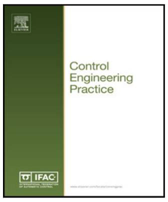
Control Engineering Practice 159 (2025) 106302

Contents lists available at ScienceDirect

Control Engineering Practice

journal homepage: www.elsevier.com/locate/conengprac



A parallel weighted ADTC-Transformer framework with FUnet fusion and KAN for improved lithium-ion battery SOH prediction

Chuang Chen [a,d], Yuheng Wu [a], Jiantao Shi [a,*], Dongdong Yue [a], Ge Shi [b], Dongzhen Lyu [c]

[a] College of Electrical Engineering and Control Science, Nanjing Tech University, Nanjing 211816, China
[b] School of Geomatics Science and Technology, Nanjing Tech University, Nanjing 211816, China
[c] College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou, 325035, China
[d] Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China

ARTICLE INFO

Keywords:
Lithium-ion batteries
State of health prediction
Transformer
Multi-scale feature extraction
Feature fusion

ABSTRACT

This paper delves into the extraction and integration of local and global features for lithium-ion battery State of Health (SOH) prediction, proposing an innovative parallel weighted architecture—ADTC-Transformer. This framework combines Adaptive Dilated Temporal Convolution (ADTC) with a Transformer encoder to effectively capture and balance local and global dependencies while dynamically optimizing feature contributions through a weighted fusion mechanism. Additionally, the traditional U-shaped network (Unet) is enhanced by incorporating a Feature Pyramid Network (FPN), forming the FUnet module, which significantly strengthens the fusion and utilization of multi-scale features. Building on this, the Kolmogorov–Arnold Network (KAN) is introduced as the final prediction module, leveraging Kolmogorov–Arnold representation theory to model complex high-dimensional features through local interpolation and global nonlinear transformations. This enables the KAN module to capture intricate temporal dependencies and interactions across a wide range of feature scales, thus improving the model's ability to predict long-term SOH. Experimental results demonstrate that the proposed method markedly improves prediction accuracy across NASA, CALCE, and WRBD datasets, excelling particularly in long-term SOH prediction for lithium-ion batteries. This provides robust support for battery health management and performance optimization.

1. Introduction

Lithium-ion batteries are widely used in portable devices, electric vehicles, and renewable energy storage systems due to their exceptional energy density, long lifespan, and low self-discharge rate (Zhang, Liu, Li, et al., 2023). However, as the number of charge–discharge cycles increases, the performance of these batteries gradually degrades. In recent years, the rise of deep learning has brought significant breakthroughs in handling time-series data and similar tasks (Chen, Lu, et al., 2021; Chen, Shi, Shen, Feng, & Tao, 2023; Han, Qi, Wang, et al., 2024; Wang, Jiang, Mu, et al., 2025; Zhao, Jing, Wu, Li, & Todoh, 2024; Zhao, Liu, et al., 2025), providing numerous successful examples in data-driven methodologies. Given that lithium-ion battery SOH prediction is inherently related to time-series data, these advancements in deep learning offer valuable insights and inspiration for improving prediction accuracy. Traditional approaches for SOH prediction can be categorized into three main types: physics-based models, semi-empirical models, and data-driven models (Lipu, Hannan, Hussain, et al., 2023).

Physics-based models utilize complex partial differential equations to describe the electrochemical mechanisms within batteries (Chen, Chai, Dogru, et al., 2021; Lyu, Liu, Chen, et al., 2025; Lyu, Niu, Zhang, et al., 2020; Vignesh, Che, Selvaraj, et al., 2024; Wen, Ye, Li, et al., 2024), enabling accurate representation of aging dynamics. However, their computational complexity and the challenges of parameterization limit their applicability in real-time scenarios. In contrast, semi-empirical models provide a simplified framework by building models based on experimental data (Cai, Holdstock, Morganti, et al., 2024; Singh, Chen, Tan, et al., 2019), though their reliance on data from specific scenarios hinders broad generalization. Data-driven approaches, on the other hand, typically estimate battery health by leveraging historical data and aging trends (Lin, Tang, & Wang, 2015). With the advancements in artificial intelligence, data-driven models have emerged as the dominant research direction in recent years (Chen, Chai, Dogru, et al., 2021; Chen, Jiang, & Ding, 2020; Chen, Tao, Shi, Shen, & Zhu, 2024; Demirci, Taskin, Schaltz, et al., 2024). These models analyze historical operational data, demonstrating high predictive accuracy and scalability.

[*] Corresponding author.
E-mail address: sjt11@tsinghua.org.cn (J. Shi).

https://doi.org/10.1016/j.conengprac.2025.106302
Received 25 November 2024; Received in revised form 19 February 2025; Accepted 20 February 2025
Available online 5 March 2025

In data-driven models, various benchmark models are widely utilized (Lyu, Liu, et al., 2025; Lyu, Zhang, Liu, Yang, & Xiang, 2025). Recurrent Neural Networks (RNNs) are among the earliest tools explored for sequence modeling due to their memory-state mechanism, which effectively captures temporal features. For instance, Lu, Xiong, Tian, et al. (2022) proposed an RNN-based framework for SOH and Remaining Useful Life (RUL) prediction, using upcoming current schedules and a limited initial capacity-voltage dataset as inputs to model battery degradation processes. However, RNNs are prone to issues like gradient explosion and vanishing gradients, which limit their prediction accuracy. To address these shortcomings, improved variants such as Gated Recurrent Unit (GRU) (Zhang, Wang, Yuan, & Liang, 2022) and Long Short-Term Memory (LSTM) (Ren, Dong, Wang, et al., 2020) have been developed. Ungurean, Micea, and Cârstoiu (2020) designed a GRU-based online SOH prediction method and demonstrated, through comparative experiments with LSTM, that both exhibit similar prediction capabilities. Furthermore, Ma, Shan, Gao, and Chen (2022) proposed a model combining an enhanced LSTM with Health Indicators (HIs), where hyperparameters were optimized using the Differential Evolution Grey Wolf Optimizer (DEGWO), resulting in significantly improved accuracy and robustness. Additionally, Bidirectional LSTMs (BiLSTMs) have gained popularity for their ability to capture bidirectional temporal dependencies. For example, Li, Luo, Zhang, and Liu (2023) integrated Incremental Energy Analysis (IEA) with BiLSTMs to evaluate the mapping relationship between battery degradation and peak characteristics, constructing a high-precision SOH prediction model. Despite their effectiveness in capturing temporal dependencies, RNNs and their variants often exhibit high computational complexity, which becomes a limiting factor, especially when handling long-term time series data.

Additionally, traditional Neural Network (NN) models are not well-suited for handling spatial patterns in data. As a result, Convolutional Neural Networks (CNNs) have emerged as effective solutions for capturing spatial patterns and short-term dependencies in time series data. For instance, Valant, Wheaton, Thurston, et al. (2019) utilized 1D CNNs to detect degradation patterns, while Durmus and Karagol (2024) optimized CNN models using genetic algorithms, achieving significant improvements in battery capacity prediction accuracy. CNNs excel in identifying spatial patterns and short-term temporal dependencies; however, their fixed receptive field limits their ability to model long-term dependencies. This constraint reduces their effectiveness in scenarios that require capturing global temporal dependencies. Building upon CNNs, Temporal Convolutional Networks (TCNs) were introduced to overcome the limitations of fixed receptive fields. TCNs enhance traditional CNNs for temporal modeling by incorporating causal convolutions to ensure that each output at a given time step depends only on the current and preceding time points, thus avoiding future information leakage. Furthermore, they leverage dilated convolutions with exponentially increasing receptive fields to capture long-term dependencies while reducing computational complexity. These improvements enable TCNs to efficiently model long time-series data while maintaining parallel computation efficiency. For instance, Chen, Liu, et al. (2022) proposed the AdTCN-BO model, which integrates Bayesian optimization to achieve high accuracy in SOH prediction. Similarly, Liu, Li, Zhang, et al. (2021) combined TCNs with transfer learning, effectively improving the model's applicability across various battery types. Despite their strong performance in time-series tasks, particularly in capturing long-term dependencies and ensuring computational efficiency compared to RNNs and CNNs, TCNs also exhibit limitations. Their capacity to handle complex multimodal data and multi-phase degradation features remains suboptimal. Specifically, the dilated convolution mechanism struggles to effectively address global dependencies among features in high-dimensional data. Additionally, TCNs primarily rely on layer-by-layer propagation during feature fusion, lacking the ability to dynamically model global features, which can hinder predictive performance in highly complex time-series tasks.

To further enhance temporal modeling capabilities, Transformers have gained widespread adoption in battery SOH prediction due to their unique advantage in capturing global dependencies through self-attention mechanisms. For instance, Shen, Zhou, Wang, et al. (2022) leveraged Transformers to effectively capture complex global temporal dependencies, significantly improving the stability of SOC predictions. Han, Li, Zheng, et al. (2023) designed a Denoising Transformer Neural Network (DTNN) that combines residual learning with multi-head attention mechanisms, demonstrating outstanding performance on complex degradation datasets. Similarly, Zhao and Wang (2024) proposed an enhanced Transformer architecture tailored for multi-condition battery SOH prediction. By integrating self-attention mechanisms with graph representation techniques, their model achieved improved adaptability and robustness for handling complex battery data. However, these studies also highlight certain limitations. The reliance on extensive training datasets can hinder Transformer performance in data-scarce scenarios. Additionally, the strong emphasis on global features may lead to the neglect of locally significant features. Furthermore, the high computational complexity and substantial data requirements make Transformers less suitable for resource-constrained environments. Lastly, the modeling of input features in Transformers often depends on relatively linear embedding processes, which struggle to fully exploit the multi-scale features present in multimodal data.

To further optimize global feature modeling and multi-scale feature fusion, U-Net and its variants, with their unique encoder–decoder structure and skip connection capability, provide valuable approaches for addressing complex feature extraction challenges. For instance, Fan, Yang, and Hou (2024) proposed the Mixed Attention U-Net (MMAU-Net), which utilizes a multi-stage degradation decomposition strategy to effectively extract battery aging features, significantly enhancing the accuracy of SOH prediction. Similarly, Song, Hu, Shu, et al. (2024) introduced the Dynamically Re-weighted U-Net (EADRU-Net), incorporating edge-emphasis loss and context-aware attention mechanisms, demonstrating strong capabilities in capturing fine-grained features. However, despite these advancements in specific U-Net capabilities, their ability to model global temporal dependencies remains limited. Moreover, the relatively high computational complexity of U-Net variants reduces their efficiency in large-scale applications. Additionally, U-Net lacks flexibility when handling multimodal data and high-dimensional complex features, making it challenging to balance global feature representation with local detail extraction, a critical requirement for modern lithium-ion battery SOH prediction tasks.

Meanwhile, hybrid models have become a focal point of research, building on the strengths of benchmark models. By combining the advantages of various models — such as RNNs for temporal sequence modeling, CNNs for local feature extraction, and Transformers for capturing global dependencies — hybrid models achieve enhanced performance. For example, Ren et al. (2020) developed the Auto-CNN-LSTM model, which integrates autoencoders, CNNs, and LSTMs to effectively fuse deep features with sequence modeling capabilities. Similarly, Bao, Chen, Lopes, et al. (2023) proposed a hybrid model that combines CNNs, an improved LSTM variant (VLSTM), and dimensional attention mechanisms, achieving high-precision SOH prediction for lithium-ion batteries by optimizing temporal and key feature weights. Additionally, Wang, Amogne, Chou, and Tseng (2022) introduced a hybrid model that integrates BiLSTM and an attention mechanism (AM). Experimental results on six lithium-ion battery datasets demonstrated significant improvements in both prediction accuracy and robustness. Compared to single models, hybrid models exhibit greater adaptability and robustness in complex scenarios. However, they also face challenges such as increased architectural complexity and higher computational costs, which must be addressed for broader application.

