Journal of *Marine Science and Engineering*

MDPI

Article

RUL Prediction for Lithium Battery Systems in Fuel Cell Ships Based on Adaptive Modal Enhancement Networks

Yifan Liu, Huabiao Jin, Xiangguo Yang *, Telu Tang, Jiaxin Luo, Lei Han, Junting Lang and Weixin Zhao

School of Naval Architecture, Ocean and Energy Power Engineering, Wuhan University of Technology, Wuhan 430070, China; 345426@whut.edu.cn (Y.L.); yx_jhb@163.com (H.J.); 331552@whut.edu.cn (T.T.); whut-ljx@whut.edu.cn (J.L.); 359814@whut.edu.cn (L.H.); 359819@whut.edu.cn (J.L.); wx.zhao@whut.edu.cn (W.Z.)

* Correspondence: yxglyr@whut.edu.cn

Abstract: With the widespread application of fuel cell technology in the fields of transportation and energy, Battery Management Systems (BMSs) have become one of the key technologies for ensuring system stability and extending battery lifespan. As an auxiliary power source in fuel cell systems, the prediction of the Remaining Useful Life (RUL) of lithium-ion batteries is crucial for enhancing the reliability and efficiency of fuel cell ships. However, due to the complex degradation mechanisms of lithium batteries and the actual noisy operating conditions, particularly capacity regeneration noise, accurate RUL prediction remains a challenge. To address this issue, this paper proposes a lithium battery RUL prediction method based on an Adaptive Modal Enhancement Network (RIME-VMD-SEInformer). By incorporating an improved Variational Mode Decomposition (VMD) technique, the RIME algorithm is used to optimize decomposition parameters for the adaptive extraction of key modes from the signal. The Squeeze-and-Excitation Networks (SEAttention) module is employed to enhance the accuracy of feature extraction, and the sparse attention mechanism of Informer is utilized to efficiently model long-term dependencies in time series. This results in a comprehensive prediction framework that spans signal decomposition, feature enhancement, and time-series modeling. The method is validated on several public datasets, and the results demonstrate that each component of the RIME-VMD-SEInformer framework is both necessary and justifiable, leading to improved performance. The model outperforms the state-of-the-art models, with a MAPE of only 0.00837 on the B0005 dataset, representing a 59.96% reduction compared to other algorithms, showcasing outstanding prediction performance.

Keywords: fuel cell; hybrid power system; lithium-ion battery; remaining useful life prediction; informer

check for updates

Academic Editor: Rosemary Norman

Received: 17 January 2025
Revised: 30 January 2025
Accepted: 4 February 2025
Published: 5 February 2025

Citation: Liu, Y.; Jin, H.; Yang, X.; Tang, T.; Luo, J.; Han, L.; Lang, J.; Zhao, W. RUL Prediction for Lithium Battery Systems in Fuel Cell Ships Based on Adaptive Modal Enhancement Networks. *J. Mar. Sci. Eng.* **2025**, *13*, 296. https://doi.org/10.3390/jmse13020296

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Ship emissions during the shipping process have a significant environmental impact. In response to China's energy transition goals [1], it is imperative to drive the energy system toward cleaner, greener, and low-carbon alternatives. Currently, fuel cell systems are considered one of the most promising maritime power technologies [2]. Among them, Proton Exchange Membrane Fuel Cells (PEMFC) have become the dominant technology for fuel cell-powered ships owing to their zero emissions, high power density, and quick start-up advantages. PEMFCs are particularly suitable for low-power vessels, such as small yachts and passenger ships. Compared to traditional diesel engines, fuel cells can

significantly reduce greenhouse gas emissions and pollutants, providing technical support for achieving green development in the shipping sector [3].

Hybrid power systems integrate different energy conversion technologies [4] and represent a promising solution for improving efficiency and reducing emissions, particularly in the maritime sector. Typically, these systems combine PEMFC with lithium batteries to provide a cleaner and more sustainable energy alternative [5]. The PEMFC system offers significant advantages in delivering low-carbon energy, while lithium batteries serve as additional energy storage, enhancing the overall system performance, particularly under fluctuating load conditions. Table 1 presents a comparison of the key parameters between diesel engine power systems and PEMFC + lithium battery hybrid power systems. Diesel engine systems have lower installation and operating costs, but exhibit higher emissions, particularly under low-load conditions where NO_x and particulate matter (PM) emissions are more pronounced. In contrast, the PEMFC + lithium battery hybrid power system has near-zero emissions [6]; however, due to the high costs associated with fuel cells and hydrogen storage, its installation and operating expenses are significantly higher.

Table 1. Comparison table of parameters between diesel drive system and PEMFC + lithium battery hybrid system.

Comparison Item	Diesel Engine Power System	PEMFC + Lithium Battery Power System
Installation Cost (USD/kW)	500–1500	2000–4000
Operating Cost (USD/h)	10–40	20–60
CO_2 Emissions (kg/kWh)	0.2–0.3	Almost 0
Fuel Consumption	0.2–0.3 L/kWh	0.5–1.0 kg/kWh
Power Output (kW)	50–30000	50–5000
Range (Nautical Miles)	1000–3000	50–500

Regarding fuel consumption, the PEMFC system has lower energy efficiency than diesel engines due to the lower energy density of hydrogen [7,8], leading to higher hydrogen consumption. Furthermore, in terms of power output, diesel engine power systems offer a broad power range, making them suitable for various vessel types, from small boats to large ocean-going ships. Conversely, the PEMFC + lithium battery hybrid system is currently primarily deployed in small- to medium-sized vessels. Due to the limited power and energy density of fuel cells and lithium batteries, large ships have yet to fully adopt this hybrid power system. However, with advancements in technology and improvements in hydrogen storage techniques, this system is expected to gradually expand to vessels with a higher power demand in the future. Figure 1 Topology Diagram of the Power System for Fuel Cell Ships illustrates the distribution structure of a certain hybrid power system. It consists of four Fuel Cell Systems (FCSs), four Boost Converters, two Lithium Battery Storage (LBS) units, two Buck/Boost Converters, propulsion loads, propulsion drivers, inverters, service loads, and other modules. The distribution structure of the HPS includes a main distribution board with DC 750V, an auxiliary distribution board with AC 380V, a daily use distribution board with AC 220V, and a control distribution board with DC 24V. Additionally, Electrical Interlocks are present in both the DC 750V main distribution board and the AC 380V auxiliary distribution board. This structure is integrated into a container-based distribution cabinet with approximate dimensions of 8 m in length, 3 m in height, and 3 m in width.

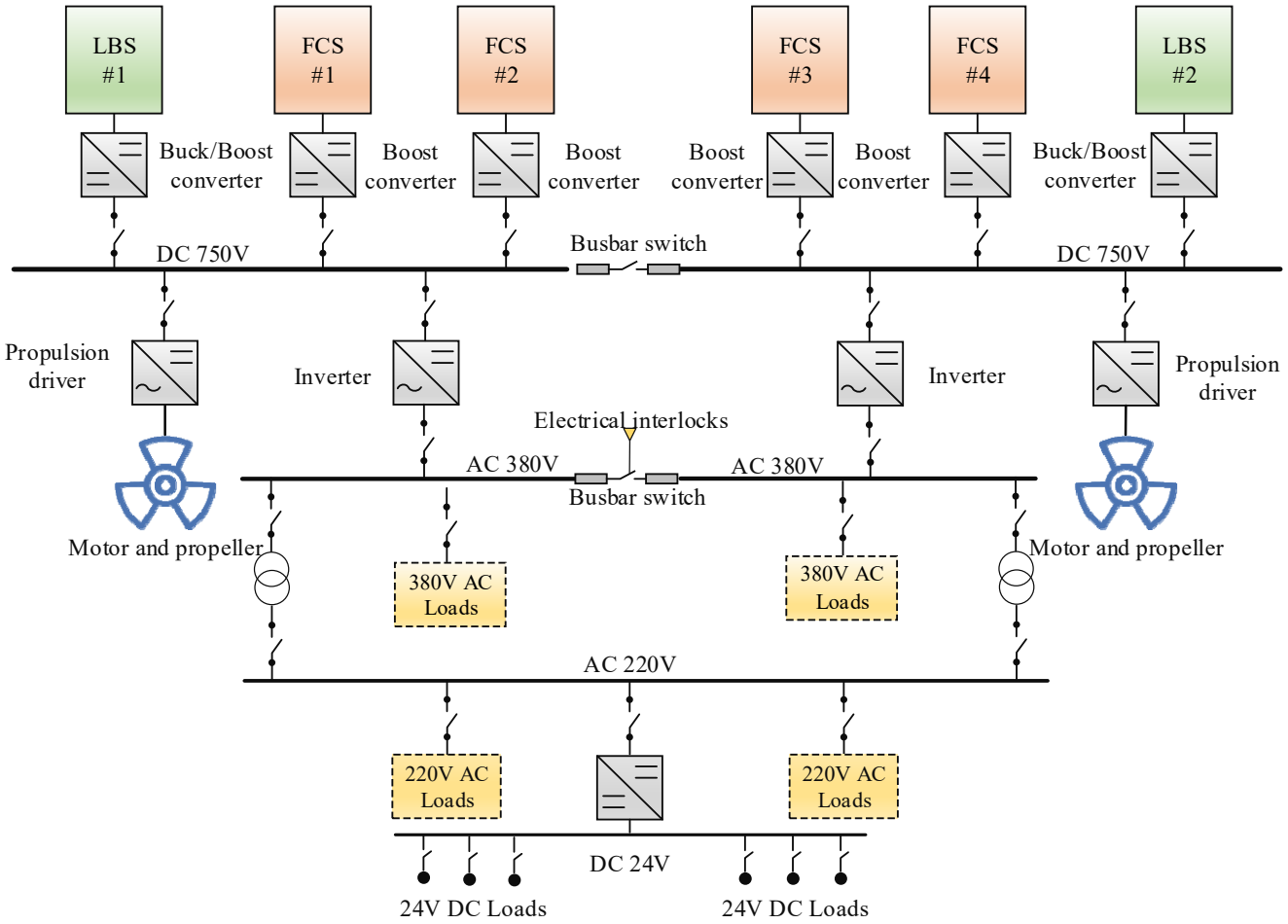


Figure 1. Topology diagram of the power system for fuel cell ships.

Lithium-ion batteries outperform other energy storage technologies in hybrid power systems due to their high energy density (150–250 Wh/kg), long cycle life (1000–5000 cycles), fast charging, and high efficiency (90–95%) [9,10]. Compared to lead-acid batteries, they last 3–5 times longer and charge much faster. While supercapacitors and flywheels offer rapid charging, they lack sufficient energy storage. Solid-state batteries and hydrogen fuel cells promise better performance but remain expensive and less commercially viable. With a good balance of cost, performance, and recyclability (60–80%), lithium-ion batteries are currently the most practical and widely used solution for electric vehicles, ships, and energy storage applications. However, lithium batteries inevitably experience aging and performance degradation over time, leading to capacity loss, reduced power output, and other issues. This not only reduces the energy efficiency of the ship but may also pose a threat to navigational safety. Therefore, accurately assessing the health status of lithium batteries and predicting their RUL has become critical for ensuring the stable operation of ship power systems [11]. RUL refers to the number of cycle periods required for the battery's maximum available capacity to degrade to a predefined failure threshold under specified charge and discharge conditions. In RUL prediction, extracting RUL information from smoothed capacity or State of Health (SOH) curves [12], which have undergone noise reduction, is more efficient than directly handling raw test data with noise. Wang et al. [13] proposed a prediction method based on decomposed terminal voltage data using the Wavelet Decomposition Technique (WDT), further expanding the application of WDT in RUL prediction. Mainstream methods for RUL prediction can be categorized into three types: model-based methods, data-driven methods, and hybrid approaches.

For model-based methods, the literature [14] established a three-parameter capacity degradation model based on the double-exponential formula, demonstrating superior prediction performance compared to the double-exponential formula alone. Additionally, the literature [15] incorporated the impact of the cutoff voltage on the capacity and internal resistance of LFP batteries, exploring the coupling effects of dual influencing factors on battery aging. The literature [16] proposed a capacity degradation model based on the formation mechanism of the solid electrolyte interface (SEI) film, considering factors such as the average State of Charge (SOC) and DOD. However, semi-empirical models with stress factors are no longer applicable when the battery operates under drastic changes in conditions, and most studies have focused on aging characteristics when the battery capacity is above 80%, with limited research on situations where the capacity is below 80%.

