat{y}_{i,p}|, \quad (32)$$

$$\bar{y} = \frac{1}{NP} \sum_{i=1}^N \sum_{p=1}^P y_{i,p}, \quad (33)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N \sum_{p=1}^P (y_{i,p} - \hat{y}_{i,p})^2}{\sum_{i=1}^N \sum_{p=1}^P (y_{i,p} - \bar{y})^2}. \quad (34)$$

Where, RMSE is more sensitive to large errors and thus reflects robustness under late-life regions or abrupt changes, while MAE provides an intuitive measure of the average absolute deviation. The R^2 score quantifies the goodness of fit by measuring how much variance in the ground-truth RUL can be explained by the predictions. By uniformly aggregating errors across all forecasting steps, the above metrics provide a holistic evaluation of multi-step performance, faithfully reflecting the practical scenario of “extrapolating future lifetime based on a historical window”.

4. Experimental Analysis

4.1. Experiment I: Comparison Under the Small-Sample Setting

4.1.1. Experiment Comparison on B005 and B007

To evaluate the relative advantage of the proposed TF-RULNet, we conduct comparative experiments on two NASA cells (B005 and B007) against a diverse set of baselines, including CNN, CNN-LSTM, Transformer, XGBoost, Random Forest (RF), and PatchTST. All models are trained and tested under the same data split strategy, identical cycle-level input features, and the same evaluation protocol to ensure a fair comparison. For deep learning baselines, we align the training configuration (e.g., optimizer, batch size, learning rate schedule, and early-stopping strategy) with TF-RULNet as closely as possible. For traditional machine learning methods (XGBoost and RF), the same cycle-level features are used as inputs, and the same supervised RUL labels are adopted for regression learning. The quantitative results are summarized in Table 3.

Table 3. Small-sample comparison results on B005 and B007.

Model	B005			B007		
	RMSE	MAE	R^2	RMSE	MAE	R^2
CNN	0.0273	0.0226	0.9129	0.0124	0.0092	0.9746
CNN-LSTM	0.0145	0.0115	0.9753	0.0115	0.0088	0.9781
Transformer	0.0131	0.0096	0.9798	0.0120	0.0101	0.9759
XGBoost	0.0249	0.0212	0.9273	0.0131	0.0107	0.9713
RF	0.0284	0.0245	0.9052	0.0173	0.0128	0.9504
PatchTST	0.0128	0.0102	0.9808	0.0111	0.0079	0.9794
TF-RULNet	0.0096	0.0074	0.9892	0.0084	0.0069	0.9883

As shown in Figure 6 and Table 3, the overall results under the small-sample setting indicate that **TF-RULNet achieves the best comprehensive performance on both batteries**. Using the strongest baseline, PatchTST, as a reference, TF-RULNet further reduces the RMSE on B005 by approximately 25.0% and the MAE by about 27.5%, while yielding an absolute improvement of about 0.0084 in R^2 . On B007, TF-RULNet continues to reduce the RMSE by around 24.3% and the MAE by about 12.7%, with an absolute R^2 gain of roughly 0.0089. These results demonstrate that the proposed model not only converges more stably with smaller errors, but also better fits the overall lifetime trajectory, exhibiting a stronger ability to *extrapolate the subsequent degradation evolution from limited early-cycle observations*.