In response to these challenges, this paper proposes an innovative hybrid model that fully integrates the advantages of multiple existing approaches while introducing several key improvements. The model employs an Adaptive Dilated Temporal Convolution (ADTC) module to efficiently capture long-sequence dependencies, balancing local feature extraction and global feature recognition capabilities. A parallel weighted architecture is introduced to optimize the representation of both local and global features by balancing the contributions of ADTC and Transformer outputs. Additionally, the model incorporates an enhanced U-Net module (FUnet) that leverages a Feature Pyramid Network (FPN) to improve the efficiency of multi-scale feature fusion while reducing computational complexity. Finally, a Kolmogorov–Arnold Network (KAN) is utilized to model the complex relationships among high-dimensional features and generate predictions, significantly enhancing the model's accuracy and robustness. These advancements collectively enable the proposed model to achieve remarkable improvements in prediction performance and reliability.

The main contributions of this paper are as follows:

1. A novel ADTC module is proposed to effectively capture long-term dependencies and local features, enhancing adaptability to multi-phase degradation data. Together with the Transformer encoder, it forms a parallel weighted fusion mechanism that balances local and global features, improving representation precision and consistency.
2. A novel FUnet module is designed to enhance feature extraction by integrating an FPN and global average pooling into the U-Net structure. It enables hierarchical multi-scale feature fusion while simplifying the network, reducing computational costs, and ensuring applicability to high-dimensional data and industrial scenarios.
3. The favorable KAN is introduced for final prediction. Through the use of learnable nonlinear weighting functions, it adeptly captures the complexity of fused high-dimensional features and ensures efficient modeling, resulting in significant improvements in prediction accuracy and robustness.

Next, Section 2 presents the proposed hybrid model architecture, detailing the fundamental principles and specific training processes of the ADTC, Transformer, FUnet, and KAN modules. Section 3 validates the model through experiments conducted on the NASA, CALCE, and WRBD datasets (the latter collected in the authors' laboratory), providing a comprehensive analysis of the results. Finally, Section 4 concludes the paper and proposes directions for future research.

2. Methodology

2.1. Overall framework

As shown in Fig. 1, the overall framework is divided into four stages: data preprocessing, feature extraction, model training and evaluation. During the preprocessing stage, data normalization and the division of training and testing datasets are completed. In the feature extraction stage, the ADTC and Transformer modules run in parallel. During the model training phase, the FUnet module performs multi-scale fusion of the weighted high-dimensional features, while the KAN module applies nonlinear modeling to optimize the final output. Finally, in the testing phase, both single-step and multi-step prediction tasks are conducted. The modules within the framework collaborate efficiently, forming a complete closed-loop process from raw data input to SOH prediction. The following sections provide a detailed explanation of the design and implementation of each module.

2.2. ADTC module

TCNs have demonstrated exceptional performance in time-series modeling due to their computational efficiency. Their core mechanisms, causal convolution and dilated convolution, enable the capture of temporal dependency features effectively. However, as application scenarios grow increasingly complex, traditional TCNs reveal certain limitations. Their convolution operations often rely on fixed strategies, such as exponentially increasing dilation factors (e.g., 1, 2, 4, 8) and predefined kernel sizes (e.g., 2, 3). This rigidity reduces the model's flexibility when dealing with diverse temporal features. Specifically, short-term dependencies may be overlooked, while an excessively large receptive field can dilute the focus on long-term dependencies. Furthermore, TCNs are unable to dynamically adjust their structure based on the characteristics of the input sequence, significantly limiting their generalization capability for varying time-series data.

To address these limitations, this paper proposes an ADTC module, as illustrated in Fig. 2. This module integrates a multi-branch structure, an adaptive dilation strategy, and nonlinear activation mechanisms, enabling it to extract local features with high precision while flexibly modeling global dependencies. These enhancements significantly improve the module's feature representation capabilities.

The core design of the ADTC module is a multi-branch convolution based on a dilated Inception structure. Within this structure, different branches use convolution kernels of varying sizes (e.g., 2, 4, 8, and three-fourths of the input temporal dimension, rounded to the nearest integer) combined with dynamically adjusted dilation factors. Specifically, temporal signals often exhibit inherent periods such as 7, 12, 24, 28, and 60 (Wu, Pan, Long, et al., 2020). By leveraging kernel sizes derived from dilation factors (e.g., 2, 4, 8, and three-fourths of the input temporal dimension), the resulting kernel sizes correspond to 6, 9, 18, and 21. This combination enables a nonlinear, multi-scale receptive field, significantly enhancing the model's ability to extract local features. Smaller kernel branches capture detailed features between local time steps, enhancing the representation of short-term dependencies, while larger kernel branches, aided by dynamic dilation factors, expand the receptive field to effectively capture long-term dependencies, thereby modeling global trends. The dynamic adjustment of dilation factors further enables the module to flexibly adapt to different temporal ranges of feature modeling based on the length of the input sequence.

The multi-branch design may result in temporal inconsistencies due to differences in convolution operations across branches. To address this, the ADTC module employs a dynamic temporal alignment strategy, which ensures consistency among all branch features by truncating them to the shortest temporal length. The aligned features are then concatenated along the channel dimension and further refined through a two-layer fully connected network, enhancing the expressiveness of the features. This ensures the effective fusion of multi-branch features for improved overall performance.

To enhance feature selection and representation capabilities, the ADTC module incorporates a nonlinear activation mechanism at its end. This mechanism is composed of a dual-dilated convolution structure: one dilated convolution layer is followed by a hyperbolic tangent activation function (tanh), acting as a filter to extract key features; the other dilated convolution layer is followed by a sigmoid activation function, serving as a gating mechanism to dynamically control the amount of information passed through the filter. The final module output is generated by element-wise multiplication of the filter and the gating mechanism, enabling dynamic feature selection and efficient information flow. This activation mechanism combines the strengths of both tanh and sigmoid, demonstrating exceptional flexibility and robustness in modeling complex temporal dependencies.

2.3. Transformer module

As a structure centered on the self-attention mechanism, the Transformer has been widely applied to sequence modeling tasks in recent years. Its core advantage lies in its ability to capture global dependencies in time-series data while significantly improving computational efficiency through parallel processing. this paper adopts the encoder portion of the Transformer, stacking multiple encoder layers to achieve

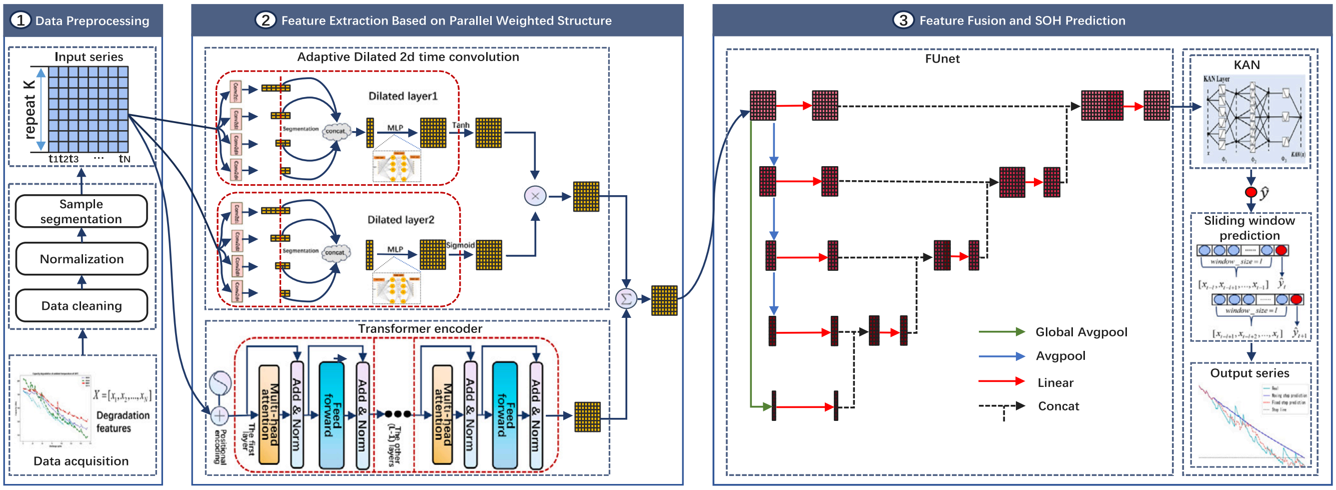


Fig. 1. Proposed parallel weighted ADTC-Transformer framework with FUnet fusion and KAN.

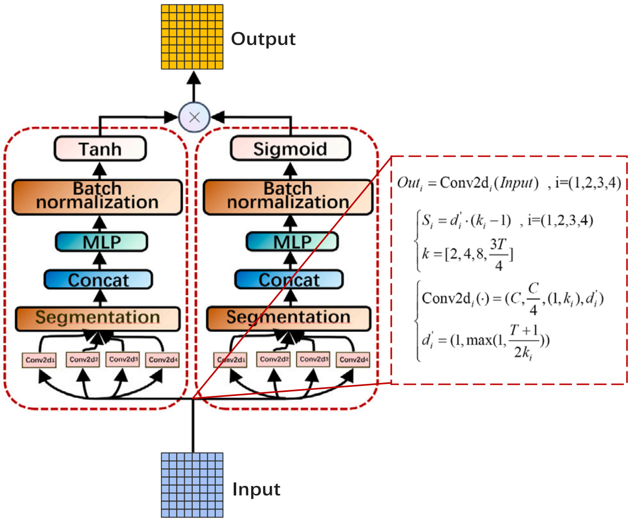

Output
Tanh
Batch normalization
MLP
Concat
Segmentation
Conv2d1
Conv2d2
Conv2d3
Conv2d4
Sigmoid
Batch normalization
MLP
Concat
Segmentation
Conv2d1
Conv2d2
Conv2d3
Conv2d4
$Out_i = \mathrm{Conv2d}_i(Input)$, i=(1,2,3,4)
$S_i = d_i' \cdot (k_i - 1)$, i=(1,2,3,4)
$k = [2,4,8,\frac{3T}{4}]$
$\mathrm{Conv2d}_i(\cdot) = (C, \frac{C}{4}, (1,k_i), d_i')$
$d_i' = (1, \max(1, \frac{T+1}{2k_i}))$
Input

Fig. 2. Structure of the ADTC module.

efficient extraction of global-temporal dependency features, as illustrated in Fig. 3. The encoder consists of three main components: the multi-head attention module, the Add & Norm module, and the feedforward network module. The design logic and functionalities of these components are as follows:

The multi-head attention module effectively uncovers global relationships among input sequence features by combining positional encoding with the self-attention mechanism. First, the positional information of the sequence is encoded using sinusoidal and cosine functions:

$$\begin{cases} PE(pos, 2u) = \sin(pos/10000^{2u/h}), \\ PE(pos, 2u+1) = \cos(pos/10000^{2u/h}), \end{cases} \tag{1}$$

where pos represents the position of an object within the input sequence, u denotes its position within the corresponding vector, and h indicates the dimensionality of the embedding space in the Transformer model's output.

The positional information is added to the original input $x^{(i)}$, forming the initial input for the multi-head attention mechanism of the Transformer model:

$$H^0 = x^{(i)} + PE^{(i)}, \tag{2}$$

where $x^{(i)}$ represents the ith feature sample as input, and $PE^{(i)}$ denotes the positional encoding corresponding to $x^{(i)}$.