Data-driven methods directly mine the degradation information and evolution of the health state of lithium-ion batteries from historical data. Most studies have focused on the trend analysis of battery degradation curves. Xiong et al. [17] applied semi-supervised learning, directly inputting Electrochemical Impedance Spectroscopy (EIS) data into a Convolutional Neural Network (CNN) for feature extraction, estimating capacity without the need for capacity labels. Current research on CNNs focuses on hyperparameter optimization, combination with Long Short-Term Memory (LSTM) networks, the application of dilated CNNs, and transfer learning. Hyperparameter optimization improves model performance through methods such as Bayesian optimization. The CNN-LSTM model captures both spatial and temporal information, significantly improving RUL prediction accuracy [18]. Ren et al. [19] introduced autoencoders to enhance the CNN-LSTM model by increasing data dimensionality. Zhang et al. [20] combined CNN-LSTM with Convolutional Recurrent Generative Adversarial Networks (GANs) to generate more realistic time-series data, reducing prediction errors. Yang et al. [21] combined CNN with Bidirectional LSTM (Bi-LSTM) to enhance learning effectiveness and generalization capabilities. The dilated CNN, which increases the receptive field by expanding the convolution kernel, effectively improves the model's ability to capture long-term dependencies. Hong et al. [22] proposed a five-layer dilated CNN framework, significantly improving the RUL prediction performance. Zhou et al. [23] introduced Temporal Convolutional Networks (TCNs), which utilize causal and dilated convolution techniques to enhance the ability to capture local capacity regeneration phenomena, improving prediction accuracy.

Transfer learning provides an effective approach for addressing small dataset issues [24]. Due to the advantages and limitations of both model-based and data-driven methods, some researchers have combined multiple methods to estimate the SOC of lithium-ion batteries, thereby leveraging complementary strengths. Hybrid approaches include combinations of model-based and data-driven methods, as well as different data-driven methods. The capacity data of lithium-ion batteries in practical applications are often affected by factors such as instrument errors, measurement noise, uncertain load conditions, and capacity regeneration phenomena. Directly using these data for Remaining Useful Life prediction may lead to inaccurate results. To address these challenges, VMD, an adaptive and non-recursive signal processing method, can effectively decompose complex non-stationary time series and provide stable multi-scale frequency subseries, thus reducing noise interference. Building upon this foundation, integrating advanced optimization algorithms, attention mechanisms, and deep learning models can enhance prediction performance. For model-based methods, establishing an accurate physical/chemical model requires a large amount of experimental data and computational resources, making the calculation complex. Moreover, these models cannot adapt to multiple operating conditions, and their performance deteriorates when the battery degrades below 80%. Data-driven methods have poor time dependency handling and are also significantly affected by noise.

In contrast, the model proposed in this paper can extract both the global trends and local fluctuations of battery lifespan data, reducing the impact of noise, better capturing long-term time dependencies, and improving prediction stability. It is suitable for complex operating conditions and has a stronger noise reduction capability. The main contributions of this paper are summarized as follows:

1. This paper presents a novel hybrid method for RUL prediction that encompasses data processing and network training. The data processing step utilizes the RIME-optimized VMD algorithm to decompose real-world capacity data, distinguishing the overall trend from local fluctuations, thereby improving prediction accuracy.
2. To process the decomposed capacity data, SEAttention and RIME-VMD interact by learning channel-level attention weights, enabling the adjustment of input features based on the importance of different channels at each time step. This enhances the learning of intrinsic mode functions (IMFs) across time-series data, which is crucial for capturing varying signals at different moments.
3. The Informer model enhances the information flow and multi-scale modeling capability, allowing it to better capture the complex long-term dependencies within time-series data. When combined with the RIME-VMD-SE module, the input data can further improve both prediction accuracy and efficiency.

2. Basic Algorithm Theory

This section describes the methods involved in the hybrid prediction model proposed in this paper.

2.1. Theoretical Basis of RIME-VMD Method

2.1.1. VMD

The core idea of VMD is to iteratively optimize and decompose the original signal into several modes with distinct frequency and amplitude characteristics [25], which are mutually orthogonal in the frequency domain [26]. In BMSs, battery capacity degradation is usually non-stationary and heavily influenced by various factors such as temperature, load, and charging strategies, all of which have significant impacts on RUL evaluation. VMD offers a method for decomposing battery capacity signals into intrinsic mode functions (IMFs), which extract frequency components with varying resolutions and distinct properties, providing an effective way to analyze battery capacity changes. The optimization problem for VMD can be expressed as

$$\min_{\{u_k\},\{\omega_k\}} \sum_{k=1}^{K} \int_{-\infty}^{+\infty} \left| \frac{d}{dt} \left(f(t) - \sum_{k=1}^{K} u_k(t) \right) \right|^2 dt \tag{1}$$

In this equation, $u_k(t)$ represents the k-th intrinsic mode function (IMF), and ω_k is the center frequency of the mode. The optimization aims to minimize the difference between the signal and the sum of the extracted IMFs.

During the decomposition process, the IMFs are iteratively extracted to represent the signal with various frequency bands, with each IMF capturing different frequency components. The decomposition of battery capacity signals allows the capture of long-term degradation trends as well as short-term fluctuations. To solve the VMD optimization problem, VMD introduces Lagrange multipliers and a quadratic penalty term, which can be expressed as

$$L = \sum_{k=1}^{K} \left| \frac{d}{dt} (f(t) - u_k(t)) \right|^2 + \lambda_k \left(\left| \frac{d}{dt} u_k(t) \right|_2 - C \right) \tag{2}$$

Here, λ_k is the Lagrange multiplier, and C is a constant that controls the frequency constraint of each IMF. This formulation allows for a more flexible decomposition while adhering to the constraints of the frequency and signal components. By introducing these constraints, VMD can effectively capture multiple frequency modes in battery capacity degradation signals [27], especially capturing noise components and low-frequency trends that are crucial for RUL prediction.

ADMM is used to solve the VMD optimization problem, and it iteratively updates the IMF components and their corresponding center frequencies. The parameters, including the number of modes K and penalty factor α, are critical for ensuring accurate decomposition. If K is set too small, important features of the battery signal may be lost, while a larger K can lead to excessive computational complexity and redundancy in the decomposition. Similarly, the penalty factor α must be optimized; too large a value may lead to overly smooth modes (underfitting), and too small a value may result in overly flexible modes (overfitting). Therefore, adaptive optimization of K and α is essential for enhancing the efficiency and accuracy of VMD decomposition.

In summary, VMD offers a rigorous and effective signal processing method for analyzing battery capacity degradation. By decomposing the signal into various frequency components, VMD can capture both long-term trends and short-term fluctuations. The optimization of parameters using methods like ADMM ensures accurate decomposition, making it a valuable tool for RUL prediction in battery management systems.

2.1.2. RIME Optimization Algorithm

The RIME algorithm is based on a physics-inspired optimizer that utilizes the growth process of soft and hard ice needles under simulated frost-ice conditions to explore and develop optimization methods. It performs global optimization through an improved forward greedy selection mechanism. Su et al. [28] divided this algorithm into three key stages: initialization of frost-ice crystallization points, introduction of soft frost-ice search strategies, and the hard frost-ice puncturing greedy mechanism. Furthermore, they improved the greedy selection strategy to construct a mathematical model.

RIME is a physics-based optimization method that simulates the growth processes of soft and hard frost under frost-ice conditions to achieve global optimization. A flowchart of the algorithm implementation is shown in Figure 2. The algorithm primarily consists of three steps: initialization of the frost-ice crystallization points, soft frost search strategy, and hard frost puncturing mechanism.

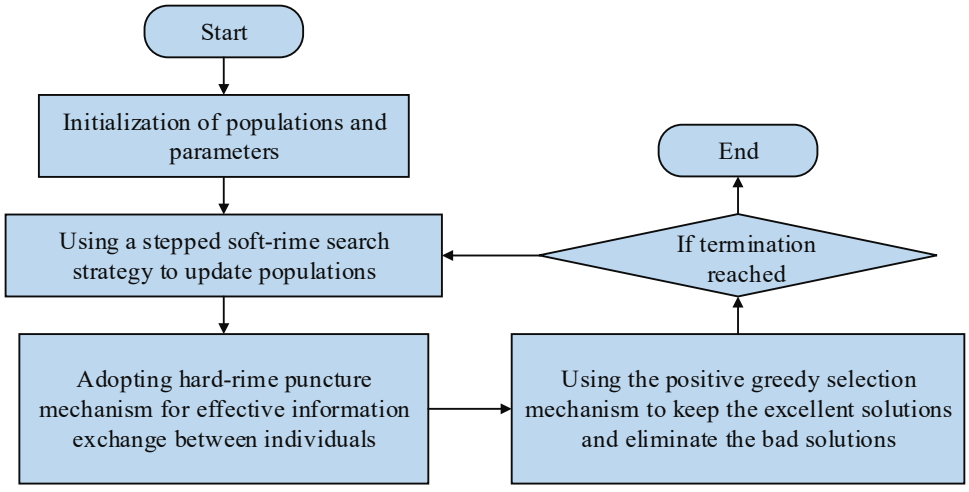


Figure 2. RIME process diagram.

First, the initial solution is randomly generated within the search space, simulating the starting point of the frost-ice crystallization process. Next, the random movement of soft frost particles is simulated to enhance the algorithm's exploration capability, preventing it from falling into local optima. Finally, the rapid growth of hard frost in a specific direction is simulated to strengthen the exploitation capability of the algorithm, accelerating convergence. Additionally, RIME improves the greedy selection strategy by choosing better solutions during each iteration, thus enhancing the global optimization performance. The effectiveness and superiority of the algorithm have been validated through multiple benchmark tests and real-world engineering problems. The position of the RIME frost-ice particle population, denoted by R, is as follows:

$$R = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} \\ x_{21} & x_{22} & \cdots & x_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} \end{bmatrix} \tag{3}$$

In the equation, the frost-ice particle population can be represented by R, where i represents the index of the frost-ice crystal and j represents the index of the frost-ice particle. Furthermore, the fitness value of all frost-ice particles is given by the following equation:

$$F(S_i) = \begin{bmatrix} f\left(\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} \end{bmatrix}\right) & & & \\ f\left(\begin{bmatrix} x_{21} & x_{22} & \cdots & x_{2j} \end{bmatrix}\right) & & & \\ \vdots & & \vdots & \ddots & \vdots \\ f\left(\begin{bmatrix} x_{i1} & x_{i2} & \cdots & x_{ij} \end{bmatrix}\right) & & & \end{bmatrix} \tag{4}$$

In the equation, f represents the fitness of each frost-ice particle. Each frost-ice particle x moves according to a specific pattern before condensation. Particles in a free state will move toward soft ice for condensation, but if they exceed a certain range, condensation will not occur. During the formation of soft ice, the randomness of each particle increases the condensation area, and environmental factors eventually lead to a stable state. Based on the movement characteristics of the frost-ice particles, the position of a frost-ice particle is given by the following equation:

$$\begin{cases} R_{ij}^{new} = R_{best,j} + r_1 \cdot \cos\theta \cdot \beta \cdot \left(\psi \cdot \left(Ub_{ij} - Lb_{ij}\right) + Lb_{ij}\right) \\ r_2 < E \end{cases} \tag{5}$$

In this equation, R_{ij}^{new} represents the updated position of the particle. R denotes the frost-ice population, with R including the particle of the optimal individual in R. r_1 is a random number within the range $(-1, 1)$ that controls the movement direction of the particle and dynamically changes with the particle's movement direction $\cos\theta$ during iterations. β is the environmental factor representing the influence of external conditions. ψ is the particle adhesion factor, a random number within the range $(0, 1)$, used to control the distance between particles. Ub_{ij} and Lb_{ij} represent the upper and lower bounds of the particle's movement space, respectively.

$$\theta = \pi \cdot \frac{h}{10 \cdot H} \tag{6}$$

$$\beta = 1 - \left[\frac{w \cdot h}{H}\right] / w \tag{7}$$

In this equation, h represents the current iteration number. β is a step function. w controls the number of segments in the step function. E is the adhesion coefficient, which increases with the number of iterations and influences the condensation probability of the particles. The expression for E is given as

$$E = \sqrt{(h/H)} \tag{8}$$

As the condensation area of soft frost expands, it exhibits strong randomness and wide coverage, enabling the algorithm to quickly explore the entire search space while balancing global and local optimizations. During the soft frost phase, the algorithm leverages its high randomness and extensive coverage to rapidly traverse the search space, ensuring both global exploration and local exploitation. In the hard frost phase, the algorithm simulates the condensation characteristics of hard frost growing in the same direction, promoting cross-interaction and information exchange among particles, thereby improving the solution accuracy. This cross-exchange mechanism effectively avoids local optima traps and enhances the algorithm's convergence. As hard frost grows and punctures, the cross-replacement of particles is given by the following equation:

$$R_{ij}^{new} = R_{best,j}, r_3 < F^{normr}(S_i) \tag{9}$$

F^{normr} represents the normalized current fitness value, which indicates the probability of the i-th particle being selected. r_3 is a random number within the range $(-1, 1)$.