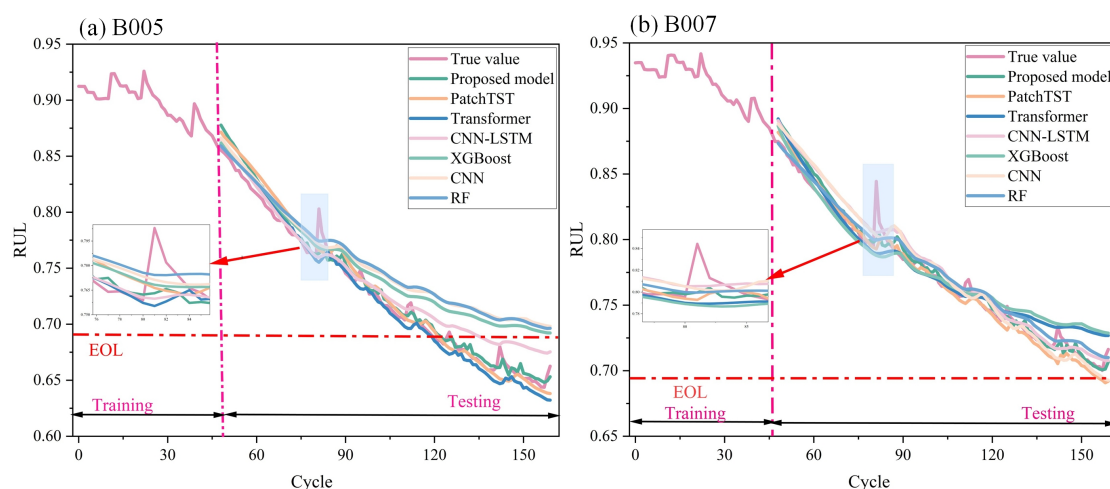


Figure 6. Performance comparison of TF-RULNet against baseline models under the small-sample setting on NASA B005 and B007.

Further comparisons with representative deep baselines show that TF-RULNet reduces RMSE by approximately 26.7%/30.0% and MAE by about 22.9%/31.7% on B005/B007 relative to the standard Transformer; compared with CNN-LSTM, RMSE decreases by roughly 33.8%/27.0% and MAE by about 35.7%/21.6%. These improvements suggest that models relying solely on time-domain attention or local convolutions are prone to overfitting local fluctuations and are less capable of capturing the global degradation rhythm in small-sample scenarios. In contrast, TF-RULNet leverages MSTC to provide multi-scale degradation cues, employs MHSI to enhance informative spectral structures in the frequency domain, and further uses CGF to dynamically fuse time–frequency information. Consequently, it mitigates error accumulation in multi-step forecasting and improves robustness across different batteries.

4.1.2. Ablation Experiment Under the Small-Sample Setting

To validate the effectiveness of each module in TF-RULNet, we conduct ablation experiments under the same training configuration as the comparative study. The ablations focus on three core components: the multi-scale time-domain convolution encoder (MSTC), the multi-head spectral interaction module (MHSI), and the cross-gated fusion module (CGF). Specifically, we remove the frequency-domain branch (w/o MHSI), remove the multi-scale time-domain branch (w/o MSTC), and replace the cross-gated fusion with a static fusion scheme (w/o CGF, implemented as Concat+Linear) to examine the necessity of *dynamic time–frequency trade-off*. The results are reported in Table 4.

As shown in Figure 7 and Table 4, the ablation results demonstrate that all three components contribute clearly to performance improvements, while their impact varies with dataset characteristics. Overall, removing any module leads to higher errors and lower goodness-of-fit. Among all ablations, removing MSTC (w/o MSTC) has the most pronounced effect on B005: compared with the full model, RMSE and MAE increase by approximately 31.3% and 40.5%, respectively. On B007, the degradation caused by w/o MSTC is relatively mild, with RMSE and MAE increasing by about 10.7% and 7.25%, indicating that the multi-scale time-domain convolution branch is particularly critical for capturing global–local degradation patterns.

