A Novel Supercapacitor Degradation Prediction Using a 1D Convolutional Neural Network and Improved Informer Model

Hao Zhang, Zhenxiao Yi, Le Kang, Yi Zhang, and Kai Wang

***Abstract*—Safety and reliability are crucial for the next-generation supercapacitors used in energy storage systems, while accurate prediction of the degradation trajectory and remaining useful life (RUL) is essential for analyzing degradation and evaluating performance in energy storage systems. This study proposes a novel data processing and improved one-dimensional convolutional neural network (1D CNN)-informer framework for robust RUL prediction. In data preprocessing, all data from two structures are adjusted to a unified format, and cross-entropy loss is used to couple the 1D CNN and informer. Then, the minimum-maximum feature scaling method is used for normalization to accelerate the training process in reaching the minimum cost function. A relative position encoding algorithm is introduced to improve the Informer model, enabling it to better learn the sequence relationships between data and effectively reduce prediction variability. Supercapacitor data in different working conditions are used to validate the proposed method. Compared with other existing methods, the maximum root mean square error is reduced by 32.71%, the mean absolute error is reduced by 28.50%, and R^2 is increased by 4.79%. The strategy considers the complementarity between two single models, which can extract features and enrich local details, as well as enhance the model's global perception ability. The experimental results demonstrate that the proposed model achieves high-precision and robust RUL prediction, thereby promoting the industrial application of supercapacitors.**

***Index Terms*—Supercapacitors, remaining useful life, convolution neural network, transformer, informer.**

I. Introduction

Supercapacitors are efficient, practical, and environmentally friendly energy storage devices, and have been widely applied. Compared to traditional batteries, supercapacitors offer advantages such as high power density, fast response times, wide temperature ranges, and long lifetimes [1]–[3]. Currently, supercapacitors are extensively used in high-power pulse applications such as electromagnetic launchers, laser weapons, and wind power frequency regulation [4]–[6]. They are also employed in grid frequency regulation, rail transportation, buses, and port energy recovery systems [7]–[10]. In hybrid energy storage systems, the integration of supercapacitors can increase the voltage and current of the overall energy storage system, improving its energy utilization efficiency [11]–[13].

Because of varying structural parameters and operating conditions, supercapacitors can exhibit significant differences in their remaining useful life (RUL), which refers to the remaining charge-discharge cycles from the current time until the end of life (EOL). This can impact the reliability, safety, and aging process of the entire energy storage system [14], [15]. The capacity loss of supercapacitors is widely accepted as a degradation indicator for RUL estimation, reflecting the state of health (SOH) of the supercapacitors, which is crucial for device durability [16]–[20]. Typically, when the capacity of a supercapacitor decreases below a certain threshold, known as the EOL, it can no longer meet the requirements and needs to be replaced promptly [21]–[25]. Therefore, accurately predicting the RUL of supercapacitors has significant implications for design, management, and maintenance. However, most existing research focuses on batteries [26]–[33], with limited studies on supercapacitors. Because of substantial differences in the internal structures and principles between batteries and supercapacitors, the existing methods cannot be readily applied to supercapacitors, necessitating the development of new approaches to address these issues.

The current methods for predicting the RUL of supercapacitors primarily include end-to-end and trajec-

Received: September 22, 2023
Accepted: December 10, 2023
Published Online: July 1, 2024
Hao Zhang and Le Kang are with the College of Materials Science and Engineering, Xi'an University of Science and Technology, Xi'an 710054, China (e-mail: hao0204@xauat.edu.cn; kangle20140805@126.com).
Zhenxiao Yi and Kai Wang (corresponding author) are with the School of Electrical Engineering, Qingdao University, Qingdao 266071, China (e-mail: yzx@qdu.edu.cn; wkwj888@163.com).
Yi Zhang is with the School of Energy Sciences and Engineering, Nanjing Tech University, Nanjing 211816, China (e-mail: zhangy@njtech.edu.cn).
DOI: 10.23919/PCMP.2023.000167