2.1.3. RIME-VMD

The RIME-VMD algorithm integrates the global search capability of RIME with its efficient parameter updating mechanism to optimize the decomposition layer number K and penalty factor α in VMD [29], thereby determining the optimal parameter combination [K, α]. Its core objective is to use the minimum envelope shrinkage as the fitness function to identify the best parameters. The specific implementation flowchart is shown in Figure 3. The algorithm steps are as follows:

(a) Parameter Initialization:

Set the frost-ice population of RIME, the maximum number of iterations H, the decomposition layer number K, and the upper and lower bounds of the optimization problem to provide the initial conditions for subsequent optimization.

(b) Input of Raw Signal:

Introduce the signal to be processed as input data for the optimization process.

(c) Adhesion Coefficient Calculation:

Based on the current iteration number h and the fitness value from the last iteration, combined with the random adhesion coefficient $r_2 < E$, update the particle's new position to enhance the algorithm's diversity.

(d) Fitness Value Normalization:

Calculate the normalized fitness value and use the minimum envelope shrinkage to adjust the particle positions, balancing global and local searches.

(e) Position Update:

Compare the fitness values of the new position and the current particle, select the better position as the updated particle state, and refresh the best fitness value.

(f) Loop Termination Condition:

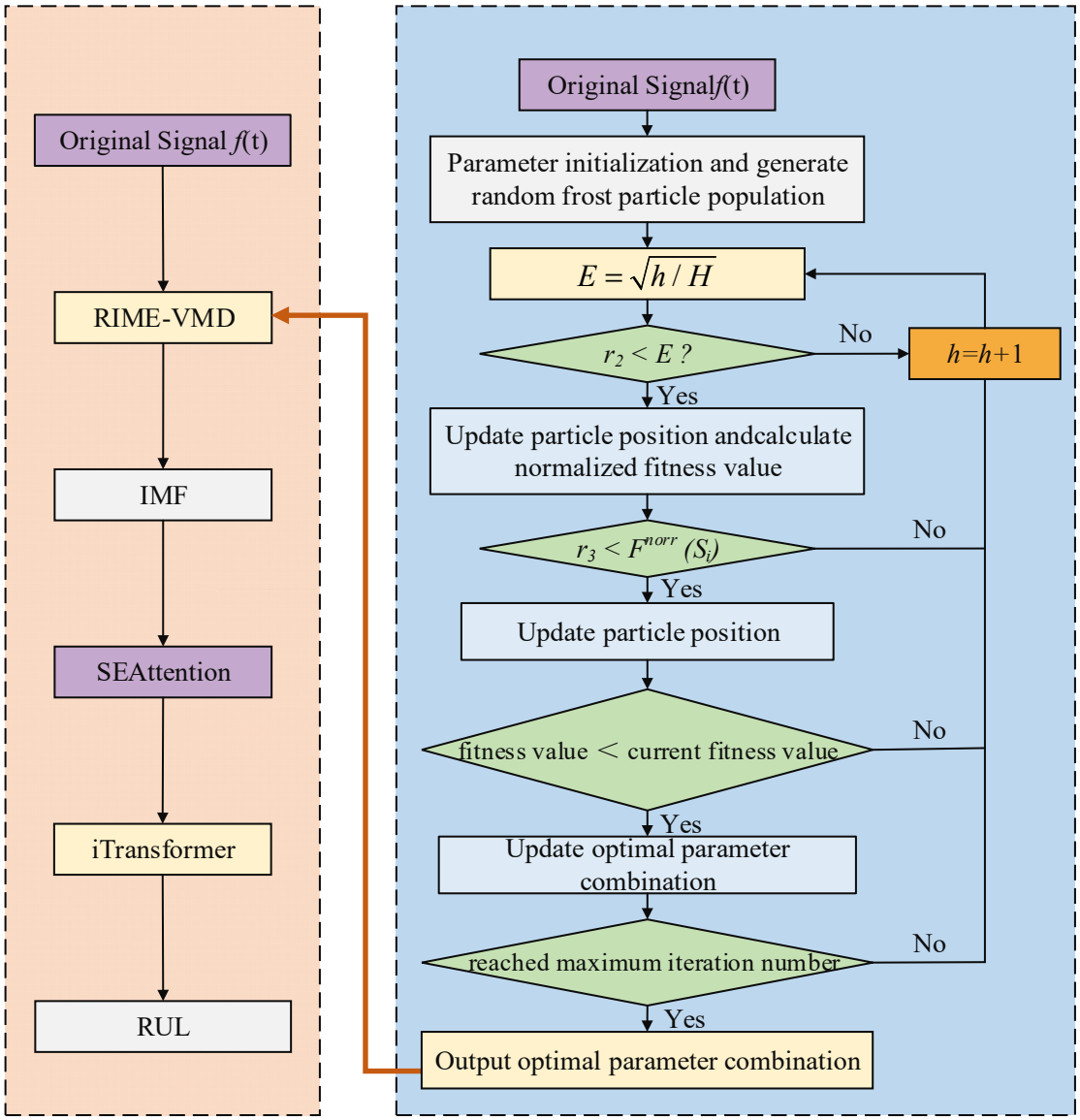


Figure 3. RIME-VMD process diagram.

Check whether the termination conditions are satisfied. If h $<$ E and the target error has not been reached, proceed to the next iteration. If the conditions are satisfied, output the final parameter combination (K, α), substitute it into the VMD for signal decomposition, and output the final result.

RIME-VMD offers significant advantages over the traditional VMD in terms of battery capacity decomposition and RUL prediction. By incorporating the RIME optimization algorithm, it adaptively adjusts key parameters, such as the number of modes K and penalty factor α, improving the decomposition accuracy and robustness of battery capacity signals, reducing noise interference, and accurately capturing both long-term degradation trends and short-term fluctuations, thus enhancing RUL prediction accuracy. At the same time, RIME-VMD reduces computational complexity and improves computational efficiency through global search and local optimization mechanisms, making it more suitable for real-time battery management systems.

2.2. Squeeze-and-Excitation Networks

SEAttention [30] is an effective channel attention mechanism that explicitly models inter-channel dependencies through feature recalibration, dynamically adjusting feature

responses. Compared to traditional convolution operations, SEAttention can highlight important features while suppressing irrelevant or redundant information, thereby enhancing the model's feature representation capabilities. A structural diagram of this module is shown in Figure 4, and its specific implementation steps are described as follows:

(a) Feature Mapping

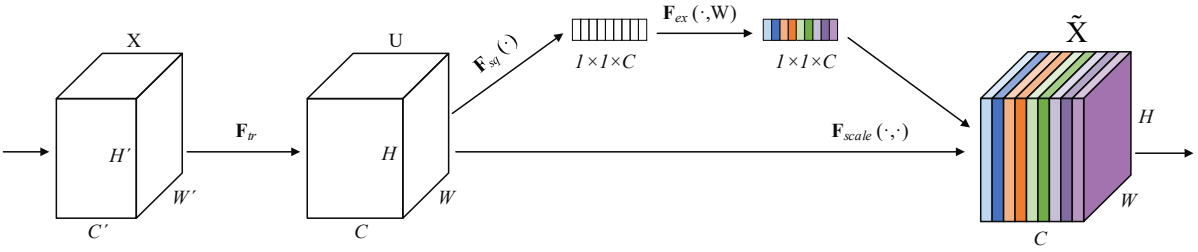


Figure 4. Schematic diagram of SEAttention module structure.

First, the input feature map $\mathbf{X} \in R^{H\times W\times C}$ is mapped to a specific spatial representation through convolution operations as follows:

$$\mathbf{U} = F_{\text{conv}}(\mathbf{X}) \tag{10}$$

Here, U has the same dimensions as X, with H and W representing the height and width of the feature map and C denoting the number of channels. This step essentially performs convolutional computation, providing input for the subsequent attention mechanism.

(b) Global Information Compression

A global average pooling operation compresses the spatial dimension into the channel dimension to obtain a global description for each channel as follows:

$$\mathbf{Z}_c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\mathbf{U}_{i,j,c} \quad \forall c \in \{1,2,\ldots,C\} \tag{11}$$

(c) Weight Generation

The compressed feature vector Z passes through a two-layer fully connected network to generate the weight for each channel, which is calculated as follows:

$$\begin{cases} \mathbf{S} = F_{\text{fc}}(\mathbf{Z}) = \sigma(\mathbf{W}_2\delta(\mathbf{W}_1\mathbf{Z})) \\ \mathbf{W}_1 \in R^{\frac{C}{r}\times C} \\ \mathbf{W}_2 \in R^{C\times\frac{C}{r}} \end{cases} \tag{12}$$

where, δ is the ReLU activation function, introducing nonlinearity. σ is the sigmoid activation function, normalizing the weight range $\mathbf{W}_1$ and $\mathbf{W}_2$ are learnable parameters of connected layers. r is the channel reduction ratio used to reduce computational complexity. Through this process, the generated weight S describes the importance of each channel.

(d) Feature Reweighting

The generated channel weights S are multiplied element-wise with the original feature map along the channel dimension, achieving a dynamic adjustment of the channel importance:

$$\begin{cases} \mathbf{S} = F_{\text{fc}}(\mathbf{Z}) = \sigma(\mathbf{W}_2\delta(\mathbf{W}_1\mathbf{Z})) \\ \mathbf{W}_1 \in R^{\frac{C}{r}\times C} \\ \mathbf{W}_2 \in R^{C\times\frac{C}{r}} \end{cases} \tag{13}$$

The reweighted output feature $\mathbf{X}'$ highlights features that are more relevant to the task while suppressing less significant features.

Thus, in data processing, artificial intelligence and neural networks play a crucial role [31]. The SEAttention model can dynamically model the dependencies between channels, enhancing feature selection capabilities. With minimal parameter and computational overhead, the model significantly improves the network's ability to focus on critical features, providing strong support for enhancing task performance.

2.3. Informer Based on Transformer Improved Model

2.3.1. Transformer Model

The Transformer model is a deep learning architecture based on the self-attention mechanism [32]. Unlike traditional recurrent neural networks (such as LSTM and GRU), its distinctive feature lies in using the self-attention mechanism to capture dependencies between elements in the input and output sequences while supporting parallel processing of the entire sequence. This design enables the Transformer to efficiently capture long-range dependencies and reduces the reliance on step-by-step computations inherent in traditional sequential models, thereby significantly enhancing performance. The key components of the Transformer model are introduced below:

(a) Self-Attention Mechanism

The self-attention mechanism is the core of the Transformer, designed to capture global dependencies among elements in the input sequence. Calculating the correlations between any two positions in the input sequence generates attention weights for each position and applies these weights to reweight the input features. The structure diagram of the Transformer is shown in Figure 5. The calculation process is as follows:

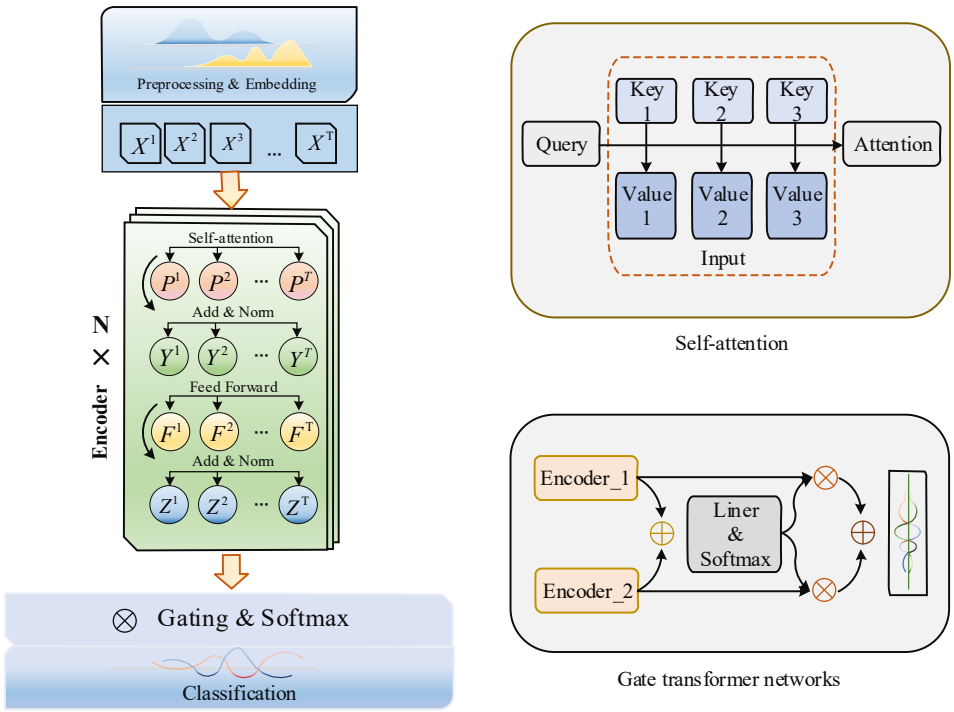


Figure 5. Structure and self-attention mechanism in Transformer.