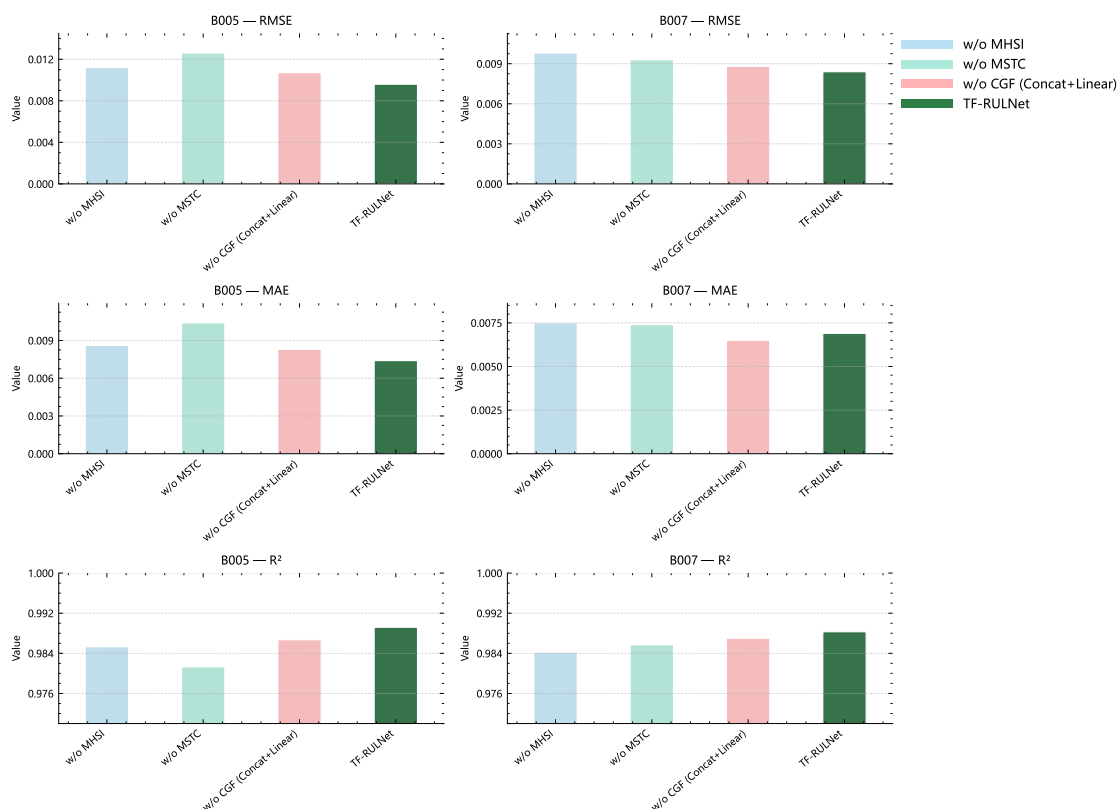


Figure 7. Ablation results of TF-RULNet on NASA B005 and B007 under the small-sample setting.

Table 4. Ablation results of TF-RULNet on B005 and B007 under the small-sample setting.

Model	B005			B007		
	RMSE	MAE	R^2	RMSE	MAE	R^2
w/o MHSI	0.0112	0.0086	0.9853	0.0098	0.0075	0.9842
w/o MSTC	0.0126	0.0104	0.9813	0.0093	0.0074	0.9857
w/o CGF (Concat+Linear)	0.0107	0.0083	0.9867	0.0088	0.0065	0.9870
TF-RULNet	0.0096	0.0074	0.9892	0.0084	0.0069	0.9883

Further comparisons show that the frequency-domain module MHSI provides consistent gains across datasets. After removing MHSI (w/o MHSI), RMSE increases by about 16.7% on both B005 and B007, while MAE increases by approximately 16.2% (B005) and 8.7% (B007). These results suggest that MHSI effectively supplements time-domain modeling by introducing spectral interactions and adaptive band refinement, alleviating issues such as frequency-band mixing and limited local receptive fields. This frequency-domain enhancement can also help the model better characterize localized recovery/regeneration-related variations.