tory prediction methods [34]. End-to-end prediction directly maps relevant features to the lifespan of the supercapacitor to predict its RUL. For example, reference [35] maps the collected aging conditions, as well as capacitance and resistance thresholds, to RUL. The particle filter (PF) algorithm is then used to achieve accurate prediction of supercapacitor RUL with high precision. A capacity degradation model is established by capturing in-situ parameters corresponding to the RUL, while multi-scale extended Kalman filters and Gauss-Hermite particle filters are used to update the in-situ parameters, ultimately achieving the prediction of RUL [36]. Reference [37] extracts several features related to RUL from data such as current, voltage, and temperature obtained from cycling experiments, while regularized regression is used to establish the relationship between these features and RUL, achieving prediction errors below 10%. Additionally, methods based on deep learning for predicting the RUL of supercapacitors can establish high-dimensional mappings without the need for manual feature extraction. For example, a convolutional neural network (CNN) can directly learn the relationship between historical data and RUL. Reference [38] proposes an end-to-end RUL prediction method based on a CNN. It establishes the mapping between charge-discharge data and corresponding RUL, and can directly learn aging features from raw data with high prediction accuracy. However, end-to-end prediction methods only provide scalar RUL values and lack a comprehensive characterization of the capacity degradation trajectory of supercapacitors.

The trajectory prediction method involves extrapolating the capacity trajectory to EOL to determine the RUL of supercapacitors. The degradation trajectory of capacity can be seen as a time series. Recurrent neural networks (RNNs), because of their structure with hidden neurons that facilitate recurrent connections and their ability to extract and update temporal data correlations, have been considered an effective means of handling time series data. However, simple RNNs suffer from issues such as vanishing or exploding gradients, which limit their practicality in terms of time dependency. To address this problem, long short-term memory (LSTM) neural networks introduce additional interactions to each module (or unit), thus removing this limitation. For instance, reference [39] predicts the capacity degradation trajectory of supercapacitors using LSTM neural networks, demonstrating high prediction accuracy and robustness. Reference [40] proposes to combine fully connected layers (FCL) with LSTM to estimate the SOH, resulting in accurate and efficient estimates. Many researchers have achieved better results by combining LSTM networks with other methods, leveraging the advantages of multiple algorithmic blends to improve estimation accuracy [41]. Reference [42] proposes an improved robust multi-timescale singular filtering-Gaussian process regression-long short-term memory (SF-GPR-LSTM) modeling method for residual capacity estimation. An optimized multi-task training strategy is constructed to achieve a fine mathematical dynamic representation of the physical carrier transport mapping relationship. The proposed model effectively achieves carrier transport collaborative optimization and is of great significance for estimating the remaining capacity of batteries over the entire life cycle at extremely low temperatures. Reference [34] proposes a method combining the Harris's hawks optimization (HHO) algorithm and a long short-term memory (LSTM) neural network to predict the RUL of supercapacitors. This approach achieves higher prediction accuracy than traditional LSTM models. The method also enhances system stability and reliability. Reference [43] proposes an improved anti-noise adaptive long short-term memory (ANA-LSTM) neural network with high robustness feature extraction and optimal parameter representation for accurate RUL prediction based on an improved dual closed-loop observation modeling strategy. The strategy considers changing current rates, environmental temperature, and other influencing parameters, while the established multi-feature collaborative model realizes multi-scale parameter optimization and robust RUL prediction. This predictive method not only provides a scalar RUL but also has the advantage of predicting future trajectories since capacity trajectories contain more practical and useful information. Therefore, it offers more comprehensive support for the overall management of supercapacitors. However, this approach requires a large amount of historical data from supercapacitors for training.

One-dimensional convolutional neural network (1D CNN) is capable of automatically extracting features based on the characteristics of sequential data, and therefore it can capture more details without the need for manual feature engineering, making it a current research hotspot. However, when dealing with long sequential data, 1D CNN may overlook the dependencies between sequence information. Therefore, it is generally used in conjunction with other neural networks to predict the RUL. Reference [44] uses a hybrid model combining CNN and LSTM to predict the RUL of batteries, achieving a high prediction accuracy. In [45], a variant attention-based spatiotemporal LSTM (AST-LSTM) neural network is proposed to determine both new and old data simultaneously while actively tracking the state of the battery through fixed connections. Experimental results demonstrate that this method predicts well. In addition, a CNN-ASTLSTM model, incorporating a convolutional neural network, is proposed. This model automatically extracts hierarchical features during battery degradation, thus avoiding the need for manual intervention and minimizing its impact on prediction accuracy [46].

However, the previously mentioned methods suffer

from limited scalability in modeling long sequences and time-consuming training. When it comes to very long sequential data like supercapacitors, learning long-term dependencies becomes challenging. This is because the increased length of paths that forward and backward signals must traverse affect the model's ability to learn long dependencies, resulting in reduced efficiency in capturing distant dependencies.

Therefore, an attention mechanism has been proposed to address the limitations of modeling long sequences. The main idea is to change the focus on information, allowing the model to pay more attention to crucial details. This enhances the model's performance and generalization capabilities [47]. However, because of the sequential nature of the processing, where each element needs to be processed one by one and the processing of the tth step can only start after the completion of the $(t-1)$th step, it is time-consuming with inefficient computation when dealing with large datasets [48].