The input features are first mapped into Query, Key, and Value through linear transformations as follows:

$$\begin{aligned} Q_i &= W_Q x_i \\ K_i &= W_K x_i \\ V_i &= W_V x_i \end{aligned} \tag{14}$$

The weight matrices $\mathbf{W}_Q$,$\mathbf{W}_K$,$\mathbf{W}_V$ are learnable parameters used to map the input features into subspaces for queries, keys, and values. The similarity between the query and the key is computed using the dot product, followed by Softmax normalization to ensure that the sum of the weights is 1:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V, \tag{15}$$

Here, d_k represents the dimensionality of the key vectors used to scale the dot product results and prevent numerical instability. The output of the self-attention mechanism is a weighted combination of the input sequence, dynamically adjusting the importance of each element.

(b) Multi-Head Attention

To further enhance the model's representation capability, the Transformer employs a multi-head attention mechanism. This mechanism splits queries, keys, and values into multiple heads, allowing each head to independently compute attention. The results are then concatenated and projected back into the output space, as follows:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \text{head}_2, \ldots, \text{head}_h)\mathbf{W}_O \tag{16}$$

Each attention head is computed as

$$\text{head}_i = \text{Attention}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) \tag{17}$$

Among them, $\mathbf{W}_O$ is the weight matrix used for the final linear transformation. Concat represents concatenating the outputs of each head together.

(c) Positional Encoding

Since the Transformer does not rely on sequence order, additional positional information is introduced through positional encoding. A common method is to add fixed sinusoidal and cosinusoidal function values to the input features, as follows:

$$\begin{aligned} \text{PE}(pos)_{2i} &= sin\left(\frac{pos}{10{,}000^{\frac{2i}{d_{obs.}}}}\right) \\ \text{PE}(pos)_{2i+1} &= cos\left(\frac{pos}{10{,}000^{\frac{2i}{d_{obs.}}}}\right) \end{aligned} \tag{18}$$

where pos is the position of the vector, i is the index of the position vector, and dim is the dimensionality of the observation data.

(d) Feedforward Network

Each layer of the Transformer includes a fully connected feedforward network (FFN) for nonlinear feature transformation. The FFN consists of two linear transformations, with an activation function between them:

$$\text{FFN}(\mathbf{x}) = \max(0, \mathbf{x}\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2 \tag{19}$$

where $\mathbf{W}_1$,$\mathbf{W}_2$,$\mathbf{b}_1$ and $\mathbf{b}_2$ are learnable parameters. Through the above modules, the Transformer eliminates recursive dependencies and can efficiently process large-scale data. The self-attention mechanism simultaneously models global information, making it well suited for applications such as lithium battery capacity prediction.

2.3.2. Informer Model

The Informer model is an improved Transformer framework designed specifically for long-sequence time-series forecasting tasks. Its overall architecture is based on the standard encoder-decoder structure, but it introduces innovative mechanisms such as the ProbSparse attention mechanism, self-attention distillation, and a generative decoder to significantly enhance computational efficiency and predictive performance. The detailed structure of the Informer is shown in Figure 6, with its main components described as follows:

(a) ProbSparse Self-Attention Mechanism

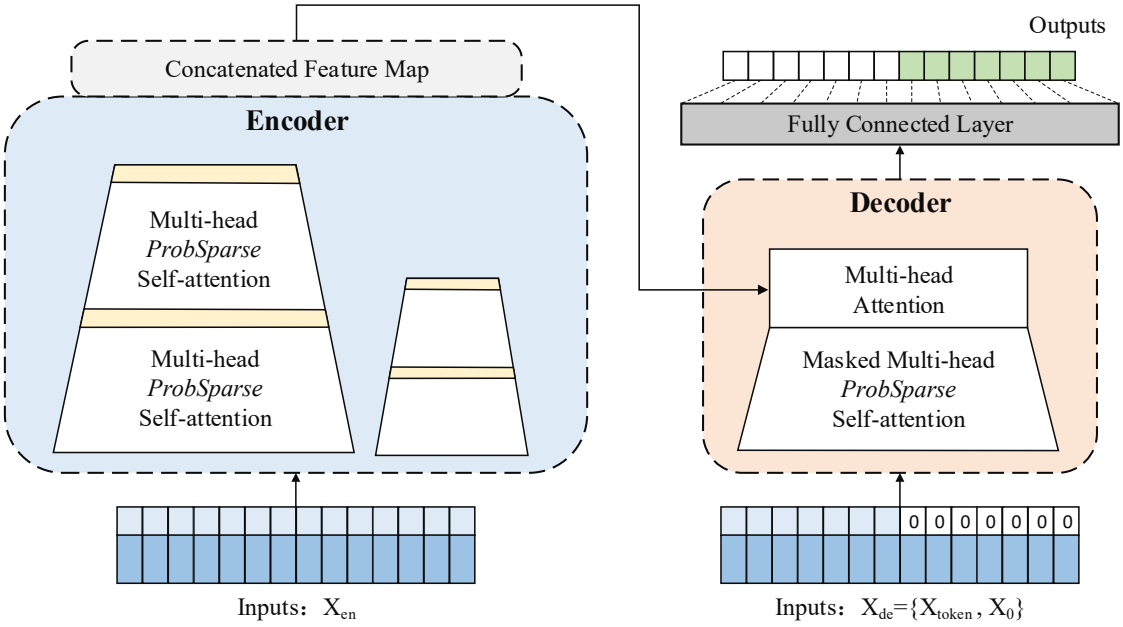


Figure 6. Structure of Informer.

The ProbSparse self-attention mechanism proposed in the model leverages the long-tail distribution of attention scores, computing only the most critical queries. It measures the importance of queries using KL divergence, and the sparsity measurement formula is as follows:

$$M(q_i, K) = \ln \sum_{j=1}^{L_K} e^{q_i k_j^\top / \sqrt{d}} - \frac{1}{L_K} \sum_{j=1}^{L_K} q_i k_j^\top / \sqrt{d} \tag{20}$$

Based on the values of $M(q_i, K)$, the top-u most critical queries are selected. The sampling strategy is defined as

$$u = c \cdot \ln L_Q \tag{21}$$

where c is a constant factor, and L_Q represents the length of the query sequence. The final formula for ProbSparse attention is given as

$$A(Q, K, V) = \text{Softmax}\left(\frac{\hat{Q}K^\top}{\sqrt{d}}\right)V \tag{22}$$

Here, Q is a subset of queries selected after sparsity filtering. This approach ensures efficient computation by focusing only on the most significant query-key pairs, thereby reducing redundancy without compromising accuracy.

(b) Self-Attention Distillation

To further reduce redundant features introduced by multi-layer stacking in the encoder, Informer applies distillation operations at each layer, progressively reducing the temporal dimension of the input. The distillation formula for each layer is

$$X_{j+1} = \text{MaxPool}\big(\text{ELU}(\text{Conv1D}(X_j))\big) \tag{23}$$

Here, Conv1D is a one-dimensional convolution operation used to extract local temporal patterns. MaxPool reduces the sequence length by half, optimizing spatial complexity. With self-attention distillation, the encoder's overall spatial complexity is optimized to

$$O_{Informer} = O((2-\epsilon)L\log L) \tag{24}$$

where ϵ is a very small constant. The specific structure of the Informer encoder is shown in Figure 7.

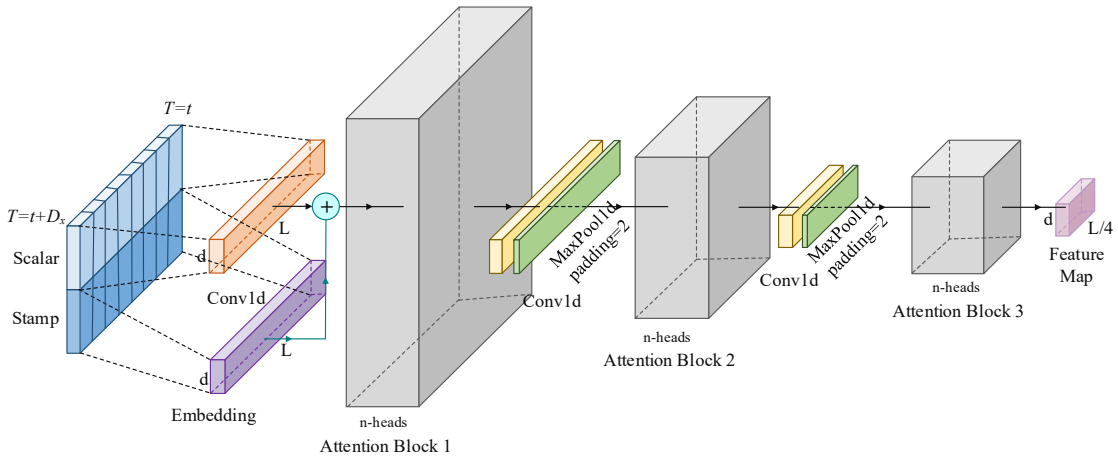


Figure 7. Structure of the Informer Encoder.

(c) Multivariate Correlation Processing

By applying the attention mechanism along the variable dimension, Informer generates a multivariate correlation map, enabling the model to capture the interdependencies among different variables. The attention score matrix A is calculated as

$$A_{ij} = \text{softmax}\left(\frac{Q_i K_j^\top}{\sqrt{d_k}}\right) \tag{25}$$

where A describes the correlation between variable i and variable j. This is critical for multivariate time-series forecasting, as it helps the model understand how different variables influence one another.

3. Description of the Proposed Hybrid Model

The hardware platform used in this experiment consists of an Intel(R) Core(TM) i5-9300H CPU @ 2.30 GHz with 16.0 GB RAM. The programming language used is Python 3.9, and the development environment is PyCharm. All deep learning models are implemented using the PyTorch1.9.0 framework.

The proposed RIME-VMD-SEInformer model integrates the unique attributes and advantages of RIME, VMD, SEAttention, and Informer to predict the RUL of lithium-ion batteries. This section describes the proposed hybrid ensemble learning framework for RUL prediction, including the data preprocessing phase and the SeInformer ensemble learning phase. The detailed processes for each phase are introduced below. The proposed RUL prediction framework is illustrated in Figure 8.