Regarding the fusion strategy, replacing cross-gated fusion with static fusion (w/o CGF, Concat+Linear) results in RMSE/MAE increases of about 11.5%/12.2% on B005, and an RMSE increase of about 4.8% on B007. This indicates that the benefit of CGF is not merely an “accidental” improvement on a single metric; rather, it enables adaptive weighting between time- and frequency-domain information. Consequently, the full TF-RULNet achieves the most balanced and stable performance on both datasets.

4.2. Cross-Dataset Generalization Experiment (NASA → Maryland)

To examine the robustness of the proposed model under distribution shifts caused by cross-device and cross-operating-condition variations, we further construct a cross-dataset setting from NASA to Maryland. Specifically, the model is trained on NASA B005 to learn a transferable degradation representation, and then directly evaluated on the Maryland dataset (CS-35 and CS-37). The results are reported in Table 5.

Table 5. Cross-dataset generalization results (NASA B005 → Maryland CS-35/CS-37).

Model	CS-35			CS-37		
	RMSE	MAE	R^2	RMSE	MAE	R^2
CNN	0.0541	0.0511	0.9179	0.0360	0.0238	0.9635
CNN-LSTM	0.0485	0.0283	0.9339	0.0399	0.0253	0.9552
Transformer	0.0492	0.0275	0.9321	0.0339	0.0199	0.9676
XGBoost	0.0814	0.0543	0.8139	0.0535	0.0310	0.9194
RF	0.0667	0.0416	0.8751	0.0775	0.0543	0.8309
PatchTST	0.0293	0.0248	0.9758	0.0301	0.0196	0.9744
TF-RULNet	0.0181	0.0121	0.9908	0.0165	0.0097	0.9923

As illustrated in Figure 8 and Table 5, TF-RULNet remains consistently superior in the NASA → Maryland transfer scenario, demonstrating that the learned representation is more transferable across devices and operating protocols. Using PatchTST—the strongest deep baseline in this cross-dataset setting—as a reference, TF-RULNet further reduces RMSE and MAE on CS-35 by **38.23%** and **51.21%**, respectively; on CS-37, RMSE and MAE are reduced by **45.18%** and **50.51%**, respectively. In addition, when measured by the *unexplained error ratio* ($1 - R^2$), TF-RULNet decreases ($1 - R^2$) by approximately **61.98%** on CS-35 and **69.92%** on CS-37 compared with PatchTST, indicating substantially improved cross-domain generalization for battery lifetime prognostics.

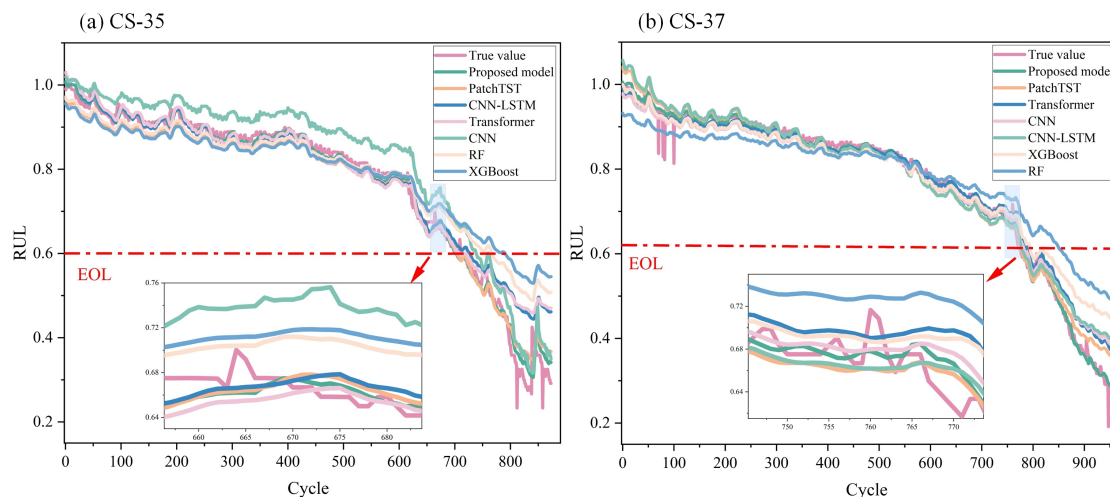


Figure 8. Performance comparison of TF-RULNet against baseline models in the cross-dataset generalization setting (NASA B005 → Maryland CS-35/CS-37).