To address the issues, the transformer model eliminates the sequential dependencies present in RNNs by introducing an attention mechanism [49]. The core component is self-attention, which enables the model to capture global dependencies between input and output using an encoder-decoder architecture. Additionally, it achieves high efficiency and accuracy through highly parallelizable computations. The output at each position can simultaneously consider information from all positions in the input, enabling better capturing of long-range dependencies in the input sequence and the relationships between different features appearing at different positions. It can also learn varying degrees of attention at different positions, allowing for more flexible and precise modeling [50]–[53]. However, the transformer model has limitations such as high time complexity, large memory usage, and sudden drops in prediction accuracy, making it unsuitable for direct application in time series forecasting. To address these challenges, a modified variant called informer is proposed in [54] based on the transformer architecture. Informer has been tested on four large-scale datasets and has shown excellent performance, providing an alternative solution for time series forecasting problems. However, although Informer is an improved and novel forecasting model, there still exist problems where local feature details can be easily overlooked, leading to a reduction in discernibility within a limited number of timestamps [55].

In this study, an innovative 1D CNN-informer neural network structure is proposed. By combining local features based on 1D CNN with global variables based on informer, the degradation trajectories and RUL of supercapacitors in different operating conditions can be accurately predicted. The main contributions of this research are:

1) 1D CNN possesses powerful feature extraction capabilities, excelling in collecting local variables through convolutional operations and preserving all local cues as feature maps, while Informer integrates global properties among compressed enhanced embeddings using its unique attention mechanism. Therefore the proposed 1D CNN-informer model retains the structures and generalization advantages of both 1D CNN and Informer, enabling high-precision and rapid prediction of supercapacitor degradation trajectories and RUL. The experimental results demonstrate the enormous potential of this approach.

2) Improvements have been made to the transformer model to obtain the informer model. Also, a relative position encoding method is introduced to optimize the Informer model.

3) All data from the 1D CNN and informer structures have been adjusted to a unified format, and cross-entropy loss is used to couple the style variables between them.

4) The proposed hybrid model effectively leverages information from the observed data. This is then applied to supercapacitors in different operating conditions. It addresses the limited positional information-capturing ability of a single informer network. This method exhibits strong generalizability and has reference value for estimating the SOH of lithium-ion batteries.

The remaining sections of this paper are organized as follows. Section II introduces the basic principles and overall prediction framework of 1D CNN, transformer, informer, and 1D CNN-informer. In Section III, the aging test platform, factors influencing the aging of supercapacitors, and the dataset of supercapacitors collected in different operating conditions for testing are presented. In Section IV, the proposed 1D CNN-Informer and three other mainstream methods are trained and analyzed based on online testing. Section V concludes the study and provides prospects for future research.

II. Methodology

This section introduces the working principles and usage methods of each component in the proposed hybrid model, including the 1D CNN and the improved informer network.

A. 1D Convolutional Neural Network

A CNN has outstanding capability in handling multidimensional data and has gained extensive attention in fields such as image recognition, natural language processing, and time series classification. A CNN consists of an input layer, an output layer, and hidden layers, providing powerful feature extraction and non-linear relationship modeling abilities. 1D CNN mainly includes convolutional layers, pooling layers, and FCL, with the convolutional layer being the core component. The structural diagram of a 1D CNN is shown in Fig. 1. The convolutional layer captures local attributes from

higher-level inputs and passes all information to lower levels to obtain more complex features. The pooling layer reduces dimensionality while preserving relevant features, and the fully connected layer yields the prediction results. The activation function used is the rectified linear unit (ReLU), which mitigates the vanishing gradient problem and improves the trainability of the network. The key factors contributing to the success of a 1D CNN include local connections, weight sharing, pooling, and the use of multiple layers. This article uses 1D CNN to capture the spatial characteristics of supercapacitor variables. Add convolutional layers so that the input information undergoes convolutional operations and activation functions before flowing to the next layer h_k:

$$h_k = \sigma_{\text{cnn}}(W^*_{\text{cnn}} x_k + b_{\text{cnn}}) \tag{1}$$

where σ_{cnn} represents the sigmoid activation function; $*$ denotes the discrete convolution between the input signal x_k and the filter weight W_{cnn}; and b_{cnn} is a bias parameter which shall be learned during training.

Finally, all neurons in each layer are connected to every neuron in the output layer through the fully connected layer [56], [57].