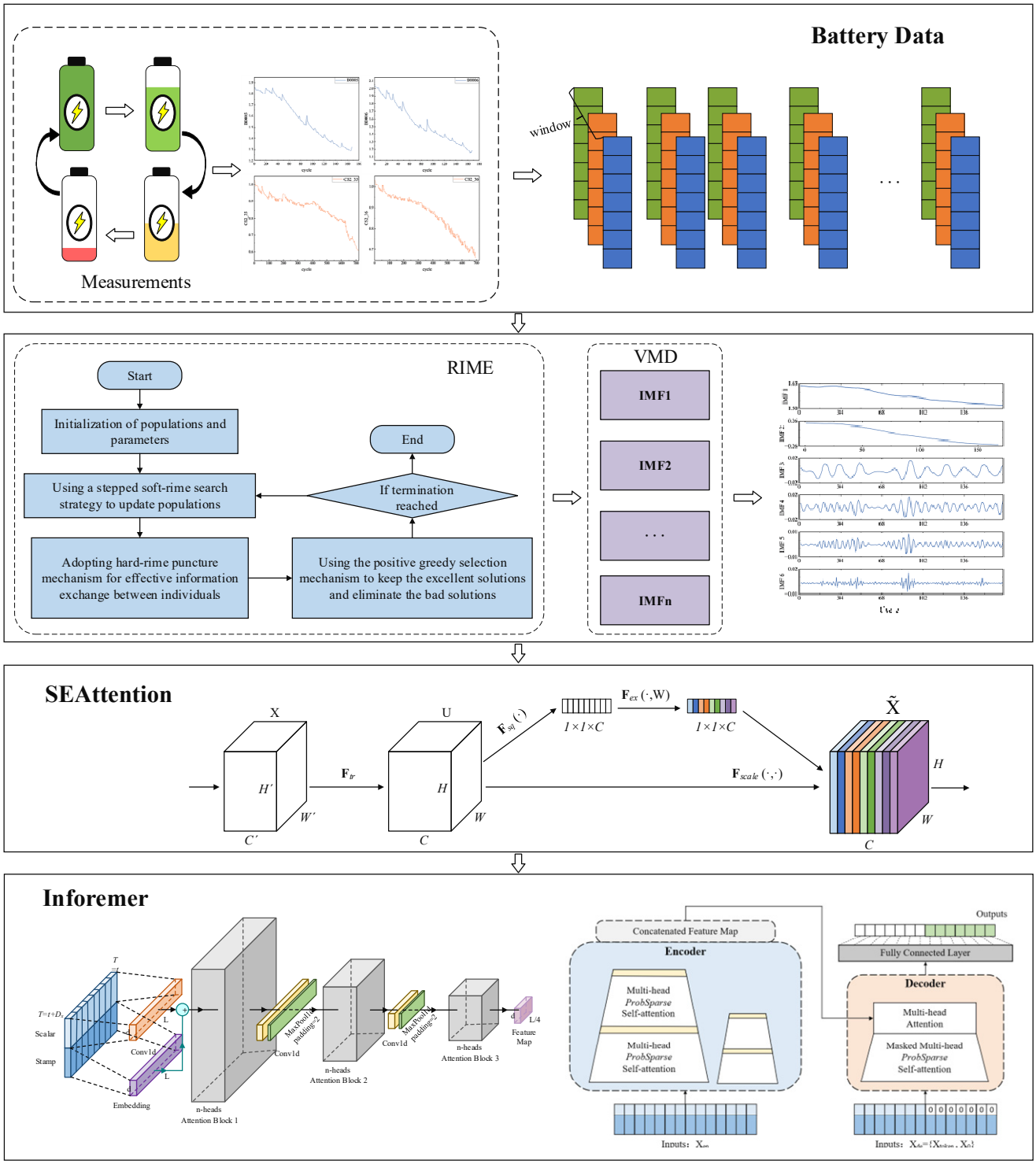


Figure 8. Based on the RIME-VMD-SEInformer RUL prediction framework.

Specifically, RIME simulates the growth processes of soft and hard frost, employing multi-stage search strategies to enhance parameter optimization efficiency and global search capability. It utilizes dynamic search strategies to adaptively optimize VMD parameters, ensuring the physical interpretability and stability of the decomposition results. Additionally, through particle interaction and multi-point exploration, RIME overcomes the tendency of traditional optimization algorithms to become trapped in local optima.

VMD decomposes the original signal into multiple intrinsic mode functions (IMFs), each corresponding to distinct frequency components, thereby making complex time-series features more prominent. Compared to traditional EMD, VMD is built on a rigorous mathematical foundation, avoiding endpoint effects and modal aliasing issues, and providing more reliable feature decomposition. Furthermore, VMD captures the multi-scale dynamic characteristics of signals, offering robust support for analyzing complex battery degradation behaviors.

SEAttention dynamically assigns weights to each feature through squeeze-and-excitation mechanisms, highlighting important features while suppressing redundant information. With low computational complexity, the module does not add significant complexity to the model, while substantially enhancing the feature representation capability. By explicitly modeling inter-channel dependencies, SEAttention further adapts to complex multivariate time series, making it particularly suitable for modeling the multimodal features of lithium batteries.

Informer models long- and short-term dependencies in sequences using attention mechanisms along the temporal dimension and captures variable interactions with feed-forward networks along the feature dimension, improving the precision of time-series predictions. Compared to traditional sequence models, Informer excels at capturing decay patterns and trend variations over extended time horizons.

In summary, the proposed model leverages RIME's optimization capabilities, VMD's signal decomposition strengths, SEAttention's channel-wise feature weighting, and Informer's efficient sequence modeling. This integrated framework effectively extracts key features and captures complex dependencies, significantly improving the accuracy of RUL predictions for lithium batteries.

3.1. RUL Prediction Model

To address the complex fluctuations and capacity regeneration phenomena in battery capacity data, this study employs the RIME-VMD technique to decompose the battery capacity sequence into N intrinsic mode functions (IMFs). Each IMF represents a single-frequency signal component that aligns with the frequency characteristics of the original signal, providing the model with rich multi-frequency feature information.

To ensure model training stability and consistency in feature scaling, all input data are normalized using the following formula:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{26}$$

where $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of the data, respectively.

Subsequently, the study applies a sliding time window technique to segment the time-series data into fixed-length subsequences for constructing input-output data pairs. The principle of the sliding window is as follows: Given a window size W and a step size S, the raw capacity data is divided into time-series segments of length W. The starting position of each window shifts forward by S, generating multiple subsequences that span the entire dataset. Each subsequence is treated as an independent sample and labeled with its corresponding RUL to form input-output pairs. The decomposed IMFs are used as input

features, which are sequentially passed through the hybrid prediction model to complete the RUL prediction process.

By combining the sliding window technique for constructing time-series data, VMD for extracting multi-frequency features, and the hybrid prediction model leveraging deep learning methods, this study achieves a high-accuracy RUL prediction for lithium-ion batteries. The multi-module design of the framework ensures robustness and predictive performance, even under capacity regeneration conditions.

3.2. Lithium Battery Data Set Description

This study utilized three publicly available battery datasets: the CALCE Battery Dataset, the NASA Battery Dataset, and the SNL Battery Dataset. A detailed description of the datasets and battery parameters is provided in Table 2. Below is an introduction to the NASA and CALCE datasets.

Table 2. This article selects detailed battery parameters.

Type	Capacity (Ah)	Cutoff Voltage (v)	Discharge Current (A)	Temperature (°C)
B0005	2	2.7	2	24
B0006	2	2.5	2	24
CS2 35	1.1	2.7	1.1	1
CS2 36	1.1	2.5	1.1	1

(a) NASA

The dataset used in this study was obtained from NASA's publicly available battery cycling test data. The testing equipment includes Arbin BT2000, Cadex C8000, and Agilent 34970A. The dataset contains six lithium-ion batteries (B5, B6, B29, B30, B45, and B46). These batteries were divided into three groups, with two batteries in each group, based on the experimental conditions. To investigate the performance variations under different environmental temperatures, three distinct test modes were employed, with cyclic aging tests conducted for each mode.

This study focuses on the B5 and B6 datasets, which reflect the aging behavior under room temperature conditions. During charging, the batteries were charged in the constant current (CC) mode with a current of 1.5 A. Once the voltage reached 4.2 V, the charging mode switched to constant voltage (CV), maintaining 4.2 V until the charging current dropped to 20 mA, at which point the charging process was complete. During discharging, batteries B5 and B6 were discharged with a constant current of 2 A to cutoff voltages of 2.7 V and 2.5 V, respectively. This mode simulates battery operation under conventional environmental conditions.

(b) CALCE

The CALCE dataset was tested using Maccor Series 4000, Arbin BT2000, and NI DAQ. Cyclic aging tests were conducted on multiple prismatic lithium-ion batteries with a nominal capacity of 1.1 Ah. For this study, two batteries (CS2_35 and CS2_36) were selected for analysis. Experiments were conducted at an ambient temperature of 25 °C, utilizing constant current-constant voltage (CC-CV) charging and constant current discharging.

During charging, the batteries were charged with a constant current of 0.5 C until the voltage reached 4.2 V, after which they were switched to constant voltage charging while maintaining 4.2 V until the charging current dropped to 0.05 A. Following this, the batteries were discharged at a constant current of 1C to a cutoff voltage of 2.7 V, completing one full cycle.

3.3. Evaluation Indicators

This study employs four common metrics to evaluate the performance of the hybrid model: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2). The definitions are as follows:

$$\mathrm{MAE} = \frac{1}{T}\sum_{t=1}^{T}\left|C_t - \hat{C}_t\right| \tag{27}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(C_t - \hat{C}_t\right)^2} \tag{28}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{t=1}^{T}\left|\frac{C_t - \hat{C}_t}{C_t}\right| \tag{29}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{T}\left[C_t - \hat{C}_t\right]^2}{\sum_{t=1}^{T}\left[C_t - \overline{C_t}\right]^2} \tag{30}$$

where: C_t represents the actual value, and $\hat{C}_t$ represents the predicted value. T represents the number of time points. MAE measures the average absolute difference between the predicted and actual values, directly reflecting the magnitude of the prediction errors. RMSE represents the square root of the mean squared errors, emphasizing larger errors, and is more sensitive to outliers or data points with significant deviations. MAE indicates the percentage of prediction errors relative to the actual values, providing a dimensionless metric that facilitates comparison across datasets of different scales. R^2 reflects the model's ability to explain the variance in the data and serves as a comprehensive indicator of the model's overall performance.

4. Experimental Results and Analysis

In the RUL prediction experimental phase, two sets of experiments were conducted. First, ablation experiments were performed on the selected datasets to validate the necessity of each component in the RIME-VMD-SE-Informer model for RUL prediction. Second, state-of-the-art (SOA) comparative experiments were carried out to evaluate the performance advantages of the proposed method, verifying its effectiveness and applicability in RUL prediction.

4.1. RIME-VMD Decomposition of Battery Capacity Data

Battery capacity data typically exhibit complex, non-stationary time-series characteristics, including long-term capacity degradation trends, short-term fluctuations (such as capacity regeneration phenomena), and high-frequency noise. The core function of RIME-VMD is to decompose these complex signals into a set of intrinsic mode functions (IMFs) with different frequency ranges. Specifically, VMD operates on a rigorous mathematical optimization framework, minimizing the signal's variational model to decompose it into several stationary mode components, thereby reducing noise interference. The frost-ice optimization algorithm further enhances the accuracy and stability of the decomposition by performing global searches and adaptive parameter adjustments, mitigating modal aliasing and endpoint effects. For the RIME-VMD-based decomposition model, the noise tolerance was set to 0, and the convergence error tolerance for iterations was set to 1×10^{-7}. The modal decomposition results based on RIME-VMD are shown in Figure 9.

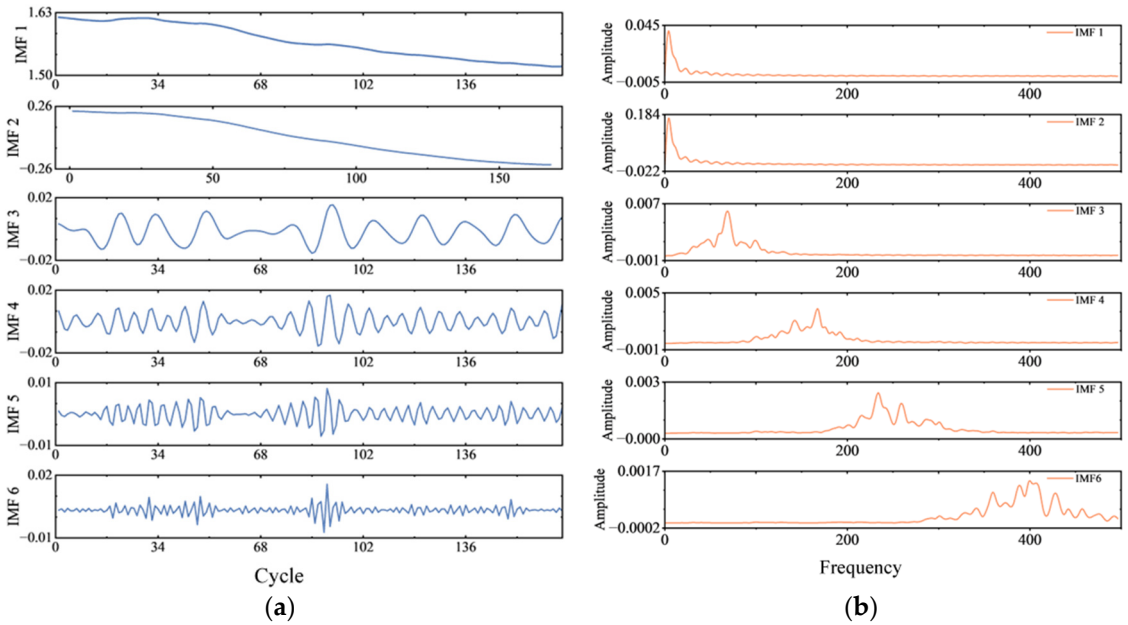


Figure 9. Taking the b0005 battery as an example, the IMFs (**a**) and corresponding frequencies (**b**) after RIME-VMD decomposition.