The advantage becomes even more pronounced when compared with conventional deep models and classical machine learning regressors. Compared with the standard Transformer, TF-RULNet achieves RMSE reductions of **63.21%** (CS-35) and **51.33%** (CS-37), with MAE reductions of **56.00%** and **51.26%**, respectively. Relative to CNN, the RMSE decreases by **66.54%** and **54.17%**, while the MAE decreases by **76.32%** and **59.24%** on CS-35 and CS-37, respectively. For XGBoost and RF, which are typically more sensitive to feature distribution shifts, TF-RULNet yields RMSE drops of **69.16%**–**77.76%** (vs. XGBoost) and **72.86%**–**78.71%** (vs. RF), accompanied by MAE reductions of **68.71%**–**82.14%**. Collectively, these gains support the conclusion that, under cross-dataset transfer, TF-RULNet—via *multi-scale time-domain encoding*, *spectral interaction in the frequency domain*, and *cross-gated dynamic fusion*—more effectively suppresses error amplification induced by domain shifts, thereby maintaining the best overall predictive performance in a setting closer to real-world deployment.

5. Conclusions and Future Work

5.1. Conclusions

This paper addresses the multi-step remaining useful life (RUL) prediction problem for lithium-ion batteries and proposes a time–frequency fusion prognostic model, namely TF-RULNet (*Time-Frequency RUL Network*). TF-RULNet consists of a multi-scale time-domain convolution encoder (MSTC), a multi-head spectral interaction module (MHSI), and a cross-gated fusion module (CGF). The core idea is to (i) capture degradation patterns at different temporal scales in the time domain, (ii) leverage FFT with linear-logarithmic complexity to model local spectral structures and band-wise patterns in the frequency domain, and (iii) dynamically balance and complement the time–frequency information via cross-gating, thereby enhancing representation capability and generalization for complex non-stationary degradation processes.

Extensive experiments under both the small-sample setting and the cross-dataset generalization setting demonstrate that TF-RULNet consistently achieves superior performance in terms of RMSE/MAE compared with a diverse set of baselines, including CNN, CNN-LSTM, standard Transformer, XGBoost, Random Forest, and PatchTST. Moreover, ablation studies further verify the effectiveness of the three core modules, indicating that multi-scale time-domain modeling, spectral interaction in the frequency domain, and dynamic gated fusion are all indispensable and jointly contribute to the overall performance gains of TF-RULNet.

5.2. Future Work

Although TF-RULNet has shown strong accuracy and robustness in both in-domain and cross-domain evaluations, several directions remain promising for further improving battery life prognostics in more complex, engineering-oriented scenarios:

1. **Uncertainty quantification for risk-aware decision making.** In practical BMS applications, *reliable* predictions are often more valuable than a single-point optimum. Future work will extend TF-RULNet to probabilistic forecasting, providing confidence intervals and risk indicators for multi-step RUL, which can facilitate safety margin assessment and maintenance planning.
2. **More systematic cross-domain adaptation and online updating.** While cross-dataset experiments validate the transferability of TF-RULNet, real-world operating conditions (e.g., protocols, temperatures, loads) may drift continuously. Future research will investigate parameter-efficient adaptation and online fine-tuning strategies to enable rapid updates with limited new data while mitigating catastrophic forgetting.
3. **From macro cycle-level features to fine-grained segment-level mechanistic feature fusion.** The current study mainly uses macro cycle-level statistical features. Future extensions will incorporate finer-grained information, such as charging/discharging curve segments and IC/DV peak characteristics, into a unified representation framework, aiming to better capture stage-wise behaviors (e.g., capacity regeneration) and non-linear degradation acceleration near the end of life.