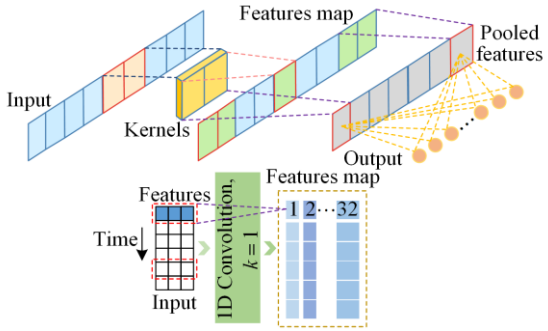


Fig. 1. Internal structure and feature extraction schematic diagram of a 1D CNN.

B. Transformer

The transformer model was proposed by Vaswani et al. in 2017 [58]. Its core idea is to model the dependency relationships between input sequences based entirely on a self-attention mechanism. It can solve the problem of parallel computation in traditional RNN when dealing with sequential data, enabling the model to process the entire input sequence simultaneously and greatly improving computational efficiency. The transformer model is an overall architecture that includes stacked self-attention and fully connected layers for both the encoder and decoder. Additionally, it relies on a self-attention mechanism composed of scaled dot-product and multi-head attention. This fundamentally changes the implementation of attention mechanisms.

The main components of the Transformer model include a self-attention mechanism, encoder-decoder structure, a multi-head attention mechanism, and a feed-forward neural network.

In the encoder, the input sequence $X = (x_1, x_2, \cdots, x_n)$ is mapped to a given continuous representation sequence $Z = (z_1, z_2, \cdots, z_n)$. The previously generated symbols are used as additional inputs, and the output sequence $Y = (y_1, y_2, \cdots, y_n)$ is generated one element at a time in the decoder. The structure of the transformer model is shown in Fig. 2.

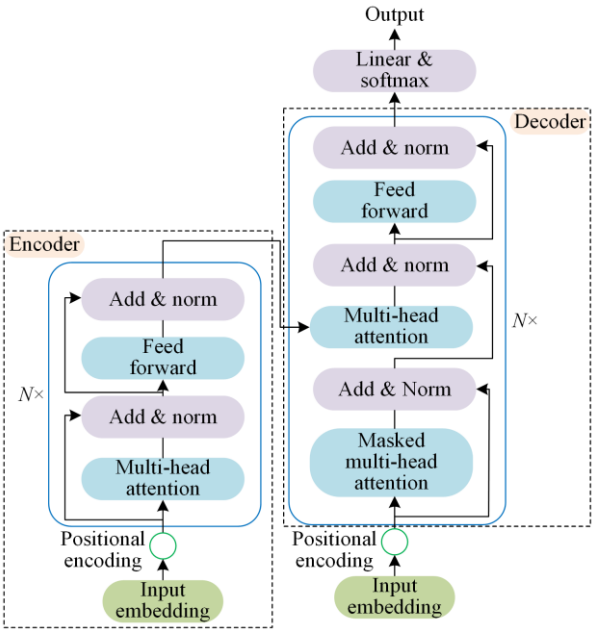


Fig. 2. Diagram illustrating the structure of the transformer model.

The encoder consists of a stack of n identical layers, while each layer includes a multi-head self-attention mechanism and a fully connected feed-forward network. The decoder also comprises a stack of identical layers, with an additional sub-layer that performs output through a multi-head attention mechanism. Both the encoder and decoder layers use residual connections followed by layer normalization. The masking in the self-attention mechanism prevents early exposure to future information during the decoding process. The calculation of the continuous representation sequence is defined by the standard self-attention mechanism as:

$$Z = \text{Attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\text{T}}}{\sqrt{d}}\right)\boldsymbol{V} \tag{2}$$

where $\boldsymbol{Q}$, $\boldsymbol{K}$, $\boldsymbol{V}$ are the query matrix, the key matrix, and the value matrix, respectively; $\sqrt{d}$ is the dimension size of $\boldsymbol{K}$.

The Transformer model addresses the limitations of traditional RNN, such as poor parallel computing capability and weak modeling of long-term dependency relationships. It also has the advantage of capturing global information. However, there are challenges including high computational complexity, and significant time and resource requirements for training and prediction. Additionally, the position information of input sequences is

encoded using positional encoding, which may not accurately represent the position information for long sequences.

C. Informer

The informer model proposed in [59] is an optimized version of the transformer model, demonstrating superior performance in sequence prediction tasks. Similar to the transformer model, the informer model consists of two main components: the encoder and the decoder. To address the limitations of the transformer model, the informer model introduces improvements in following three ways.