Additionally, IMFs can be dynamically learned by SEAttention. While RIME-VMD's role is to decompose the raw signal into multiple modal components, extracting features across different frequency ranges, SEAttention dynamically determines which modes are critical and which are not. This eliminates the need to analyze each IMF's frequency range or specific contribution individually. SEAttention automatically adjusts weights according to the model's optimization objectives, such as minimizing error metrics and ensuring that the model prioritizes the most impactful features.

4.2. Selection of Experimental Parameters

To ensure the predictive accuracy of the model, appropriate parameters must be selected. In the SEAttention module, the compression ratio was set to 2. This section specifically discusses the determination of the training sample size m. The value of m was determined using an iterative experimental approach utilizing the baseline Informer model. To evaluate the model's feature extraction and prediction performance under limited training data conditions, capacity data from four batteries—B0005, B0006, CS2-35, and CS2-36—were selected as test samples. This study investigated the model's performance across different scales of training data. The training sample size m ranged from 5% to 30%, and the effect of increasing the training data proportion on the model performance was observed [33]. Figure 10 illustrates the prediction errors for different training data proportions.

The results indicate that the prediction errors, including the RMSE, MASE, and MSE, decrease as the proportion of training data increases. Notably, the proposed model achieves the highest accuracy when trained on 30% of the training set. Therefore, the proposed model uses 30% of the total degradation data for training.

For the RIME-VMD-SEInformer model developed in this study, the selected parameters for the prediction phase are listed in Table 3.

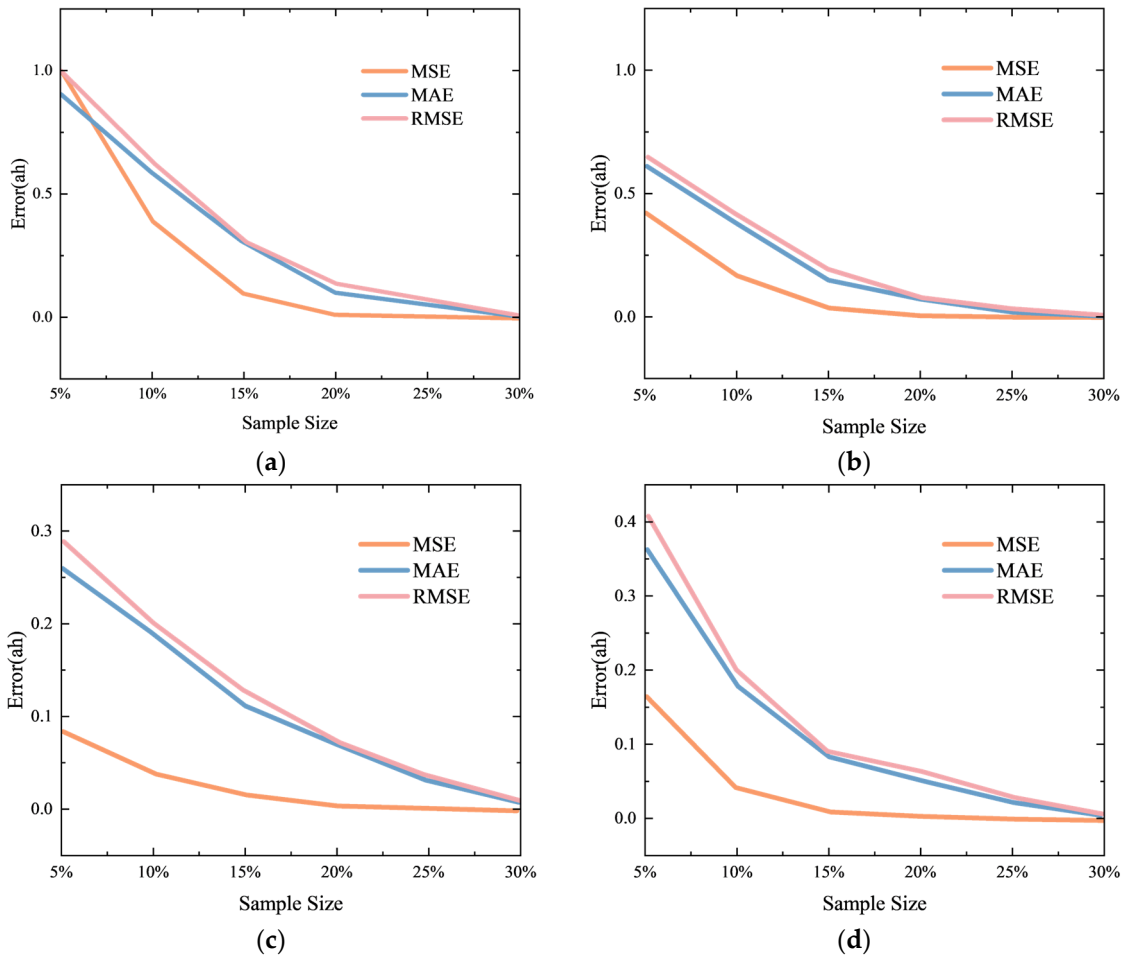


Figure 10. Errors of different training sets on b0005 (**a**), b0006 (**b**), cs2_35 (**c**), cs2_36 (**d**).

Table 3. Forecast frame parameters.

Parameters	Value
Batch size	32
Training-to-testing split	3:7
Encoder layers	2
Decoder layers	1
Label length	4
Attention heads	8
Dropout rate	0.2
Learning rate	0.0005

4.3. Ablation Comparative Experiment

To evaluate the contribution of each component in the proposed RIME-VMD-SEInformer framework, this study conducted ablation experiments to systematically analyze the impact of each module on the overall model performance using the selected datasets. The following six variant models were designed and evaluated:

- w/o RV: To validate the role of the decomposition module (RIME-VMD), this model removes the RIME-VMD module and evaluates its effect on model performance.
- w/o S: To demonstrate the effectiveness of SEAttention, this model excludes the SEAttention module, leaving only the RIME-VMD and Informer components.
- w/o Informer: To verify the importance of the Informer module, this model replaces the Informer with a standard Transformer to observe its impact on model performance.

The results of the ablation experiments are detailed in Table 4, and the findings are shown in Figure 11.

Table 4. The results of ablation comparative experiments.

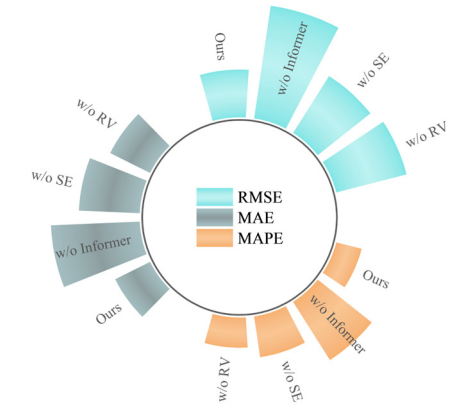
Datasets	Model	R2	RMSE	MAE	MAPE
B0005	w/o RV	0.96561	0.023205	0.014267	0.00968
	w/o SE	0.96668	0.021481	0.019512	0.01316
	w/o Informer	0.89785	0.036396	0.028458	0.02200
	Ours	0.98513	0.01526	0.012455	0.00837
B0006	w/o RV	0.96521	0.024593	0.014746	0.01089
	w/o SE	0.94287	0.02974	0.026062	0.01878
	w/o Informer	0.70969	0.064060	0.062069	0.04349
	Ours	0.97406	0.021233	0.015162	0.01068
CS2_35	w/o RV	0.99219	0.005495	0.003958	0.00423
	w/o SE	0.98675	0.005699	0.004450	0.00436
	w/o Informer	0.90583	0.025151	0.019977	0.02659
	Ours	0.99725	0.004324	0.00312	0.00390
CS2_36	w/o RV	0.98169	0.011318	0.007762	0.01018
	w/o SE	0.98766	0.008232	0.006393	0.00754
	w/o Informer	0.78352	0.038585	0.008374	0.04243
	Ours	0.99581	0.006523	0.004667	0.00586

Compared to models without key modules, the complete RIME-VMD-SEInformer model demonstrates significant advantages in terms of predictive performance. Taking the B0005 dataset as an example, incorporating the RIME-VMD module reduced the MAPE, MAE, and RMSE by 13.5%, 12.7%, and 34.3%, respectively. This improvement is attributed to the RIME-VMD module's efficient decomposition of complex signals during data preprocessing, particularly its ability to mitigate capacity fluctuations and noise interference. By decomposing the raw signal into distinct modes, the model more accurately captures both the long-term trends and local details of capacity degradation, significantly enhancing its predictive capability.

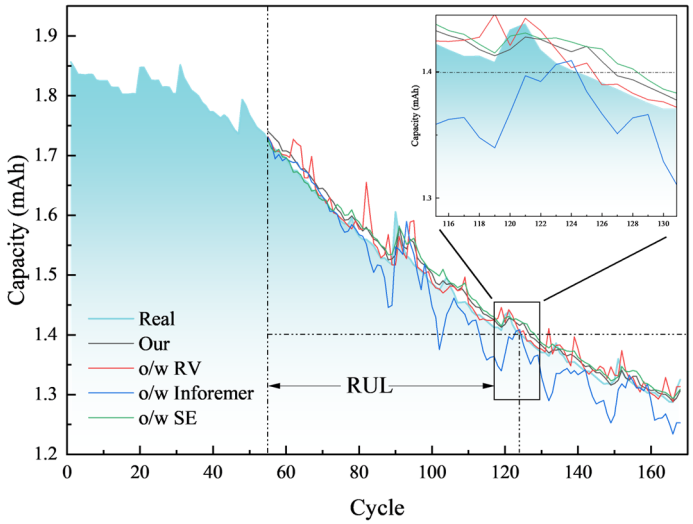
The SEAttention module further strengthens the model's focus on critical features through dynamic weight allocation. For instance, in the CS2_36 dataset, the addition of this module improved R^2 from 0.9817 to 0.9958, indicating more efficient feature selection. SEAttention highlights features closely associated with capacity degradation through squeeze-and-excitation mechanisms while effectively suppressing redundant information.

The Informer module optimizes the modeling of long-term sequences, significantly improving the model's performance in capturing complex dependencies. Experimental results show that removing the Informer module increased the RMSE on the B0006 dataset from 0.0212 to 0.0641, indicating a substantial decline in the model's ability to capture long- and short-term temporal features. Compared to the standard Transformer, the Informer's sparse attention mechanism and self-attention distillation technique reduce computational complexity while significantly enhancing prediction accuracy, particularly when handling highly volatile data.

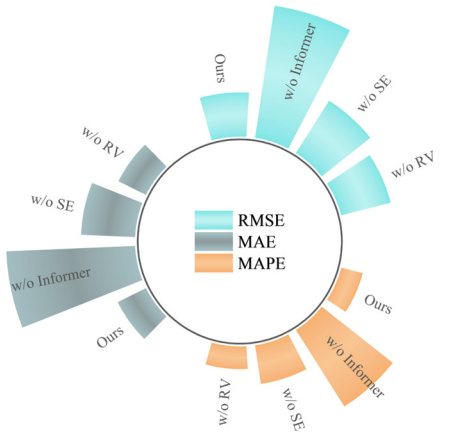
In summary, the combined contributions of RIME-VMD, SEAttention, and Informer—through signal decomposition, feature enhancement, and time-series modeling—enable the RIME-VMD-SEInformer model to achieve breakthroughs in both performance and efficiency for lithium battery RUL prediction tasks.