In summary, TF-RULNet provides a time–frequency fusion paradigm that is both accurate and robust for multi-step battery RUL prediction. Future work will focus on uncertainty-aware prognostics, continual domain adaptation, and fine-grained feature fusion to further enhance the practical applicability and reliability of the model under complex operating conditions.

References

1. Htet, M.; et al. A Review of Bayesian-Filtering-Based Techniques in RUL Prediction for Lithium-Ion Batteries. *Journal of Energy Storage* **2025**, *111*, 115371.
2. Paccha-Herrera, E.; et al. A Particle Filter-Based Approach for Real-Time Temperature Estimation in a Lithium-Ion Battery Module during the Cooling-down Process. *Journal of Energy Storage* **2024**, *94*, 112413. <https://doi.org/10.1016/j.est.2024.112413>.
3. Liu, K.; et al. A Data-Driven Approach with Uncertainty Quantification for Predicting Future Capacities and Remaining Useful Life of Lithium-Ion Battery. *IEEE Transactions on Industrial Electronics* **2020**. <https://doi.org/10.1109/TIE.2020.2973876>.
4. Zhang, S.; Liu, Z.; Su, H. A Bayesian Mixture Neural Network for Remaining Useful Life Prediction of Lithium-Ion Batteries. *IEEE Transactions on Transportation Electrification* **2022**, *8*, 4708–4721.
5. Li, X.; et al. Li-Ion Battery State of Health Prediction through Metaheuristic Algorithms and Genetic Programming. *Energy Reports* **2024**, *12*, 368–380. <https://doi.org/10.1016/j.egyr.2024.06.038>.
6. Li, Y.; et al. SOH Evaluation and RUL Estimation of Lithium-Ion Batteries Based on MC-CNN-TimesNet Model. *Reliability Engineering & System Safety* **2025**, *261*, 111125. <https://doi.org/10.1016/j.res.2025.111125>.
7. Shi, J.; et al. Battery Health Management Using Physics-Informed Machine Learning: Online Degradation Modeling and Remaining Useful Life Prediction. *Mechanical Systems and Signal Processing* **2022**, *179*, 109347. <https://doi.org/10.1016/j.ymsp.2022.109347>.
8. Wang, Y.; Hei, C.; Liu, H.; et al. Prognostics of Remaining Useful Life for Lithium-Ion Batteries Based on Hybrid Approach of Linear Pattern Extraction and Nonlinear Relationship Mining. *IEEE Transactions on Power Electronics* **2022**, *38*, 1054–1063.
9. Demirci, O.; Taskin, S.; Schaltz, E.; et al. Review of Battery State Estimation Methods for Electric Vehicles-Part II: SOH Estimation. *Journal of Energy Storage* **2024**, *96*, 112703.
10. Pournazari, J.; et al. LiRUL: A Hybrid and Lightweight Learning Framework for RUL Prediction on Edge Devices. *SSRN Electronic Journal*, 2025. <https://doi.org/10.2139/ssrn.5611930>.
11. Peng, S.; et al. State of Charge Estimation for LiFePO4 Batteries Joint by PID Observer and Improved EKF in Various OCV Ranges. *Applied Energy* **2024**, *377*, 124435. <https://doi.org/10.1016/j.apenergy.2024.124435>.