Next, the input features are processed through the multi-head attention mechanism to calculate the correlations between sub-features, resulting in the output for each attention head:

$$\begin{cases} head_i = Attention\left(H^{l-1}W_Q^{l,i}, H^{l-1}W_K^{l,i},\right. \\ \qquad\qquad \left. H^{l-1}W_V^{l,i}\right), \quad (1 \le l \le L), \\ Attention\left(H^{l-1}W_Q^{l,i}, H^{l-1}W_K^{l,i}, H^{l-1}W_V^{l,i}\right) \\ \quad = softmax\left(\dfrac{H^{l-1}W_Q^{l,i} \bullet (H^{l-1}W_K^{l,i})^T}{\sqrt{d_h}}\right) H^{l-1}W_V^{l,i}, \end{cases} \tag{3}$$

where $Attention(\cdot)$ represents the mapping function of the attention head, H^{l-1} is the output of the $(l-1)$-th encoder layer, $W_Q^{l,i}$, $W_K^{l,i}$, and $W_V^{l,i}$ are the three randomly-initialized weight matrices corresponding to the ith attention head in the lth layer of the Transformer model; L denotes the number of stacked encoder layers in the Transformer model.

Then, the outputs of all attention heads are concatenated and linearly transformed to form the output of the multi-head attention module:

$$MultiHead(H^{l-1}) = [head_1; head_2; \cdots; head_n]W^O, \tag{4}$$

where MultiHead$(\cdot)$ represents the mapping function of the multi-head attention mechanism, and W^O denotes the weight matrix of the multi-head attention module.

The output of the multi-head attention module is optimized through the Add & Norm layer to ensure gradient stability and preserve the input features. Specifically, the process involves adding the output of the multi-head attention module to the original input, followed by layer normalization:

$$H'^{l} = \mathrm{LayerNorm}(\mathrm{MultiHead}(H^{l-1}) + H^{l-1}), \tag{5}$$

where MultiHead(H^{l-1}) is the output of the multi-head attention module, H^{l-1} represents the original input, which is the output of the previous encoder layer, and MultiHead(H^{l-1}) denotes the layer normalization operation.

In the feedforward network section, the encoder employs a two-layer linear transformation network to enhance nonlinear modeling capabilities. The output of the feedforward network can be expressed as:

$$\mathrm{FFN}(H'^{l}) = \mathrm{ReLU}(H'^{l}W_1 + b_1)W_2 + b_2, \tag{6}$$

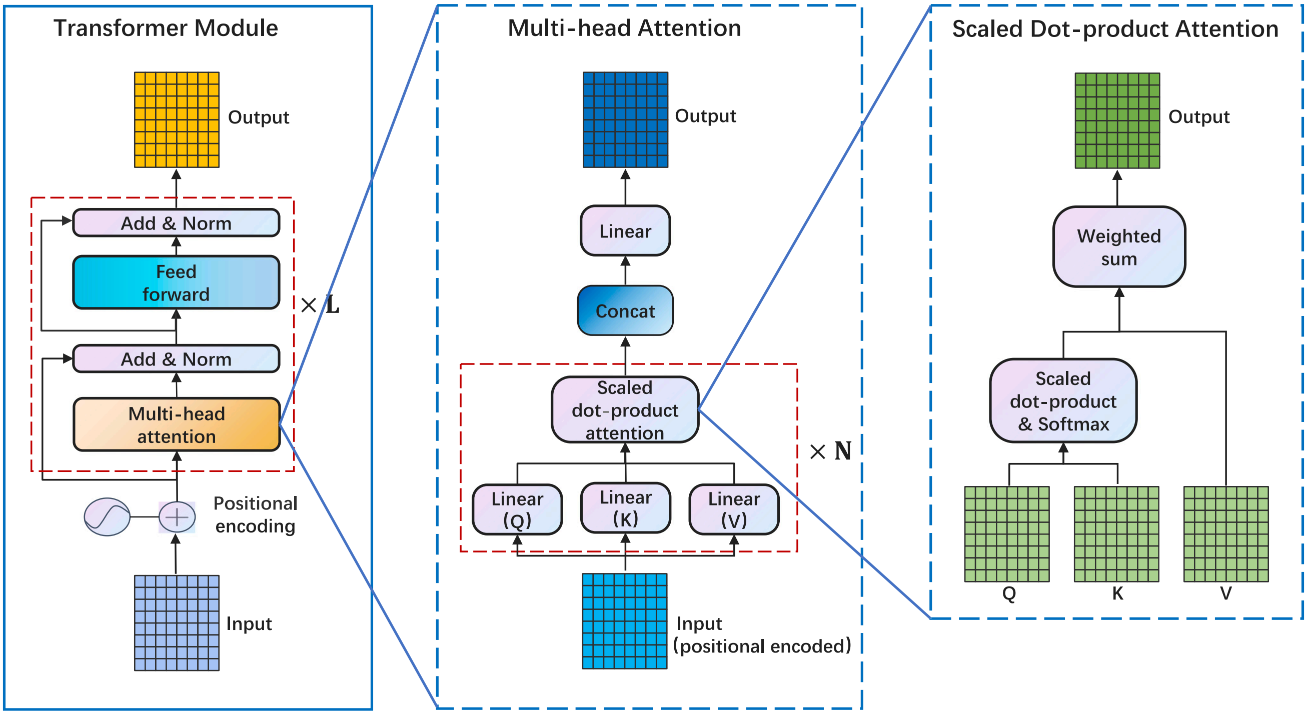


Fig. 3. Structure of the Transformer encoder module.

where FFN(⋅) represents the mapping function of the feedforward layer, ReLU(⋅) is the activation function used within the feedforward layer, W_1 and W_2 are the weight matrices of the linear layers, and b_1 and b_2 are the bias terms of the linear layers.

Then, the output of the feedforward network is further optimized through an Add & Norm layer:

$$H'''^{l} = \mathrm{LayerNorm}(\mathrm{FFN}(H''^{l}) + H''^{l}), \tag{7}$$

where H''^{l} denotes the output of the multi-head attention layer, and LayerNorm(⋅) refers to the layer normalization operation.

Through the above operations, the encoder module of the Transformer progressively extracts global temporal dependency features by stacking multiple encoder units. The output of each encoder layer is passed to the next layer for higher-level feature modeling. This process continues until the final layer, i.e., the Lth encoder layer, where the integration and optimization of global time-series features are completed:

$$H''^{L} = \mathrm{LayerNorm}(\mathrm{FFN}(H'^{L}) + H'^{L}), \tag{8}$$

Through this progressive structure, the Transformer's encoder achieves a well-balanced capability between global modeling and detail capture, providing rich and precise feature representations for the subsequent modules.

2.4. Funet module

The traditional U-Net architecture, characterized by its encoder–decoder structure and skip connections, demonstrates clear advantages in combining contextual information with detailed features. However, for certain tasks, particularly long time-series data modeling, its performance is limited by the following challenges: First, U-Net relies on local convolution and pooling operations for feature extraction, which makes it insufficient in capturing global features, especially when handling complex long-term dependencies. Second, its feature fusion strategy depends on skip connections at corresponding levels in the decoding path. This approach fails to effectively integrate deep-level global

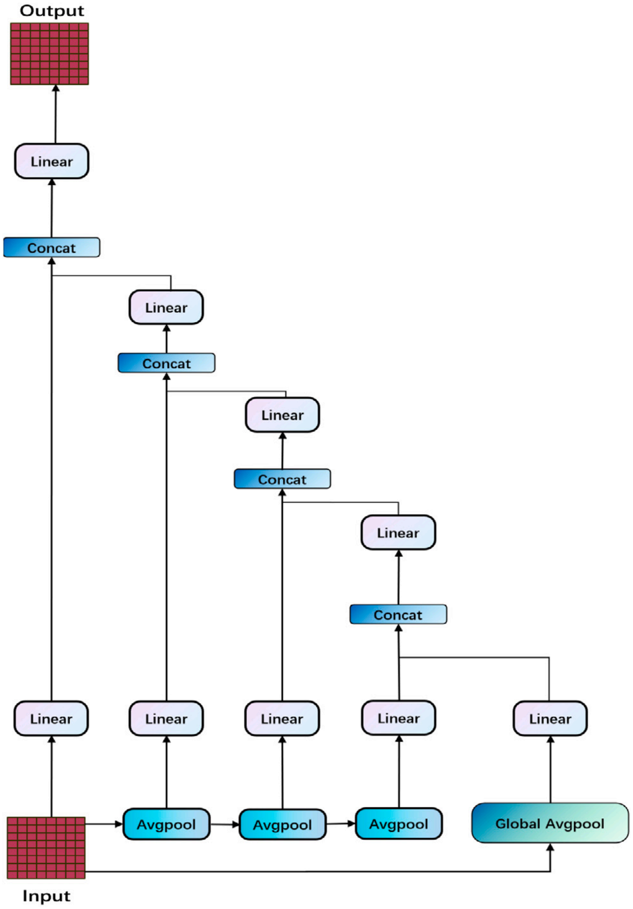


Fig. 4. Structure of the FUnet module.

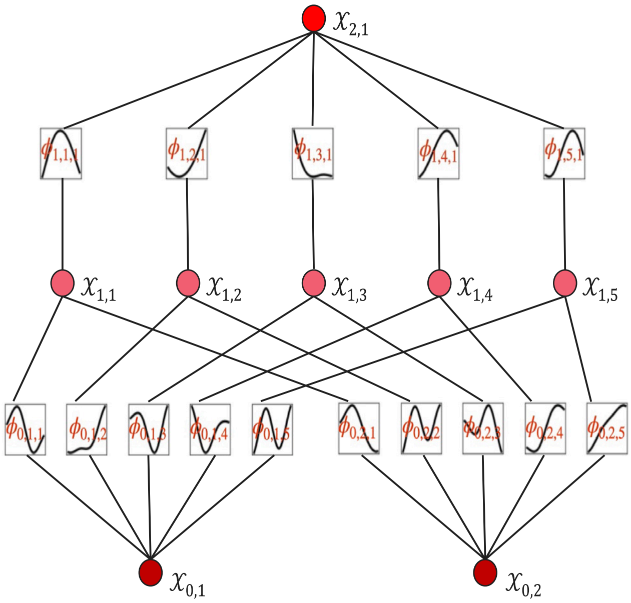


Fig. 5. Structure of the KAN module.

trends with shallow local features. Finally, while the pooling operations during downsampling help compress feature maps, they also risk losing critical detailed information.

To overcome these limitations, this paper designs the FUnet module, as illustrated in Fig. 4, and optimizes the traditional U-Net in the following ways. First, by introducing global average pooling at the bottom layer, FUnet effectively captures overall trend features from high-dimensional data. These global features are transmitted through skip connections to the decoding path to participate in the feature reconstruction process, significantly enhancing the ability to model long-term dependencies. Second, by incorporating the FPN approach, FUnet progressively extracts shallow, deep, and global features using average pooling at each layer. During the upsampling path, these multi-scale features are fused through concatenation and linear transformations. This design of multi-level feature fusion balances information across different scales, enabling the model to efficiently handle time-series data of various lengths and resolutions while significantly improving the efficiency of feature map fusion.

Additionally, to further reduce computational complexity, FUnet incorporates lightweight linear modules, decomposing certain convolution operations to accommodate the processing needs of long time-series data. This lightweight strategy not only enhances computational efficiency but also ensures the effective capture of key features.