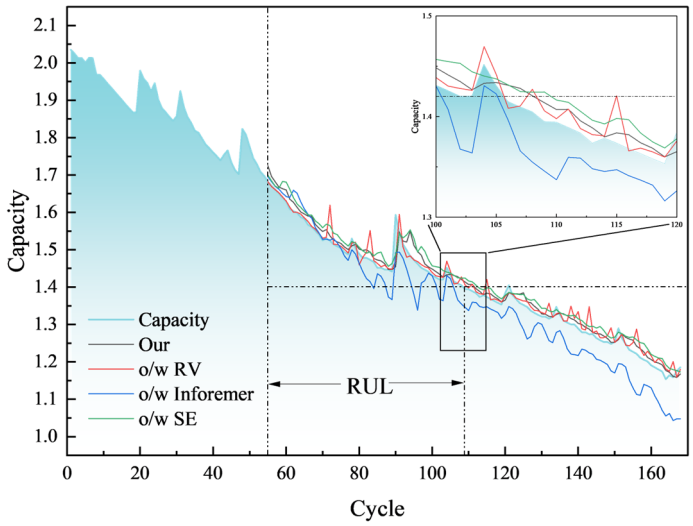
(**a**)



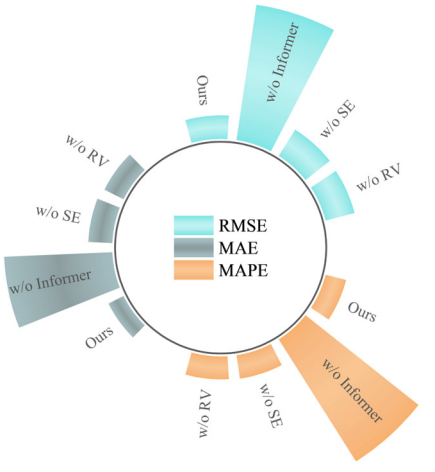
(**b**)



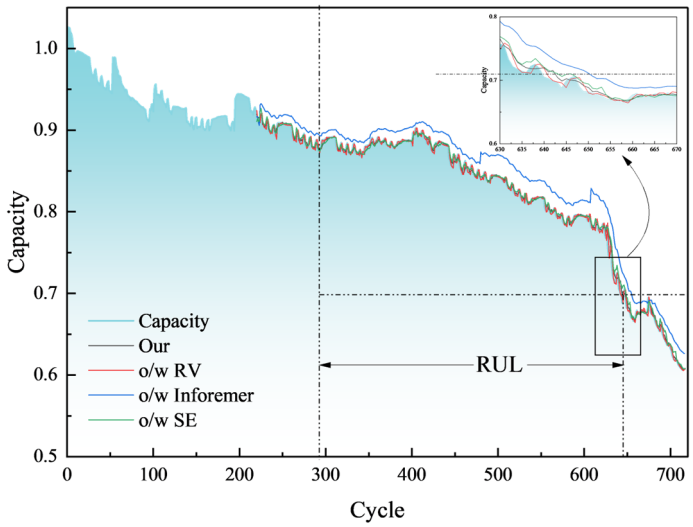
(**c**)



(**d**)



(**e**)



(**f**)

Figure 11. *Cont.*

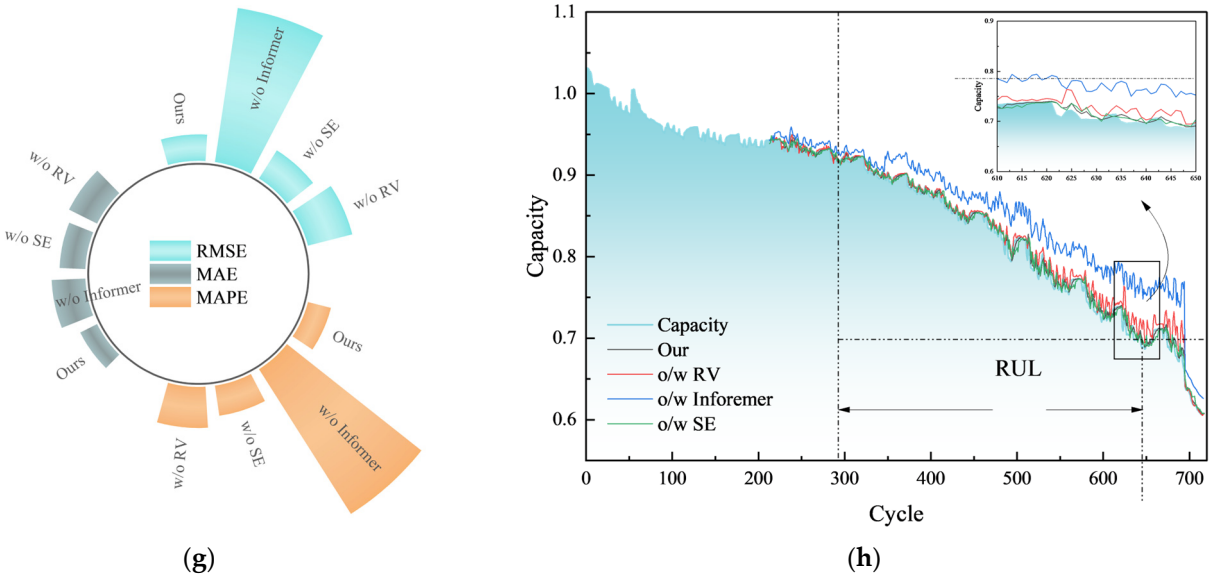


Figure 11. The comparison of evaluation indexes and the prediction effect of each model are in b0005 (**a**,**b**), b0006 (**c**,**d**), cs2_35 (**e**,**f**), and cs2_36 (**g**,**h**).

4.4. Soa Comparison Test

To validate the superior performance of the proposed RIME-VMD-SEInformer model, SOA comparative experiments were conducted. Four algorithms were selected for experimental comparison and evaluation: WOA-VMD-SSALSTM [34], De-Transformer [35], CEEMDAN-Transformer [36], and FP-ARIMA-Gaussian [37]. The detailed comparison results between RIME-VMD-SEInformer and other SOA algorithms are presented in Table 5 and illustrated in Figure 12.

Table 5. Comparison between RIME-VMD-SEInformer and other advanced models.

Datasets	Model	R2	RMSE	MAE	MAPE
B0005	WOA-VMD-SSALSTM	0.8138	0.03464	0.02881	0.02090
	De-Transformer	0.8639	0.02902	0.02439	0.01721
	CEEMDAN-Transformer	0.9160	0.02326	0.01869	0.01302
	CRP-FP-ARIMA-Gaussian	0.8400	0.03212	0.03016	0.02161
	Ours	0.9851	0.01526	0.012455	0.00837
B0006	WOA-VMD-SSALSTM	0.7632	0.04822	0.03594	0.02751
	De-Transformer	0.8762	0.02956	0.02205	0.01548
	CEEMDAN-Transformer	0.9141	0.01905	0.01869	0.01361
	CRP-FP-ARIMA-Gaussian	0.8564	0.03483	0.02897	0.02196
	Ours	0.9740	0.021233	0.015162	0.01068
CS2_35	WOA-VMD-SSALSTM	0.8054	0.09005	0.07007	0.01427
	De-Transformer	0.8493	0.07817	0.05044	0.01175
	CEEMDAN-Transformer	0.8984	0.06417	0.04327	0.00971
	CRP-FP-ARIMA-Gaussian	0.8429	0.08089	0.06414	0.01291
	Ours	0.997	0.004324	0.00312	0.00390
CS2_36	WOA-VMD-SSALSTM	0.8249	0.09926	0.06321	0.02032
	De-Transformer	0.8612	0.08669	0.05556	0.01938
	CEEMDAN-Transformer	0.9093	0.07007	0.04681	0.01571
	CRP-FP-ARIMA-Gaussian	0.8518	0.09131	0.05874	0.02201
	Ours	0.9958	0.00652	0.004667	0.00586

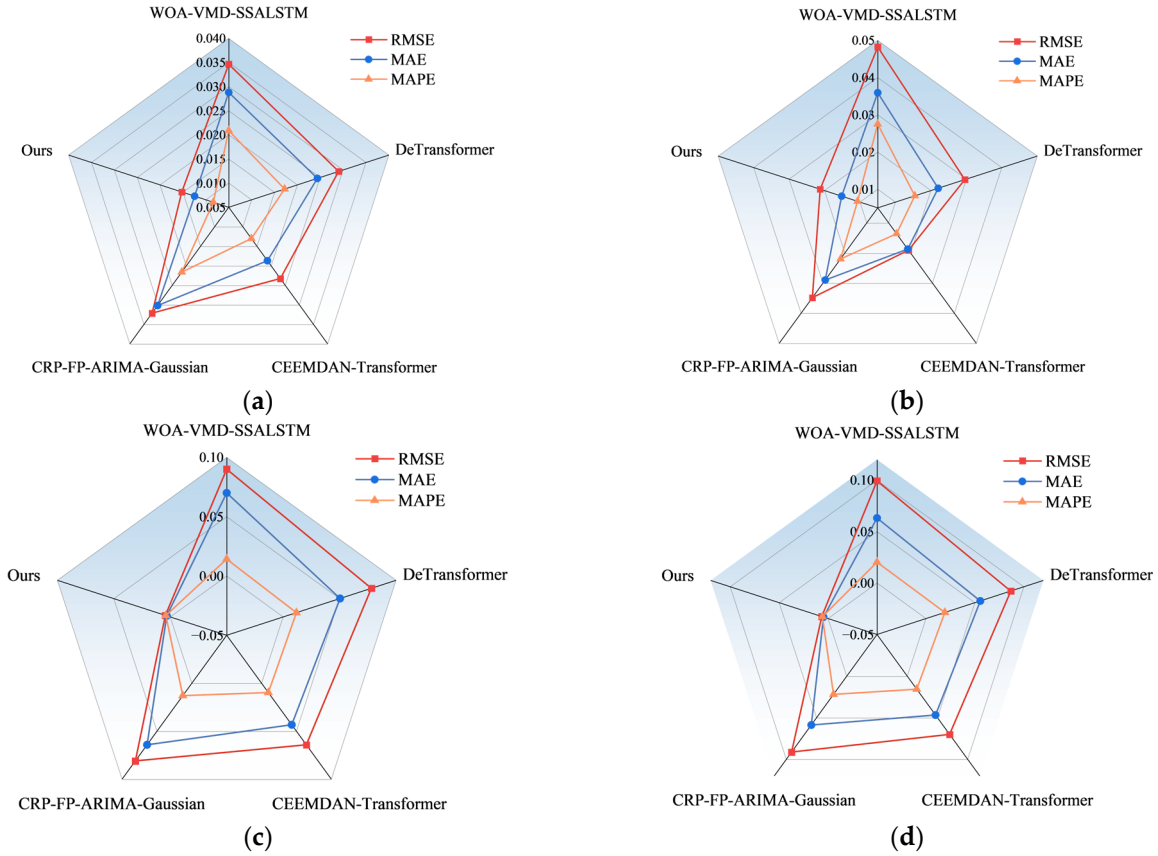


Figure 12. Comparison of RIME-VMD-SEInformer and other models on b0005 (**a**), b0006 (**b**), cs2_35 (**c**), and cs2_36 (**d**).

Across all datasets, the R^2 values of the RIME-VMD-SEInformer model consistently approach 1, with RMSE and MAPE metrics significantly lower than those of the comparative algorithms. For example, on the B0005 dataset, the R^2 reached 0.9851, and the MAPE was only 0.00837. In comparison, the closest competitor, CEEMDAN-Transformer, achieved an MAPE of 0.01302, indicating an improvement of over 35%.

Furthermore, the proposed model demonstrates superior data fitting capabilities compared to other models, with substantial reductions in both MAPE and RMSE. On the B0005 dataset, the MAPE of the proposed method was only 0.00837, significantly lower than that of the CEEMDAN-Transformer (0.01302) and De-Transformer (0.01721), and 59.96% lower than that of WOA-VMD-SSALSTM (0.02090). Additionally, the RMSE of the proposed model was 0.01526, representing a 52.47% reduction compared to CRP-FP-ARIMA-Gaussian (0.03212).

Compared to methods like CRP-FP-ARIMA-Gaussian, the RIME-VMD-SEInformer model achieves comprehensive improvements across multiple error metrics, particularly excelling in modeling long time spans and complex dependencies. These results highlight its significant advantages in terms of predictive accuracy and robustness.

5. Conclusions

This study proposes an innovative Remaining Useful Life (RUL) prediction model for lithium-ion batteries—RIME-VMD-SEInformer. By synergizing the signal decomposition, feature extraction, and time-series modeling modules, the model significantly enhances the prediction accuracy of the battery capacity degradation trends. The main contributions and conclusions of this work are summarized as follows:

1. Integration of RIME-VMD-SEInformer Components

The RIME-VMD-SEInformer model integrates RIME-optimized Variational Mode Decomposition (VMD), the SEAttention channel attention mechanism, and Informer's sparse attention mechanism. The RIME-VMD module adaptively extracts multi-frequency features by optimizing the decomposition parameters, effectively mitigating the impact of capacity regeneration phenomena and noise. The SEAttention module dynamically adjusts the channel weights to emphasize key multimodal features, thereby improving the model's ability to capture complex characteristics. The Informer module uses sparse attention and self-attention distillation to model long-term sequence dependencies while reducing computational complexity.

2. Effect of Training Data Size on Model Performance

This study analyzed the impact of the training set size on model performance. The results indicate that as the training sample proportion increases from 5% to 30%, error metrics, such as RMSE and MAPE, decrease significantly. When the training set proportion reaches 30%, the model achieves the best prediction accuracy across all datasets. This suggests that a 3:7 ratio for the training and testing sets can be adopted in future applications.