12. Zhang, Y.; et al. Application of Physics-Informed Machine Learning in Performance Degradation and RUL Prediction of Hydraulic Piston Pumps. *Reliability Engineering & System Safety* **2025**, *261*, 111108. <https://doi.org/10.1016/j.res.2025.111108>.
13. Zhou, J.; et al. Centralized Strategy Learning with Multi-Feature Scale Health Factors for Lithium-Ion Battery RUL Prediction. *Journal of Energy Storage* **2025**, *132*, 117875.
14. Chen, J.; et al. A Convolutional Neural Network for Estimation of Lithium-Ion Battery State-of-Health during Constant Current Operation. In Proceedings of the IEEE Transportation Electrification Conference, 2023. <https://doi.org/10.1109/ITEC55900.2023.10186914>.
15. Jin, Z.; et al. A Data-Driven Framework for Lithium-Ion Battery RUL Using LSTM and XGBoost with Feature Selection via Binary Firefly Algorithm. *Energy* **2024**, *314*, 134229. <https://doi.org/10.1016/j.energy.2024.134229>.
16. Liu, L.; et al. A Novel Interpretable Deep Autoencoder for Health Indicators Construction and Similarity-Based RUL Prediction. *Measurement* **2026**, p. 120571.
17. Pournazari, J.; et al. DFL-RUL: Decentralised Federated Learning for Battery Remaining Useful Life Estimation on Heterogeneous Edge-To-Cloud. *Energy and AI* **2026**, *24*, 100689. <https://doi.org/10.1016/j.egyai.2026.100689>.
18. Guo, W.; Tan, A.H.; Ong, D.S. Optimized Prediction of Remaining Useful Life of Lithium-Ion Batteries: A Voltage-Current Behavior Analysis for Enhanced Health Monitoring. *Journal of Energy Storage* **2025**, *134*, 118138.
19. Pan, W.; et al. A Health Indicator Extraction and Optimization for Capacity Estimation of Li-Ion Battery Using Incremental Capacity Curves. *Journal of Energy Storage* **2021**, *42*, 103072.
20. Wang, Z.; et al. Adaptive Self-Attention LSTM for RUL Prediction of Lithium-Ion Batteries. *Information Sciences* **2023**. <https://doi.org/10.1016/j.ins.2023.01.100>.
21. Rengarajan, V.; Anuradha, T. A Hybrid Neural Architecture for Enhanced Lithium-Ion Battery SOH Estimation and RUL Prediction. *Journal of Energy Storage* **2025**, *140*, 119019.
22. Aizpurua, J.I.; et al. Probabilistic Forecasting Informed Failure Prognostics Framework for Improved RUL Prediction under Uncertainty: A Transformer Case Study. *Reliability Engineering & System Safety* **2022**, *226*, 108676.
23. Yu, B.; et al. MDM: A Mamba-Conditioned Diffusion Model for Spacecraft Lithium-Ion Battery RUL Prediction. *IFAC-PapersOnLine* **2025**, *59*, 781–786. <https://doi.org/10.1016/j.ifacol.2025.11.247>.
24. Liu, L.; et al. PatchFormer: A Novel Patch-Based Transformer for Accurate Remaining Useful Life Prediction of Lithium-Ion Batteries. *Journal of Power Sources* **2025**, *631*, 236187. <https://doi.org/10.1016/j.jpowsour.2025.236187>.
25. Zhao, J.; et al. A Survey of Transformer Networks for Time Series Forecasting. *Computer Science Review* **2025**, *60*, 100883. <https://doi.org/10.1016/j.cosrev.2025.100883>.
26. Zhao, T.; et al. TFFormer: A Time-Frequency Domain Bidirectional Sequence-Level Attention Based Transformer for Interpretable Long-Term Sequence Forecasting. *Pattern Recognition* **2025**, *158*, 110994. <https://doi.org/10.1016/j.patcog.2024.110994>.
27. Shao, X.; et al. EnergyFormer: Dual-Supervised Adaptive Multi-Scale Transformer for Multistep Energy Forecasting. *Expert Systems with Applications* **2025**, *305*, 130750. <https://doi.org/10.1016/j.eswa.2025.130750>.
28. Li, J.; et al. A Dual-Branch Multi-Scale Encoding and Fusion Model for Multivariate Time Series Forecasting. *Engineering Applications of Artificial Intelligence* **2026**, *166*, 113610. <https://doi.org/10.1016/j.engappai.2025.113610>.
29. Guo, Y.; et al. A Multi-Scale Spatiotemporal Spiking Neural Model for Power Load Forecasting Considering Extreme Weather Impact. *International Journal of Electrical Power & Energy Systems* **2026**, *175*, 111604. <https://doi.org/10.1016/j.ijepes.2026.111604>.
30. Li, H.; et al. Data-Driven Heterogeneous Ensemble Forecasting Framework Incorporating Multi-Scale Criteria. *Expert Systems with Applications* **2026**, *302*, 130537. <https://doi.org/10.1016/j.eswa.2025.130537>.
31. Qian, Z.; et al. A Transfer Condition-Focused Model for Battery Capacity Forecast. *Reliability Engineering & System Safety* **2026**, *267*, 111928. <https://doi.org/10.1016/j.res.2025.111928>.
32. Duan, Y.; et al. Cross-Domain Time-Frequency Mamba: A More Effective Model for Long-Term Time Series Forecasting. *Knowledge-Based Systems* **2026**, *115341*. <https://doi.org/10.1016/j.knosys.2026.115341>.
33. Hao, F.; et al. FTdasc: A Frequency-Time Domain Approach with Stationarity Correction for Multivariate Time Series Forecasting. *Expert Systems with Applications* **2026**, *308*, 131164. <https://doi.org/10.1016/j.eswa.2026.131164>.