Based on these improvements, FUnet is placed downstream of the parallel weighted module integrating the ADTC and Transformer Encoder. By further refining and integrating the feature maps resulting from the weighted fusion of local and global features, FUnet successfully addresses the limitations of the traditional U-Net, providing a more comprehensive and efficient solution for time-series modeling.

2.5. KAN module

The KAN module, inspired by the Kolmogorov–Arnold representation theorem (Liu, Wang, Vaidya, et al., 2024), enables the representation of any continuous multi-dimensional function as a nested composition of single-variable functions. This provides a robust theoretical basis for processing complex high-dimensional data. Specifically, in lithium-ion battery prediction tasks, where intricate nonlinear relationships exist among high-dimensional variables, KAN leverages its unique properties — such as nonlinear activation and dynamic linear combination — to reduce the modeling complexity of these relationships. This enables efficient feature representation and utilization. The KAN module is placed after the FUnet module in the overall architecture as the final prediction layer. Its structure is illustrated in Fig. 5.

To bridge the theoretical Kolmogorov–Arnold representation and its practical implementation, KAN adopts a two-stage process involving local interpolation using B-splines and global nonlinear transformations. The process begins with dynamically generating interpolation nodes based on the distribution of each input feature x_p, and then combining the interpolation results for all features via linear and nonlinear operations. The detailed implementation steps are as follows:

(1) local feature interpolation. KAN first applies B-spline basis functions to interpolate each input feature x_p based on its distribution. The B-spline interpolation dynamically adjusts its grid node positions t_i, ensuring adaptability to varying feature distributions. The interpolation function for each feature is expressed as:

$$KAN(x) = f(x_1, x_2, \dots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \varphi_{q,p}(x_p) \right), \tag{9}$$

where k is the degree of the B-spline, and $B_{i,k-1}(x_p)$ are the lower-order B-spline functions. After obtaining the B-spline values, the final interpolation result for each feature is calculated as:

$$\varphi_{q,p}(x_p) = \sum_{i=1}^{N} w_i B_{i,k}(x_p), \tag{10}$$

where N is the number of grid nodes, w_i are learnable weights, and t_i represents the node positions. These weights w_i enable the module to adaptively learn the contribution of each interpolated feature, capturing local data variations with high precision.

(2) global nonlinear transformation. The interpolated results for all features are then aggregated and passed through a global nonlinear transformation. First, the interpolated features are summed as:

$$y_q = \sum_{p=1}^{n} \varphi_{q,p}(x_p), \tag{11}$$

where n is the number of input features. The aggregated result y_q is then transformed via a nonlinear activation function Φ_q, defined as:

$$\Phi_q(y_q) = \sigma \left(\sum_{j=1}^{m} v_{q,j} \cdot y_{q,j} + b_q \right), \tag{12}$$

where $v_{q,j}$ are learnable weights, b_q is a bias term, and $\sigma(\cdot)$ is the activation function (e.g., sigmoid or ReLU). This step enables the module to model global interactions among input features, effectively capturing long-range dependencies and complex nonlinear relationships.

(3) Kolmogorov-arnold representation integration. The overall KAN function integrates the above steps, achieving a structure that adheres to the Kolmogorov–Arnold representation theorem. The complete function is expressed as:

$$KAN(x) = f(x_1, x_2, \dots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \varphi_{q,p}(x_p) \right), \tag{13}$$

where the inner summation $\sum_{p=1}^{n} \varphi_{q,p}(x_p)$ represents local interpolation results, while the outer summation $\sum_{q=1}^{2n+1}$ incorporates global nonlinear transformations.

(4) practical advantages. KAN's implementation provides several advantages. The use of dynamically generated interpolation nodes and scalable weights ensures that the module adapts to different feature distributions, enabling robust modeling across various datasets. By reducing multi-dimensional functions to single-variable compositions, KAN significantly decreases computational complexity while retaining high accuracy. Additionally, the combination of B-spline interpolation and nonlinear transformations allows KAN to effectively model intricate dependencies in high-dimensional data, making it highly suitable for complex feature spaces.

Through these design principles, KAN serves as a powerful prediction module that captures both local and global patterns in high-dimensional feature spaces, ensuring accurate and robust SOH prediction.

2.6. Specific process

This section details the specific process of the entire framework, which is divided into four stages: data preprocessing and feature extraction, module feature fusion, global prediction, and training and testing.

In the preprocessing stage, the raw health status data X is preprocessed to obtain the initial input X_{re} for the ADTC module. By gradually increasing the convolution kernel size $[k_1, k_2, k_3, k_4]$ through the dilated convolution layers, time-dependent multi-scale features X_1, X_2, X_3, X_4 are extracted, without using padding during the convolution process.

$$X_{\text{re}} = \mathbb{R}^{C\times N\times T} = \text{reshape}(X), \tag{14}$$

$$X_i \in \mathbb{R}^{(C/4)\times N\times(T-S_i)} = \text{Conv2d}_i(X_{re}) \quad ,i = (1,2,3,4), \tag{15}$$

$$\begin{cases} S_i = d'_i \cdot (k_i - 1), i = (1,2,3,4), \\ k = [2,4,8,\frac{3T}{4}], \end{cases} \tag{16}$$

$$\begin{cases} \text{Conv2d}_i(\cdot) = (C, \frac{C}{4}, (1,k_i), d'_i), \\ d'_i = (1, \max(1, \frac{T+1}{2k_i})), \end{cases} \tag{17}$$

where X is the original health status data, X_{re} represents the reshaped X, $\in {}^{C\times N\times T}$ represents the shape of the real tensor to which it belongs, C represents the number of features, N represents the number of sequences, and T represents the time step length. $X_i, i = (1,2,3,4)$ is the set of four multi-scale features obtained by applying convolutions with different kernel sizes $k_i, i = (1,2,3,4)$ to X_{re}. Conv2d$(\cdot)$ represents the convolution operation. $S_i, i = (1,2,3,4)$ indicates the stride length of the convolution operation, while $k = [2,4,8,3T/4]$ corresponds respectively to k_1, k_2, k_3, k_4, which are the sizes of the four convolution kernels. The kernel size $\left(C, \frac{C}{4}, (1,k_i), d'_i\right)$ indicates that the input channel count for the 2D convolution is C, the output channel count is $\frac{C}{4}$, $(1,k_i)$ represents the dimensions of the 2D convolution kernel, and d'_i is the dilation factor.

After convolving with different kernel sizes, four multi-scale features X_1, X_2, X_3, X_4 are generated. To unify the time dimensions of these multi-scale features, the shortest feature along the time dimension is used as the standard. The features are truncated to a length of $(T - S_{\text{max}})$, resulting in four unified multi-scale features X'_1, X'_2, X'_3, X'_4:

$$X'_i \in \mathbb{R}^{(C/4)\times N\times(T-S_{\text{max}})} = X_i[\ldots, -(T - S_{\text{max}}) :], i = (1,2,3,4), \tag{18}$$

where $(T - S_{\text{max}})$ indicates the last $(T - S_{\text{max}})$ time steps uniformly truncated along the time dimension, and S_{max} represents the maximum value among $S_i, i = (1,2,3,4)$.

Subsequently, these features are concatenated along the channel dimension to restore the same number of features as the initial input X_{re} of the ADTC module, resulting in the concatenated feature tensor X_T:

$$X_T \in \mathbb{R}^{C\times N\times(T-S_{\text{max}})} = \text{Concat}(X'_1, X'_2, X'_3, X'_4), \tag{19}$$

Next, the concatenated feature tensor X_T is processed by a Multi-Layer Perceptron (MLP) to generate the output of the first dilated initialization layer X'_T:

$$X'_T \in \mathbb{R}^{C\times N\times T} = \text{MLP}(X_T), \tag{20}$$

Next, through a similar process, the output X'_S of the second dilated initialization layer (DIL) is obtained.

Subsequently, X'_T and X'_S are processed through the Tanh and Sigmoid activation functions respectively, and then combined via element-wise multiplication to generate the final output Y_{ADTC} of the ADTC module:

$$Y_{ADTC} \in \mathbb{R}^{C\times N\times T} = \text{Tanh}(X'_T) \odot \sigma(X'_S), \tag{21}$$

where Tanh$(\cdot)$ represents the Tanh activation function, $\sigma(\cdot)$ represents the Sigmoid activation function, and $\odot$ represents the element-wise multiplication operation.

In the parallel structure, while the input X_{re} is fed into the ADTC module, it is also fed into the Transformer encoder module. Through the processing of multi-head attention and fully-connected layers, the global feature representation Y_{TRM} is generated:

$$Y_{TRM} = TransformerEncoder(X_{re}), \tag{22}$$

where $TransformerEncoder(\cdot)$ represents the function of the Transformer encoder module, which includes Equations (1) to (8).

Next, the outputs from the ADTC module and the Transformer encoder module are combined through weighted fusion to generate the input feature tensor for the FUnet module:

$$Z = \alpha_{\text{ADTC}} \cdot Y_{\text{ADTC}} + \alpha_{\text{TRM}} \cdot Y_{\text{TRM}}, \tag{23}$$

In the FUnet module, a feature pyramid structure is constructed, starting from the second layer to extract features through pooling layers, resulting in the extracted features Z_2, Z_3, Z_4, Z_5:

$$\begin{cases} Z_1 \in \mathbb{R}^{C\times N\times T_1} = Z, \quad T_1 = T, \\ Z_i \in \mathbb{R}^{C\times N\times T_i} = AvgPool(Z_{i-1}), i = (2,3,4), \\ Z_5 \in \mathbb{R}^{C\times N\times T_5} = GlobalAvgPool(Z_1), T_5 = 1, \end{cases} \tag{24}$$

where $T_i, i = (2,3,4)$ is obtained using the calculation formula len$(\cdot)$, which can be expressed as follows:

$$\text{len}(T_i) = \frac{T_{i-1} + 2\times\text{padding} - \text{kernel\_size}}{\text{stride}} + 1, i = (2,3,4), \tag{25}$$

Subsequently, the features from each layer Z_1, Z_2, Z_3, Z_4, Z_5 are mapped through a linear layer to intermediate features E_1, E_2, E_3, E_4, E_5:

$$\begin{cases} D_i = \text{Linear}(\text{Concat}(D_{i+1}, E_i)), i = (1,2,3), \\ D_4 = \text{Linear}(\text{Concat}(E_5, E_4)), \end{cases} \tag{26}$$

Subsequently, based on the pyramid structure, intermediate features E_1, E_2, E_3, E_4, E_5 are concatenated along the time dimension from top to bottom, and the next layer of intermediate features D_{i+1} is generated through a linear layer. This process is repeated until the feature D_1 of the first layer is obtained.

$$\begin{cases} D_i = \text{Linear}(\text{Concat}(D_{i+1}, E_i)), i = (1,2,3), \\ D_4 = \text{Linear}(\text{Concat}(E_5, E_4)), \end{cases} \tag{27}$$

Next, the new feature from the first layer D_1 is taken as the fused multi-scale feature O, and after passing through a linear layer and a reshape operation, the final fused feature O' is obtained:

$$O' \in \mathbb{R}^{N\times T\times C} = \text{reshape}(W_o \in \mathbb{R}^{C\times C} \cdot O), \tag{28}$$

where $W_o \in \mathbb{R}^{C\times C}$ represents the transformation matrix of the linear layer, which is a transformation matrix of dimension $C \times C$.

Finally, based on the high-dimensional features obtained from the FUnet module, the KAN is employed to transform the fused features O' into point prediction results O'':

$$O'' \in \mathbb{R}^1 = \text{KAN}(O'), \tag{29}$$

where KAN$(\cdot)$ represents the functional process of the KAN module.