3. Ablation Study on Model Components

Ablation experiments verified the contribution of each module to the overall model performance. For instance, on the B0005 dataset, removing the RIME-VMD module decreased R^2 from 0.9851 to 0.9656 and increased MAPE from 0.00837 to 0.00968, demonstrating the module's effectiveness in signal decomposition and noise reduction for high-quality input features. Removing the SEAttention module increased the MAPE on the B0005 dataset from 0.00837 to 0.01316, highlighting its critical role in strengthening key multimodal features. Replacing the Informer with a standard Transformer caused R^2 on the CS2_36 dataset to drop from 0.9958 to 0.7835, with MAPE rising from 0.00586 to 0.04243, underscoring Informer's core role in long-term dependency modeling.

4. Comparison with State-of-the-Art (SOA) Algorithms

Experiments across multiple publicly available datasets demonstrate that the RIME-VMD-SEInformer model outperforms other SOA algorithms (e.g., CEEMDAN-Transformer, WOA-VMD-SSALSTM) in metrics such as R^2, RMSE, MAE, and MAPE. For example, on the B0005 dataset, the model achieved a MAPE of 0.00837, a 59.96% reduction compared to WOA-VMD-SSALSTM, and a 35.69% reduction compared to CEEMDAN-Transformer. On the CS2_36 dataset, R^2 reached 0.9958, significantly higher than De-Transformer's 0.8612 and WOA-VMD-SSALSTM's 0.8249, showcasing its superior performance in handling high-complexity data.

5. Limitations and Future Prospects

This study presents the RIME-VMD-SEInformer model, which significantly improves the prediction accuracy of the Remaining Useful Life (RUL) of lithium-ion batteries. The experimental results demonstrate that the proposed model outperforms state-of-the-art (SOA) methods, excelling in long-term sequence modeling, data robustness, and signal denoising. However, challenges remain, including high computational complexity, lack of integration with physical degradation models, limited transfer learning capabilities, and absence of real-world industrial validation. Future research will focus on optimizing the computational efficiency (e.g., knowledge distillation and pruning), incorporating physics-informed modeling (to enhance generalization), improving small-sample adaptability (transfer learning and self-supervised learning), and conducting real-world industrial testing (e.g., in new energy ships and electric transportation). These advancements will further enhance the practical applicability of the model, providing promising solutions for Smart Battery Management Systems (Smart BMS) and Predictive Maintenance.

Author Contributions: Conceptualization, X.Y. and Y.L.; methodology, H.J. and Y.L.; software, Y.L.; validation, Y.L., H.J. and T.T.; formal analysis, Y.L., H.J. and J.L. (Jiaxin Luo); investigation, L.H.; resources, H.J. and L.H.; data curation, W.Z.; writing—original draft preparation, X.Y. and J.L. (Junting Lang); writing—review and editing, H.J. and Y.L.; visualization, Y.L. and H.J.; supervision, X.Y.; project administration, X.Y.; funding acquisition, X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the China National Key Research and Development Project (Grant No: 2023YFB4301704) and the China National Key Laboratory of Electromagnetic Energy Technology Open Fund (Grant No: 61422172220403).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be made available upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, Z.; Li, M.; Zhao, F.; Ji, Y.; Han, F. Status and prospects in technical standards of hydrogen-powered ships for advancing maritime zero-carbon transformation. *Int. J. Hydrogen Energy* **2024**, *62*, 925–946. [CrossRef]
2. van Biert, L.; Godjevac, M.; Visser, K.; Aravind, P.V. A review of fuel cell systems for maritime applications. *J. Power Sources* **2016**, *327*, 345–364. [CrossRef]
3. Wang, Z.; Dong, B.; Wang, Y.; Li, M.; Liu, H.; Han, F. Analysis and evaluation of fuel cell technologies for sustainable ship power: Energy efficiency and environmental impact. *Energy Convers. Manag. X* **2024**, *21*, 100482. [CrossRef]
4. Milićević, S.; Blagojević, I.; Milojević, S.; Bukvić, M.; Stojanović, B. Numerical Analysis of Optimal Hybridization in Parallel Hybrid Electric Powertrains for Tracked Vehicles. *Energies* **2024**, *17*, 3531. [CrossRef]
5. Lü, X.; Qu, Y.; Wang, Y.; Qin, C.; Liu, G. A comprehensive review on hybrid power system for PEMFC-HEV: Issues and strategies. *Energy Convers. Manag.* **2018**, *171*, 1273–1291. [CrossRef]
6. Lu, J.; Abed, A.M.; Nag, K.; Fayed, M.; Deifalla, A.; Al-Zahrani, A.; Ghamry, N.A.; Galal, A.M. Optimization of a near-zero-emission energy system for the production of desalinated water and cooling using waste energy of fuel cells. *Chemosphere* **2023**, *336*, 139035. [CrossRef] [PubMed]
7. Gürbüz, H.; Akçay, H.; Demirtürk, S.; Topalcı, Ü. Techno-enviro-economic comparison analysis of a PEMFC and a hydrogen-fueled SI engine. *Appl. Therm. Eng.* **2024**, *243*, 122528. [CrossRef]
8. Balcı, Ö.; Karagöz, Y.; Kale, S.; Damar, S.; Attar, A.; Köten, H.; Dalkılıç, A.S.; Wongwises, S. Fuel consumption and emission comparison of conventional and hydrogen feed vehicles. *Int. J. Hydrogen Energy* **2021**, *46*, 16250–16266. [CrossRef]
9. Naseri, F.; Barbu, C.; Sarikurt, T. Optimal sizing of hybrid high-energy/high-power battery energy storage systems to improve battery cycle life and charging power in electric vehicle applications. *J. Energy Storage* **2022**, *55*, 105768. [CrossRef]
10. Liu, J.; Chen, X.; Cao, S.; Yang, H. Overview on hybrid solar photovoltaic-electrical energy storage technologies for power supply to buildings. *Energy Convers. Manag.* **2019**, *187*, 103–121. [CrossRef]
11. Vanem, E.; Salucci, C.B.; Bakdi, A.; Alnes, Ø.Å.S. Data-driven state of health modelling—A review of state of the art and reflections on applications for maritime battery systems. *J. Energy Storage* **2021**, *43*, 103158. [CrossRef]
12. Lin, C.P.; Cabrera, J.; Yu, D.Y.W.; Yang, F.; Tsui, K.L. SOH Estimation and SOC Recalibration of Lithium-Ion Battery with Incremental Capacity Analysis & Cubic Smoothing Spline. *J. Electrochem. Soc.* **2020**, *167*, 090537. [CrossRef]
13. Wang, J.; Liu, P.; Hicks-Garner, J.; Sherman, E.; Soukiazian, S.; Verbrugge, M.; Tataria, H.; Musser, J.; Finamore, P. Cycle-life model for graphite-$LiFePO_4$ cells. *J. Power Sources* **2011**, *196*, 3942–3948. [CrossRef]
14. Wang, D.; Miao, Q.; Pecht, M. Prognostics of lithium-ion batteries based on relevance vectors and a conditional three-parameter capacity degradation model. *J. Power Sources* **2013**, *239*, 253–264. [CrossRef]
15. Li, Z.; Lu, L.; Ouyang, M.; Xiao, Y. Modeling the capacity degradation of $LiFePO_4$/graphite batteries based on stress coupling analysis. *J. Power Sources* **2011**, *196*, 9757–9766. [CrossRef]
16. Xu, B.; Oudalov, A.; Ulbig, A.; Andersson, G.; Kirschen, D.S. Modeling of Lithium-Ion Battery Degradation for Cell Life Assessment. *IEEE Trans. Smart Grid* **2018**, *9*, 1131–1140. [CrossRef]
17. Xiong, R.; Tian, J.; Shen, W.; Lu, J.; Sun, F. Semi-supervised estimation of capacity degradation for lithium ion batteries with electrochemical impedance spectroscopy. *J. Energy Chem.* **2023**, *76*, 404–413. [CrossRef]
18. Kong, D.; Wang, S.; Ping, P. State-of-health estimation and remaining useful life for lithium-ion battery based on deep learning with Bayesian hyperparameter optimization. *Int. J. Energy Res.* **2022**, *46*, 6081–6098. [CrossRef]

19. Ren, L.; Dong, J.; Wang, X.; Meng, Z.; Zhao, L.; Deen, M.J. A Data-Driven Auto-CNN-LSTM Prediction Model for Lithium-Ion Battery Remaining Useful Life. *IEEE Trans. Ind. Inform.* **2021**, *17*, 3478–3487. [CrossRef]
20. Zhang, X.; Qin, Y.; Yuen, C.; Jayasinghe, L.; Liu, X. Time-Series Regeneration With Convolutional Recurrent Generative Adversarial Network for Remaining Useful Life Estimation. *IEEE Trans. Ind. Inform.* **2021**, *17*, 6820–6831. [CrossRef]
21. Yang, H.; Wang, P.; An, Y.; Shi, C.; Sun, X.; Wang, K.; Zhang, X.; Wei, T.; Ma, Y. Remaining useful life prediction based on denoising technique and deep neural network for lithium-ion capacitors. *eTransportation* **2020**, *5*, 100078. [CrossRef]
22. Hong, J.; Lee, D.; Jeong, E.-R.; Yi, Y. Towards the swift prediction of the remaining useful life of lithium-ion batteries with end-to-end deep learning. *Appl. Energy* **2020**, *278*, 115646. [CrossRef]
23. Zhou, D.; Li, Z.; Zhu, J.; Zhang, H.; Hou, L. State of Health Monitoring and Remaining Useful Life Prediction of Lithium-Ion Batteries Based on Temporal Convolutional Network. *IEEE Access* **2020**, *8*, 53307–53320. [CrossRef]
24. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 9. [CrossRef]
25. Duan, J.; Zuo, H.; Bai, Y.; Duan, J.; Chang, M.; Chen, B. Short-term wind speed forecasting using recurrent neural networks with error correction. *Energy* **2021**, *217*, 119397. [CrossRef]
26. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [CrossRef] [PubMed]
27. Tian, Z.; Chen, H. Multi-step short-term wind speed prediction based on integrated multi-model fusion. *Appl. Energy* **2021**, *298*, 117248. [CrossRef]
28. Su, H.; Zhao, D.; Heidari, A.A.; Liu, L.; Zhang, X.; Mafarja, M.; Chen, H. RIME: A physics-based optimization. *Neurocomputing* **2023**, *532*, 183–214. [CrossRef]
29. Liu, W.; Bai, Y.; Yue, X.; Wang, R.; Song, Q. A wind speed forcasting model based on rime optimization based VMD and multi-headed self-attention-LSTM. *Energy* **2024**, *294*, 130726. [CrossRef]
30. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. 2018. pp. 7132–7141. Available online: https://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.html (accessed on 16 January 2025).
31. Gajević, S.; Marković, A.; Milojević, S.; Ašonja, A.; Ivanović, L.; Stojanović, B. Multi-Objective Optimization of Tribological Characteristics for Aluminum Composite Using Taguchi Grey and TOPSIS Approaches. *Lubricants* **2024**, *12*, 171. [CrossRef]
32. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is All you Need. *arXiv* **2017**, arXiv:1706.03762.
33. Wang, Z.; Liu, Y.; Wang, F.; Wang, H.; Su, M. Capacity and remaining useful life prediction for lithium-ion batteries based on sequence decomposition and a deep-learning network. *J. Energy Storage* **2023**, *72*, 108085. [CrossRef]
34. Bao, Q.; Qin, W.; Yun, Z. A Multi-Stage Adaptive Method for Remaining Useful Life Prediction of Lithium-Ion Batteries Based on Swarm Intelligence Optimization. *Batteries* **2023**, *9*, 224. [CrossRef]
35. Transformer Network for Remaining Useful Life Prediction of Lithium-Ion Batteries | IEEE Journals & Magazine | IEEE Xplore, (n.d.). Available online: https://ieeexplore.ieee.org/abstract/document/9714323 (accessed on 16 January 2025).
36. Zhao, B.; Zhang, W.; Zhang, Y.; Zhang, C.; Zhang, C.; Zhang, J. Research on the remaining useful life prediction method for lithium-ion batteries by fusion of feature engineering and deep learning. *Appl. Energy* **2024**, *358*, 122325. [CrossRef]
37. He, N.; Yang, Z.; Qian, C.; Li, R.; Gao, F.; Cheng, F. Remaining useful life prediction of lithium-ion battery based on fusion model considering capacity regeneration phenomenon. *J. Energy Storage* **2024**, *85*, 111068. [CrossRef]

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.