34. Fan, H.; et al. EPformer: Unlocking Day-Ahead Electricity Price Forecasting Accuracy Using the Time–Frequency Domain Feature Learning Strategy Considering Renewable Energy. *Renewable Energy* **2026**, *261*, 125296. <https://doi.org/10.1016/j.renene.2026.125296>.
35. Wu, P.q.; et al. Long-Term Time Series Forecasting by Combining Local Periodic Change Features in Frequency Domain and Global Features in Time Domain. *Journal of the Franklin Institute* **2026**, *363*, 108304. <https://doi.org/10.1016/j.jfranklin.2025.108304>.
36. Aljohani, A.; Aljohani, S. A Hybrid GRU-MHA Model for Accurate Battery RUL Forecasting with Feature Selection. *Energy Reports* **2025**, *14*, 294–309. <https://doi.org/10.1016/j.egy.2025.05.059>.
37. Zhang, M.; et al. A Dynamic Association Multi-Attribute Fusion Graph Network for Multivariate Time Series Forecasting. *Information Processing & Management* **2026**, *63*, 104588. <https://doi.org/10.1016/j.ipm.2025.104588>.
38. Li, Z.; et al. DDformer: Transformer with Dynamic Variable Fusion and Dynamic Difference Attention for Multivariate Time Series Long-Term Forecasting. *Neurocomputing* **2026**, *672*, 132716. <https://doi.org/10.1016/j.neucom.2026.132716>.
39. Jiang, Z.; Tan, Q.; Che, J.; et al. Spatio-temporal Convolutional Attention Dynamic Fusion Network for Regional Collaborative Wind Power Forecasting. *Expert Systems with Applications* **2025**, 130291.
40. Yao, X.; Xu, H. A HybridGCN-GRU Approach for High-Accuracy Wind Speed Forecasting via Multi-Graph Feature Fusion and Dynamic Modeling. *Information Sciences* **2026**, *728*, 122805. <https://doi.org/10.1016/j.ins.2025.122805>.
41. Obisakin, I.; Ekeanyanwu, C.V. State of health estimation of lithium-ion batteries using support vector regression and long short-term memory. *Open Journal of Applied Sciences* **2022**, *12*, 1366–1382.
42. He, W.; Williard, N.; Osterman, M.; Pecht, M. Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method. *Journal of Power Sources* **2011**, *196*, 10314–10321.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.