In the testing phase, the preprocessed test data X_{test} is fed into the trained model Model$(\cdot)$. Using a sliding window to accumulate predicted points, the complete health state prediction results O''_{pred} are

Algorithm 1 Training and Testing Process of the Parallel Weighted ADTC-Transformer-FUnet-KAN Model

Input: Lithium-ion battery operation datasets

Parameters: Sequence length, sliding window size, batch size, learning rate, number of epochs, dropout rate, feature channels, kernel size, number of Transformer layers, hidden layer dimensions, and other structural hyperparameters.

Output:

- Trained Parallel Weighted ADTC-Transformer-FUnet-KAN model with determined weights, fixed-step prediction results, moving-step prediction results, and evaluation metrics: Relative Error (RE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

```
for each battery_name in Battery_list do
    Normalize the battery operation data using Min-Max normalization.
    Divide the data into training and testing sets.
end for
Store the preprocessed data as train_data (train_x, train_y) and test_data (test_x, test_y).
Initialize the ADTC, Transformer, FUnet, and KAN modules.
Define the weight combination hyperparameters (e.g., $\alpha_{ADTC}$ and $\alpha_{TRM}$).
Configure the optimizer (Adam) with the initial learning rate.
for $epoch = 1, 2, \ldots, max\_epochs$ do
    for each batch in train_data do
        Input the current batch into the ADTC and Transformer modules.
        Perform weighted fusion of extracted features.
        Use the FUnet module to extract and fuse multi-scale features.
        Use the KAN module to generate the final prediction.
        Use a sliding window to accumulate predicted points and generate the health state prediction results.
        Compute the prediction error (e.g., MSE) and update the model weights using the Adam optimizer.
    end for
    if $epoch \,\%\, 10 == 0$ then
        Record intermediate evaluation metrics (RE, MAE and RMSE).
        Save the current model state.
    end if
end for
Save the trained ADTC-Transformer-FUnet-KAN model.
for each batch in test_data do
    Select the last feature_size points of test_x as the initial input.
    for $step = 1, 2, \ldots, prediction\_length$ do
        Predict the next point using the model, record the predicted value as single-step prediction results, and append it to test_x.
    end for
end for
for each batch in test_data do
    Select the first feature_size points (ground truth) from test_x to construct the input sequence.
    for $step = 1, 2, \ldots, prediction\_length$ do
        Predict the next point using the model and record the predicted value as multi-step prediction results.
    end for
end for
Compare the predicted values with the ground truth test_y to calculate final evaluation metrics (RE, MAE and RMSE).
```

generated:

$$O''_{\text{pred}} \in \mathbb{R}^{I \times T_{\text{pred}}} = \text{Model}(X_{\text{test}}), \tag{30}$$

where $O''_{\text{pred}} \in \mathbb{R}^{I \times T_{\text{pred}}}$ represents the test results, O''_{pred} is a scalar containing T_{pred} data points, and T_{pred} denotes the number of sliding steps taken by the sliding window from the start of prediction to the endpoint (the maximum recorded discharge cycle count of the battery).

After detailing the specific workflow of the model, a concise summary of the core algorithm design is provided in the form of pseudocode to facilitate quick understanding and implementation by engineering practitioners. This approach aims to present the key steps in a clearer and more structured manner, making it directly applicable to real-world engineering practices while also serving as an operational reference for future model improvements and extensions. The pseudocode is outlined in Algorithm 1.

3. Experiments and discussion

The experimental environment is based on the Python 3.11.5 interpreter, with the model implemented using the Keras 3.3.3 and TensorFlow 2.16.1 deep learning frameworks. All experiments were conducted on an Intel 64 Family 6 Model 191 Stepping 2 processor.

The “Leave-One-Out” cross-validation method was used for training and testing, with the division of datasets shown in Table 1. This approach iterates four times, each time reserving one battery dataset for testing, enabling evaluation on all four batteries in the dataset. All data were normalized to the range $[0, 1]$ by dividing by the rated capacity of each battery. The normalized data were then input into the trained model, and the optimization was carried out using reverse-mode automatic differentiation with the Adam optimizer. The mean squared error (MSE) was employed as the loss function.

3.1. Dataset description

To validate the effectiveness of the proposed model, three major datasets were utilized: NASA, CALCE, and Wenzhou Randomized Battery Data (WRBD). These datasets provide battery degradation information under various experimental conditions.

(1) NASA dataset: This dataset includes test data from 18650 lithium-ion batteries (B0005, B0006, B0007, B0018). These tests were conducted at a room temperature of 24 °C using a constant current

Table 1
Division details of the NASA, CALCE, and WRBD datasets.

Dataset	Experiment	Trained battery	Tested battery
NASA	Experiment 1	B0006, B0007, B0018	B0005
	Experiment 2	B0005, B0007, B0018	B0006
	Experiment 3	B0005, B0006, B0018	B0007
	Experiment 4	B0005, B0006, B0007	B0018
CALCE	Experiment 1	CS2_36, CS2_37, CS2_38	CS2_35
	Experiment 2	CS2_35, CS2_37, CS2_38	CS2_36
	Experiment 3	CS2_35, CS2_36, CS2_38	CS2_37
	Experiment 4	CS2_35, CS2_36, CS2_37	CS2_38
WRBD	Experiment 1	A_003, A_004, A_006	A_001
	Experiment 2	A_001, A_004, A_006	A_003
	Experiment 3	A_001, A_003, A_006	A_004
	Experiment 4	A_001, A_003, A_004	A_006

(CC) mode for charging at 1.5 A up to 4.2 V, followed by a constant voltage (CV) mode until the current dropped to 20 mA. The discharge process was performed at a constant current of 2 A with cut-off voltages of 2.7 V, 2.5 V, 2.2 V, and 2.5 V, respectively. The end-of-life (EOL) criterion for the batteries was defined as a capacity fade to 30% of the initial rated capacity, i.e., from 2 Ah to approximately 1.4 Ah. This dataset provides detailed records of degradation trends under various conditions and is a widely used benchmark dataset in the field of battery SOH prediction. The specific trends are illustrated in Fig. 6 (a).

(2) CALCE dataset: This dataset, provided by the Center for Advanced Life Cycle Engineering (CALCE), includes data from four batteries (CS2_35, CS2_36, CS2_37, CS2_38). The tests were conducted under controlled conditions at a temperature of 1 °C. The charging was carried out in constant current (CC) mode up to 4.2 V, followed by constant voltage (CV) charging until the current dropped to 20 mA. The discharge process was conducted in constant current mode with a cut-off voltage of 2.7 V. The end-of-life (EOL) criterion was similarly defined as a 30% drop in rated capacity, i.e., from 1.1 Ah to approximately 0.77 Ah. This dataset captures degradation characteristics under rigorous laboratory conditions, providing valuable data support for SOH prediction modeling. The specific trends are illustrated in Fig. 6 (b).

(3) WRBD dataset: This dataset (Lyu, Zhang, Zio, et al., 2024) collected and processed by the laboratory, was developed to supplement publicly available datasets such as NASA and CALCE, addressing their limitations by providing unique, high-quality degradation data that enhances the diversity and robustness of training and evaluation processes. The dataset includes a diverse range of lithium-ion battery testing data under various protocols, such as bench mode, complex mode, and random mode, to ensure broad applicability and representativeness. The data collection spanned four years (2020 to 2023), covering multiple years of battery usage cycles, with a total data acquisition time exceeding 26 million minutes. The dataset contains over 0.5 million charge–discharge cycles from up to 300 batteries, capturing degradation information across a wide variety of conditions. For this paper, we specifically selected data from four batteries (A_001, A_003, A_004, A_006), tested under the standard constant current/constant voltage (CC/CV) protocol in the bench mode. The charging process used a 1C current up to 4.2 V, followed by a constant voltage phase until the current dropped to 0.05 C, while discharge was conducted at a 1C current until a cut-off voltage of 2.75 V. All tests were performed at room temperature with periodic capacity fade recordings, and the end-of-life (EOL) criterion was defined as a capacity reduction from 1.0 Ah to approximately 0.7 Ah. The experimental platform used for this dataset is shown in Fig. 7(a), and the specific degradation trends for the selected batteries are illustrated in Fig. 7(b). By focusing on the standard mode data of these four batteries, the selected dataset ensures consistency and representativeness, making it ideal for state-of-health estimation and battery degradation modeling. This carefully designed data collection process bridges the gap between controlled experimental data and real-world applications, ensuring the reliability and applicability of the dataset across various research scenarios.

3.2. Evaluation metrics

To evaluate the prediction performance of the proposed model, three widely used metrics were employed: RE, MAE and RMSE. These metrics are defined as follows:

$$\begin{cases} \text{RE} = \dfrac{1}{N-T} \displaystyle\sum_{t=T+1}^{N} \dfrac{\left| S\hat{O}H_t - SOH_t \right|}{SOH_t}, \\ \text{MAE} = \dfrac{1}{N-T} \displaystyle\sum_{t=T+1}^{N} \| SOH_t - S\hat{O}H_t \|, \\ \text{RMSE} = \sqrt{\dfrac{1}{N-T} \displaystyle\sum_{t=T+1}^{N} \left(SOH_t - S\hat{O}H_t \right)^2}, \end{cases} \tag{31}$$

where $S\hat{O}H_t$ represents the estimated SOH at time t, n denotes the total number of time points in the current battery sequence, and T is the end time point of the training data in the sequence.

The ablation experiments were conducted by comparing the following models: U-Net, FUnet, FUnet+ADTC, FUnet+ADTC+TRM, FUnet+TCN+TRM+KAN, and the full model (FUnet+ADTC+TRM+KAN), where TCN represents traditional temporal convolution and TRM represents the Transformer module. Additionally, comparative experiments were carried out by benchmarking against state-of-the-art models, including Dual-LSTM (Shi & Chehade, 2021), CNN-GRU (Sun, Wang, Xiao, Peng, & Zhou, 2024), DeTransformer (Chen, Hong, & Zhou, 2022), TCN-ECANet-GRU (Xiang, Li, Zhang, et al., 2024), Adaptive MAGNN-TCN (Ye, Wang, Yang, et al., 2025), and MEWOA-VMD and Transformer (Chen et al., 2024). Each experiment was repeated five times, and the average results were reported. The experimental outcomes are presented in the following figure.

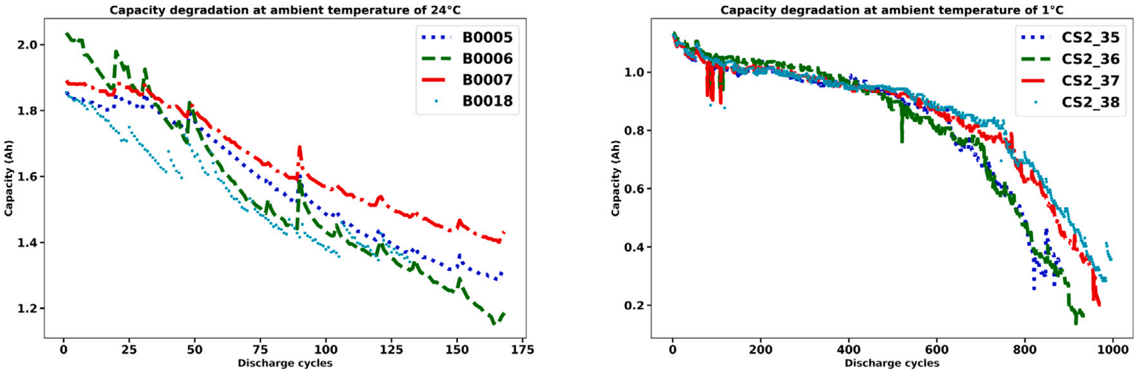
3.3. Parameter configuration

The selection and optimization of hyperparameters play a crucial role in achieving optimal performance in the process of model construction and training. In this paper, the sliding window size was fixed at 1, and the number of training epochs was set to 500. Other hyperparameters were optimized using the grid search method. The hyperparameters and their corresponding search ranges are as follows: feature channels (k) were searched within $1, 2, 4, 8, 16$, sequence length (m) within $8, 16, 32, 64, 128$, batch size (p) within $8, 16, 32, 64, 128$, number of Transformer layers (L) within $1, 2, 4, 8, 16$, number of attention heads (n) within $1, 2, 4, 8, 16$, hidden layer dimension (h) within $2, 4, 8, 16, 64$, learning rate (τ) within $0.0001, 0.0005, 0.001, 0.005, 0.01$, and dropout ratio ($d$) within $0.0001, 0.0005, 0.001, 0.005, 0.01$. The final optimized hyperparameters for each dataset are shown in Table 2.

The optimization process was guided by the root mean square error (RMSE) metric, which served as the primary evaluation criterion during grid search. RMSE was chosen over relative error (RE) because, in datasets with small rated capacities, such as CALCE and WRBD, RE scores were found to fluctuate significantly. This instability makes RE an unreliable metric for consistent optimization, particularly in datasets with limited capacity ranges. RMSE, being an absolute measure, provides a more robust and stable evaluation standard for hyperparameter optimization, ensuring more reliable model performance across diverse datasets.

3.4. Single-step and multi-step predictions

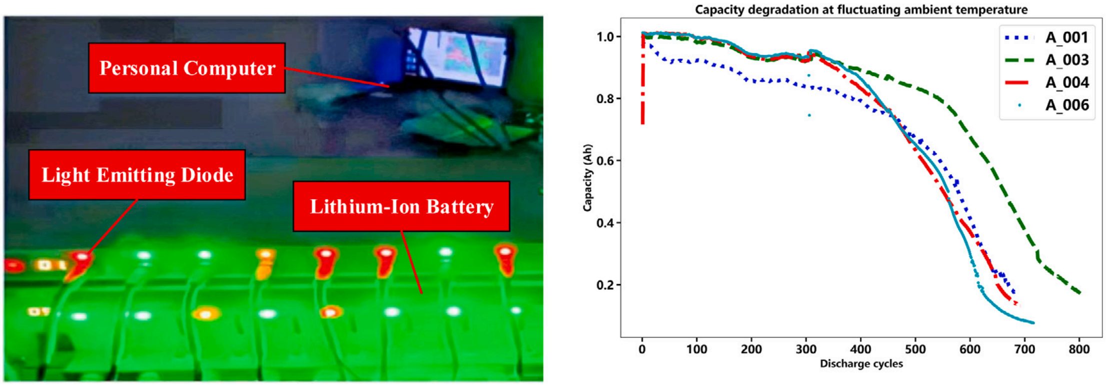
The proposed model was evaluated using single-step prediction (Fixed Step) and multi-step prediction (Moving Step), with the results shown in Fig. 8. In single-step prediction, the model predicts one time step at a time, using actual observed values as input for subsequent predictions. In contrast, multi-step prediction predicts multiple time steps at once, using the predicted values as input for subsequent



(a) NASA dataset.

(b) CALCE dataset.

Fig. 6. Capacity degradation trends of batteries in public datasets.



(a) Experiment platform.

(b) Capacity degradation trends.

Fig. 7. Experimental platform and capacity degradation trends of batteries in the private dataset.

Table 2
Optimized hyperparameters for different datasets for the proposed model.

Dataset	Hyper-parameters			Specific parameters				
	k	m	p	L	n	h	τ	d
NASA	16	16	128	3	8	16	0.005	0.0001
CALCE	16	64	128	2	8	8	0.003	0.0001
WRBD	16	64	128	2	8	8	0.005	0.0001

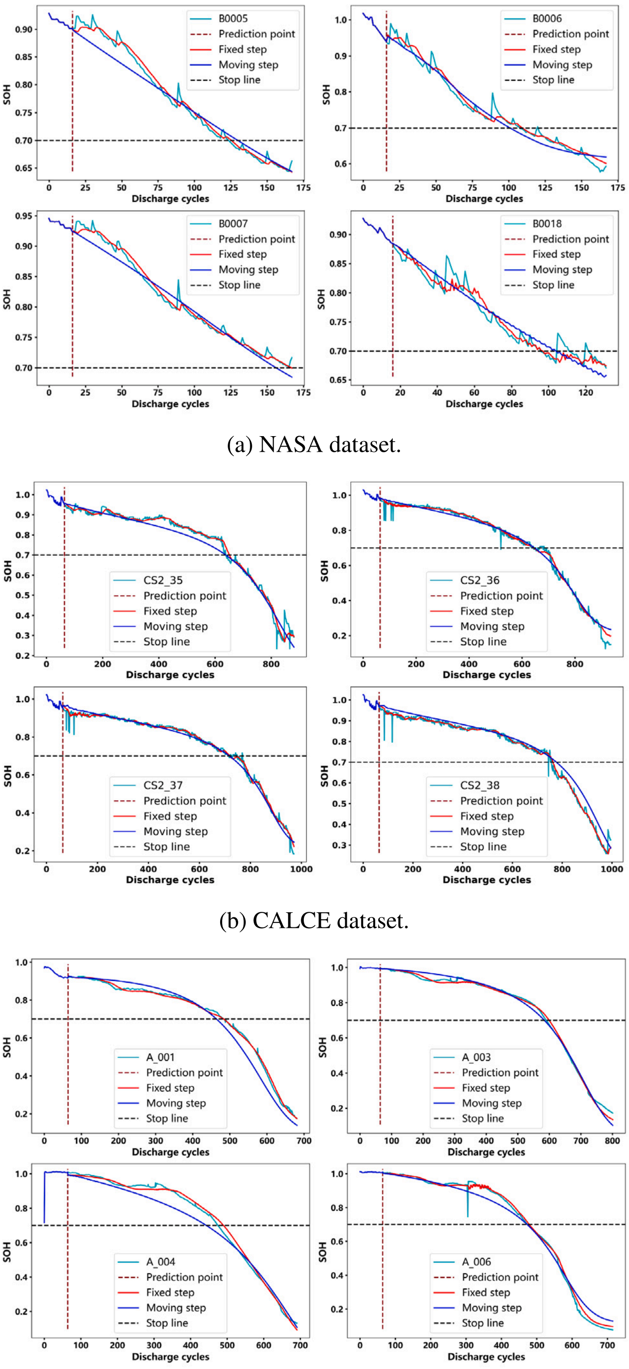
steps. When evaluating SOH prediction performance on the NASA, CALCE, and WRBD datasets, single-step prediction demonstrated exceptional accuracy and precision across all datasets. This prediction mode closely followed the actual battery discharge curves, showcasing its high responsiveness. Such accuracy makes single-step prediction particularly suitable for real-time monitoring and forecasting applications that require immediate feedback and high precision.

Multi-step prediction performed well in simulating the long-term degradation trends of battery SOH, though prediction accuracy typically decreased during the latter stages of the battery lifecycle. The smoothness of this prediction mode is highly beneficial for scenarios requiring long-term trend analysis, such as maintenance scheduling and lifespan forecasting. However, its performance is significantly affected by cumulative errors, especially during phases of rapid battery degradation.

Regarding the characteristics of the datasets, the NASA dataset exhibited relatively stable degradation processes, with prediction results showing good alignment and low error rates, highlighting the reliability of the model. The CALCE dataset, obtained under strictly controlled experimental conditions, revealed varying degradation patterns, and the model demonstrated excellent adaptability and accuracy under these conditions, particularly in handling rapid changes in battery capacity. The WRBD dataset, which provided more complex real-world battery usage scenarios, featured more variable and irregular degradation behavior. Despite these challenges, the model was still able to capture long-term trends effectively, demonstrating its broad applicability and strong adaptability.

3.5. Ablation experiments

A series of ablation experiments were conducted to evaluate the contributions of each module in the proposed SOH prediction model, as shown in Fig. 9. The baseline U-Net model demonstrated its fundamental ability to capture basic battery characteristics, but it showed limitations when handling complex battery degradation patterns. The integration of the FUnet module, with its FPN, significantly improved the processing of multi-scale information and optimized the model's ability to capture battery capacity changes across various stages. When combined with the ADTC, the FUnet further enhanced its ability to handle time-series data, particularly in recognizing long-term dependencies in battery performance. Adding the Transformer module to the FUnet+ADTC configuration enabled the model to precisely capture global dependencies, which improved prediction accuracy and stability. Finally, the inclusion of the KAN in the full model demonstrated outstanding performance in managing high-dimensional features and complex functional mappings. This addition significantly reduced key error metrics, including RE, MAE and RMSE, underscoring the superior capability of the composite model for handling challenging battery degradation data. Overall, the incremental inclusion of each module

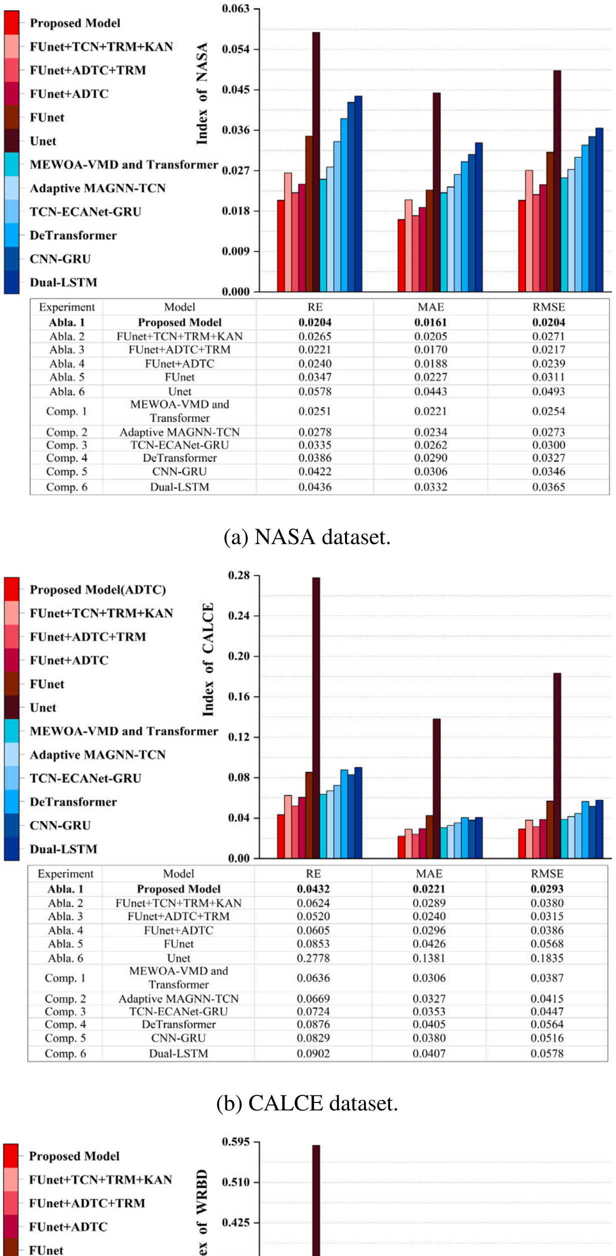


(a) NASA dataset.

(b) CALCE dataset.

(c) WRBD dataset.

Fig. 8. Results of single-step prediction and multi-step prediction.



Experiment	Model	RE	MAE	RMSE
Abla. 1	**Proposed Model**	**0.0204**	**0.0161**	**0.0204**
Abla. 2	FUnet+TCN+TRM+KAN	0.0265	0.0205	0.0271
Abla. 3	FUnet+ADTC+TRM	0.0221	0.0170	0.0217
Abla. 4	FUnet+ADTC	0.0240	0.0188	0.0239
Abla. 5	FUnet	0.0347	0.0227	0.0311
Abla. 6	Unet	0.0578	0.0443	0.0493
Comp. 1	MEWOA-VMD and Transformer	0.0251	0.0221	0.0254
Comp. 2	Adaptive MAGNN-TCN	0.0278	0.0234	0.0273
Comp. 3	TCN-ECANet-GRU	0.0335	0.0262	0.0300
Comp. 4	DeTransformer	0.0386	0.0290	0.0327
Comp. 5	CNN-GRU	0.0422	0.0306	0.0346
Comp. 6	Dual-LSTM	0.0436	0.0332	0.0365

(a) NASA dataset.

Experiment	Model	RE	MAE	RMSE
Abla. 1	**Proposed Model**	**0.0432**	**0.0221**	**0.0293**
Abla. 2	FUnet+TCN+TRM+KAN	0.0624	0.0289	0.0380
Abla. 3	FUnet+ADTC+TRM	0.0520	0.0240	0.0315
Abla. 4	FUnet+ADTC	0.0605	0.0296	0.0386
Abla. 5	FUnet	0.0853	0.0426	0.0568
Abla. 6	Unet	0.2778	0.1381	0.1835
Comp. 1	MEWOA-VMD and Transformer	0.0636	0.0306	0.0387
Comp. 2	Adaptive MAGNN-TCN	0.0669	0.0327	0.0415
Comp. 3	TCN-ECANet-GRU	0.0724	0.0353	0.0447
Comp. 4	DeTransformer	0.0876	0.0405	0.0564
Comp. 5	CNN-GRU	0.0829	0.0380	0.0516
Comp. 6	Dual-LSTM	0.0902	0.0407	0.0578

(b) CALCE dataset.

Experiment	Model	RE	MAE	RMSE
Abla. 1	**Proposed Model**	**0.0763**	**0.0341**	**0.0437**
Abla. 2	FUnet+TCN+TRM+KAN	0.0986	0.0410	0.0545
Abla. 3	FUnet+ADTC+TRM	0.0871	0.0362	0.0482
Abla. 4	FUnet+ADTC	0.0980	0.0423	0.0579
Abla. 5	FUnet	0.1408	0.0601	0.0781
Abla. 6	Unet	0.5881	0.2180	0.2608
Comp. 1	MEWOA-VMD and Transformer	0.0884	0.0410	0.0476
Comp. 2	Adaptive MAGNN-TCN	0.0927	0.0436	0.0510
Comp. 3	TCN-ECANet-GRU	0.1006	0.0472	0.0550
Comp. 4	DeTransformer	0.1522	0.0678	0.0866
Comp. 5	CNN-GRU	0.1440	0.0637	0.0793
Comp. 6	Dual-LSTM	0.1566	0.0682	0.0888

(c) WRBD dataset.

Fig. 9. Results of the ablation experiments.

not only validated its effectiveness but also enhanced the model's practicality and reliability in battery health monitoring applications.

A series of comparative experiments were conducted to evaluate the performance of the proposed complete model (FUnet+ADTC+TRM+KAN) against other state-of-the-art models, including Dual-LSTM, DeTransformer, CNN-GRU, TCN-ECANet-GRU, and MEWOA-VMD and Transformer, in lithium-ion battery SOH prediction. While simpler composite models provided reliable predictions under specific conditions, they often showed limitations when dealing with high-dimensional and complex datasets. In contrast, more sophisticated composite models excelled in prediction accuracy and stability due to their advanced

Table 3
Performance results of weight adjustment experiments for the proposed model.

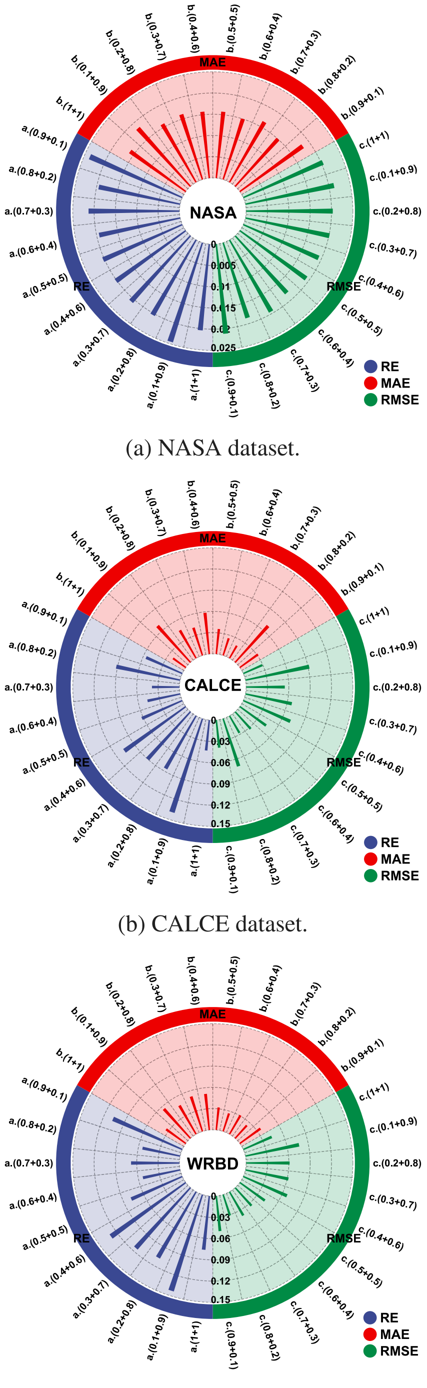
Dataset	Weight distribution		Evaluation metric		
	ADTC	TRM	RE	MAE	RMSE
NASA	1	1	0.0204	0.0161	0.0204
	0.1	0.9	0.0245	0.0184	0.0212
	0.2	0.8	0.0204	0.0158	0.0203
	0.3	0.7	0.0208	0.0161	0.0201
	0.4	0.6	0.0202	0.0157	0.0193
	0.5	0.5	0.0203	0.0157	0.0192
	0.6	**0.4**	**0.0194**	**0.0150**	**0.0181**
	0.7	0.3	0.0213	0.0164	0.0195
	0.8	0.2	0.0195	0.0150	0.0186
	0.9	0.1	0.0237	0.0182	0.0212
CALCE	1	1	0.0432	0.0221	0.0293
	0.1	0.9	0.1390	0.0692	0.0916
	0.2	0.8	0.0860	0.0456	0.0543
	0.3	0.7	0.0902	0.0412	0.0666
	0.4	0.6	0.1066	0.0585	0.0723
	0.5	0.5	0.0621	0.0355	0.0456
	0.6	0.4	0.0469	0.0255	0.0326
	0.7	**0.3**	**0.0386**	**0.0205**	**0.0279**
	0.8	0.2	0.0918	0.0692	0.0707
	0.9	0.1	0.0557	0.0315	0.0392
WRBD	1	1	0.0763	0.0341	0.0437
	0.1	0.9	0.1428	0.0558	0.0770
	0.2	0.8	0.1073	0.0468	0.0611
	0.3	0.7	0.1148	0.0518	0.0617
	0.4	0.6	0.1299	0.0512	0.0671
	0.5	0.5	0.0785	0.0320	0.0434
	0.6	0.4	0.0535	0.0267	0.0343
	0.7	0.3	0.0680	0.0308	0.0387
	0.8	**0.2**	**0.0541**	**0.0245**	**0.0310**
	0.9	0.1	0.1066	0.0358	0.0498

data processing capabilities, particularly in challenging environments like the WRBD dataset, which features substantial noise and variability. The proposed FUnet+ADTC+TRM+KAN model demonstrated the best overall performance across all comparative experiments. By integrating the Feature Pyramid Network, Adaptive Dilated Temporal Convolution, Transformer, and Kolmogorov–Arnold Network, the model significantly improved feature extraction efficiency and accuracy while optimizing the capture and processing of long-term dependencies. Experimental results revealed that the proposed model consistently outperformed its competitors in key evaluation metrics such as RE, MAE and RMSE. Its superior performance was particularly evident in datasets characterized by complex degradation patterns and noisy data. These comparative experiments not only validated the design and effectiveness of the new model but also highlighted its potential and value in real-world applications. The results provide a solid foundation for advancing battery health management systems and demonstrate the model's capability to address the challenges of future energy storage technologies.

3.6. Weight adjustment experiments

In the weight adjustment experiments, nine combinations of ADTC and Transformer weights were tested, with the sum fixed at 1 (e.g., 0.1+0.9, 0.2+0.8, etc.) to explore the balance between local and global features. The results were compared to an unweighted (1:1) distribution. This experiment aimed to investigate how weight distributions are optimized across different datasets, providing insights for model design. The results, shown in Table 3 and illustrated in Fig. 10, revealed that optimal weight distributions varied across datasets, with local features dominating for short sequences (e.g., NASA dataset) and a balance of local and global features needed for longer sequences (e.g., CALCE, WRBD datasets).

For the NASA dataset, the optimal ADTC ratio was 0.6, indicating that local features dominated. The CALCE dataset, with longer sequences, required a balance, with an optimal ADTC ratio of 0.7. The WRBD dataset, characterized by longer and more complex sequences, required a higher emphasis on local features, with an optimal ADTC ratio of 0.8. These findings demonstrate the importance of adapting weight distributions based on dataset characteristics.



(a) NASA dataset.
(b) CALCE dataset.
(c) WRBD dataset.

Fig. 10. Impact of different weight distributions on performance.

The NASA dataset showed relatively minor performance fluctuations due to its short sequences, where local features predominated. In contrast, the CALCE and WRBD datasets exhibited significant performance changes, highlighting the importance of balancing local and global features for optimal performance in longer, more complex sequences. The WRBD dataset showed even greater sensitivity to weight distribution due to its complexity.

Overall, the ADTC module excelled in local feature extraction, while the Transformer module captured global dependencies effectively. The optimal weight distribution for each dataset was found to be 0.6:0.4 for NASA, 0.7:0.3 for CALCE, and 0.8:0.2 for WRBD. These results emphasize the importance of dynamically adjusting weights based on sequence length and data complexity. Future work will focus on dynamic weight allocation and optimizing module collaboration to improve model adaptability.

4. Conclusion

This paper proposes a deep learning-based multi-stage, multi-scale feature extraction and fusion model for lithium-ion battery SOH prediction. The model integrates ADTC, Transformer Encoder, enhanced FUnet module, and KAN. By leveraging multi-stage and multi-scale feature extraction and fusion, the proposed model demonstrates significantly improved predictive performance compared to traditional models on NASA, CALCE, and WRBD datasets. The model achieves collaborative extraction of local and global features through the parallel weighting mechanism of ADTC and Transformer modules, enabling precise capture of short-term local dependencies and long-term global trends while dynamically balancing the contributions of each feature via weight optimization. Additionally, the FUnet module incorporates an FPN structure and global average pooling techniques, enhancing the efficiency of multi-scale feature fusion and improving adaptability to diverse, complex time-series data. In the final prediction phase, the KAN module, leveraging nonlinear activation functions and B-spline interpolation, effectively processes the complex relationships between high-dimensional features, significantly improving prediction accuracy and robustness. Experimental results demonstrate that the proposed model accurately captures critical changes in battery health, adapts to diverse data patterns, and achieves superior performance in prediction accuracy and stability. This paper provides an effective technical approach for efficient lithium-ion battery management and safe operation. Moreover, it introduces valuable improvements in algorithm design for time-series prediction, offering practical insights and laying the foundation for further research and applications in related fields.

Future research will focus on optimizing the model structure, including exploring more flexible dynamic weight allocation mechanisms and collaborative optimization strategies among modules to ensure stable performance under diverse operating conditions and complex sequence data scenarios. Additionally, the model will undergo extensive testing and validation in more real-world application scenarios. By integrating cutting-edge deep learning techniques, efforts will aim to further enhance the model's adaptability and reliability in SOH prediction tasks.

CRediT authorship contribution statement

Chuang Chen: Writing – original draft, Methodology, Investigation, Funding acquisition. **Yuheng Wu:** Writing – original draft, Validation, Software, Methodology. **Jiantao Shi:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Dongdong Yue:** Writing – review & editing, Conceptualization. **Ge Shi:** Writing – review & editing, Resources. **Dongzhen Lyu:** Writing – review & editing, Resources, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62303217, Grant 62373184, and Grant 52405126; in part by China Postdoctoral Science Foundation, China under Grant 2024M751977; in part by the Vice General Project of Science and Technology in Jiangsu Province under Grant FZ20241078; in part by the Natural Science Foundation of Jiangsu Province, China under Grant BK20240531; in part by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant 23KJB510006, and 24KJB120006; in part by the Fundamental Research Funds for the Central Universities, China under Grants 2242024k30037, and 2242024k30038; and in part by Nanjing Science and Technology Innovation Project for Overseas Graduates.

References

Bao, X., Chen, L., Lopes, A. M., et al. (2023). Hybrid deep neural network with dimension attention for state-of-health estimation of lithium-ion batteries. *Energy*, *278*, Article 127734.

Cai, L., Holdstock, M., Morganti, M. V., et al. (2024). A semi-empirical state of health estimation method for batteries of electric vehicles operating in varying real-world conditions. *IEEE Access*, In press.

Chen, H., Chai, Z., Dogru, O., et al. (2021). Data-driven designs of fault detection systems via neural network-aided learning. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(10), 5694–5705.

Chen, D., Hong, W., & Zhou, X. (2022). Transformer network for remaining useful life prediction of lithium-ion batteries. *IEEE Access*, *10*, 19621–19628.

Chen, H., Jiang, B., & Ding, S. X. (2020). A broad learning aided data-driven framework of fast fault diagnosis for high-speed trains. *IEEE Intelligent Transportation Systems Magazine*, *13*(3), 83–88.

Chen, Q., Liu, Y. B., Ge, M. F., et al. (2022). A novel Bayesian-optimization-based adversarial TCN for RUL prediction of bearings. *IEEE Sensors Journal*, *22*(21), 20968–20977.

Chen, C., Lu, N., Jiang, B., & Wang, C. (2021). A risk-averse remaining useful life estimation for predictive maintenance. *IEEE/CAA Journal of Automatica Sinica*, *8*(2), 412–422.

Chen, C., Shi, J., Shen, M., Feng, L., & Tao, G. (2023). A predictive maintenance strategy using deep learning quantile regression and kernel density estimation for failure prediction. *IEEE Transactions on Instrumentation and Measurement*, *72*, Article 3506512.

Chen, C., Tao, G., Shi, J., Shen, M., & Zhu, Z. H. (2024). A lithium-ion battery degradation prediction model with uncertainty quantification for its predictive maintenance. *IEEE Transactions on Industrial Electronics*, *71*(4), 3650–3659.

Demirci, O., Taskin, S., Schaltz, E., et al. (2024). Review of battery state estimation methods for electric vehicles-part II: SOH estimation. *Journal of Energy Storage*, *96*, Article 112703.

Durmus, F., & Karagol, S. (2024). Lithium-ion battery capacity prediction with GA-optimized CNN, RNN, and BP. *Applied Sciences*, *14*(13), Article 5662.

Fan, X., Yang, X., & Hou, F. (2024). Integrated mixed attention U-net mechanisms with multi-stage division strategy customized for accurate estimation of lithium-ion battery state of health. *Electronics*, *13*(16), Article 3244.

Han, Y., Li, C., Zheng, L., et al. (2023). Remaining useful life prediction of lithium-ion batteries by using a denoising transformer-based neural network. *Energies*, *16*(17), Article 6328.

Han, D., Qi, H., Wang, S. X., et al. (2024). Adaptive stepsize forward–backward pursuit and acoustic emission-based health state assessment of high-speed train bearings. *Structural Health Monitoring*, Article 14759217241271036. http://dx.doi.org/10.1177/14759217241271036.

Li, Y., Luo, L., Zhang, C., & Liu, H. (2023). State of health assessment for lithium-ion batteries using incremental energy analysis and bidirectional long short-term memory. *World Electric Vehicle Journal*, *14*(7), Article 188.

Lin, C., Tang, A., & Wang, W. (2015). A review of SOH estimation methods in lithium-ion batteries for electric vehicle applications. *Energy Procedia*, *75*, 1920–1925.

Lipu, M. H., Hannan, M. A., Hussain, A., et al. (2023). Real-time state of charge estimation of lithium-ion batteries using optimized random forest regression algorithm. *IEEE Transactions on Vehicular Technology*, *8*(1), 639–648.

Liu, Y., Li, J., Zhang, G., et al. (2021). State of charge estimation of lithium-ion batteries based on temporal convolutional network and transfer learning. *IEEE Access*, *9*, 34177–34187.

Liu, Z., Wang, Y., Vaidya, S., et al. (2024). Kan: Kolmogorov-arnold networks. arXiv Preprint arXiv:2404.19756.

Lu, J., Xiong, R., Tian, J., et al. (2022). Battery degradation prediction against uncertain future conditions with recurrent neural network enabled deep learning. *Energy Storage Materials*, *50*, 139–151.

Lyu, D., Liu, E., Chen, H., et al. (2025). Transfer-driven prognosis from battery cells to packs: An application with adaptive differential model decomposition. *Applied Energy*, *377*, Article 124290.

Lyu, D., Liu, E., Zhang, B., et al. (2025). Threshold-varying assessment for prognostics and health management. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *55*(1), 685–698.

Lyu, D., Niu, G., Zhang, B., et al. (2020). Lebesgue-time–space-model-based diagnosis and prognosis for multiple mode systems. *IEEE Transactions on Industrial Electronics*, *68*(2), 1591–1603.

Lyu, D., Zhang, B., Liu, E., Yang, T., & Xiang, J. (2025). Prognosis-enabled battery SOC estimation using a closed-loop approach with consideration of SOH degradation. *Journal of Energy Storage*, *106*, Article 113713.

Lyu, D., Zhang, B., Zio, E., et al. (2024). Battery cumulative lifetime prognostics to bridge laboratory and real-life scenarios. *Cell Reports Physical Science*, *5*(9), Article 102164.

Ma, Y., Shan, C., Gao, J., & Chen, H. (2022). A novel method for state of health estimation of lithium-ion batteries based on improved LSTM and health indicators extraction. *Energy*, *251*, Article 123973.

Ren, L., Dong, J., Wang, X., et al. (2020). A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life. *IEEE Transactions on Industrial Informatics*, *17*(5), 3478–3487.

Shen, H., Zhou, X., Wang, Z., et al. (2022). State of charge estimation for lithium-ion battery using transformer with immersion and invariance adaptive observer. *Journal of Energy Storage*, *45*, Article 103768.

Shi, Z., & Chehade, A. (2021). A dual-LSTM framework combining change point detection and remaining useful life prediction. *Reliability Engineering & System Safety*, *205*, Article 107257.

Singh, P., Chen, C., Tan, C. M., et al. (2019). Semi-empirical capacity fading model for SoH estimation of Li-ion batteries. *Applied Sciences*, *9*(15), 3012.

Song, L., Hu, Q., Shu, W., et al. (2024). An optimized visual measurement method for cell parallelism based on edge-aware dynamic re-weighted U-Net (EADRU-Net). *Signal, Image and Video Processing*, 1–10.

Sun, S., Wang, J., Xiao, Y., Peng, J., & Zhou, X. (2024). Few-shot RUL prediction for engines based on CNN-GRU model. *Scientific Reports*, *14*(1), Article 16041.

Ungurean, L., Micea, M. V., & Cârstoiu, G. (2020). Online state of health prediction method for lithium-ion batteries, based on gated recurrent unit neural networks. *International Journal of Energy Research*, *44*(8), 6767–6777.

Valant, C. J., Wheaton, J. D., Thurston, M. G., et al. (2019). Evaluation of 1D CNN autoencoders for lithium-ion battery condition assessment using synthetic data. In *Proceedings of the annual conference of the PHM Society, vol. 11, no. 1* (pp. 1–11). PHM Society.

Vignesh, S., Che, H. S., Selvaraj, J., et al. (2024). State of Health (SoH) estimation methods for second life lithium-ion battery—Review and challenges. *Applied Energy*, *369*, Article 123542.

Wang, F. K., Amogne, Z. E., Chou, J. H., & Tseng, C. (2022). Online remaining useful life prediction of lithium-ion batteries using bidirectional long short-term memory with attention mechanism. *Energy*, *254*, Article 124344.

Wang, X., Jiang, H., Mu, M., et al. (2025). A dynamic collaborative adversarial domain adaptation network for unsupervised rotating machinery fault diagnosis. *Reliability Engineering & System Safety*, *255*, Article 110662.

Wen, P., Ye, Z. S., Li, Y., et al. (2024). Physics-informed neural networks for prognostics and health management of lithium-ion batteries. *IEEE Transactions on Vehicular Technology*, *9*(1), 2276–2289.

Wu, Z., Pan, S., Long, G., et al. (2020). Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 753–763). ACM.

Xiang, X., Li, X., Zhang, Y., et al. (2024). A short-term forecasting method for photovoltaic power generation based on the TCN-ECANet-GRU hybrid model. *Scientific Reports*, *14*(1), 6744.

Ye, Y., Wang, J., Yang, J., et al. (2025). Adaptive MAGNN-TCN: An innovative approach for bearings remaining useful life prediction. *IEEE Sensors Journal*, *25*(4), 7467–7481.

Zhang, M., Liu, Y., Li, D., et al. (2023). Electrochemical impedance spectroscopy: A new chapter in the fast and accurate estimation of the state of health for lithium-ion batteries. *Energies*, *16*, Article 1599.

Zhang, L., Wang, B., Yuan, X., & Liang, P. (2022). Remaining useful life prediction via improved CNN, GRU, and residual attention mechanism with soft thresholding. *IEEE Sensors Journal*, *22*(15), 15178–15190.

Zhao, Y., Jing, S., Wu, H., Li, H., & Todoh, M. (2024). E-TRGAN: A novel transformer generative adversarial network for high-density surface electromyography signal reconstruction. *IEEE Transactions on Instrumentation and Measurement*, *73*, Article 4011013.

Zhao, Y., Liu, Z., Yu, J., Jing, S., Li, H., & López, M. B. (2025). HD-sEMG-CORE: An open-source hybrid network algorithm for efficient compression and accurate restoration of high-density surface electromyography signals. *IEEE Sensors Journal*, *25*(3), 5478–5490.

Zhao, J., & Wang, Z. (2024). Specialized convolutional transformer networks for estimating battery health via transfer learning. *Energy Storage Materials*, *71*, Article 103668.