1) ProbSparse Self-attention

ProbSparse self-attention is proposed. This can select the most important queries, significantly reducing the computational and spatial complexity.

The formula for ProbSparse self-attention can be derived as:

$$A(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{Softmax}(\frac{\bar{\boldsymbol{Q}}\boldsymbol{K}^{\text{T}}}{\sqrt{d}})\boldsymbol{V} \tag{3}$$

where $\bar{\boldsymbol{Q}}$ is a sparse matrix of the same size of $\boldsymbol{Q}$.

This enables ProbSparse self-attention to compute dot products only for each query-key pair. This approach effectively solves the quadratic computational complexity issue of the self-attention mechanism in Transformers. For the detailed derivation of the formulas, please refer to [54].

2) A Self-attention Distillation Operation

A self-attention distillation operation is introduced, using convolution and pooling operations to reduce dimensionality and network parameters, thus greatly reducing the overall spatial complexity. This overcomes the barrier of being unable to stack inputs that are too long. The distillation process from the *j*th layer to the (*j*+1)th layer is as follows:

$$X_{j+1}^{t} = \text{MaxPool}(\text{ELU}(\text{Cov1d}([X_j^t]_{\text{AB}}))) \tag{4}$$

where $[\cdot]_{\text{AB}}$ represents the attention block, it contains the multi-head ProbSparse self-attention and the essential operations; Conv1d(·) performs an 1D convolutional filters (the kernel width = 3) on time dimension with the ELU(·) activation function; and X_{j+1}^{t} represents down sampling.

This includes multi-head probsparse self-attention along with the key operations in the attention block. This will address the issue of high memory usage in the Transformer model.

3) A Generation Style Decoder

A generation style decoder is proposed. This obtains all prediction results in one step instead of the traditional step-by-step approach. This improves the speed of long sequence prediction, as only a single forward step is needed to generate the entire output sequence. It also avoids error propagation during the prediction phase. The input to the decoder is represented as:

$$X_{\text{de}}^{t} = \text{Concat}(X_{\text{token}}^{t}, X_{\text{O}}^{t}) \in \mathbb{R}^{(L_{\text{token}}+L_y)\times d_{\text{model}}} \tag{5}$$

where $X_{\text{token}}^{t} \in \mathbb{R}^{L_{\text{token}}\times d_{\text{model}}}$ is the start token; and $X_{\text{O}}^{t} \in \mathbb{R}^{L_y\times d_{\text{model}}}$ is a placeholder for the target sequence that sets the scalar as 0; L_{token} is the length of the input; L_{y} is the length of the output sequence; and d_{model} is the dimension of the model input. Masked multi-head attention is applied in the ProbSparse self-attention computing by setting masked dot products to $-\infty$. It prevents each position from attending to coming positions, which avoids auto-regressive.

Thus, the efficiency problem of predicting long-term output, which exists in transformer, is solved by calculating the output by multiple outputs together.

D. Relative Position Encoding Method

The transformer's position coding formula is given as:

$$\text{PE}(p,2i) = \sin\left(\frac{p}{10000^{2i/d_{\text{model}}}}\right) \tag{6}$$

$$\text{PE}(p,2i+1) = \cos\left(\frac{p}{10000^{2i/d_{\text{model}}}}\right) \tag{7}$$

where p is the position of the input data and i is the dimension of the data. The calculation of the attention scores for query, key, and value are:

$$\begin{cases} q_i = (W_q(\boldsymbol{E}_{x_i} + P_i)) \\ k_j = (W_k(\boldsymbol{E}_{x_j} + P_j)) \\ v_j = (W_v(\boldsymbol{E}_{x_j} + P_j)) \end{cases} \tag{8}$$

where W_q, W_k, and W_v are the query, key, and value parameters added to each head of the multi-head attention, respectively; $\boldsymbol{E}_{x_i}$ and $\boldsymbol{E}_{x_j}$ are the respective data embedding vectors for x_i and x_j; while P_i and P_j are the position embeddings for the *i*th and *j*th positions, respectively.

The calculation formula of the attention score $S_{i,j}$ of the *i*th element and *j*th element is:

$$S_{i,j} = q_i\boldsymbol{k}_j^{\text{T}} = (W_q(\boldsymbol{E}_{x_i} + \boldsymbol{P}_i))(W_k(\boldsymbol{E}_{x_j} + \boldsymbol{P}_j))^{\text{T}} \tag{9}$$

Ensure that the output probability value $a_{i,j}$ is between 0 and 1:

$$\begin{cases} a_{i,j} = \text{softmax}(S_{i,j}) \\ A_{i,j} = \sum_j S_{i,j}v_j \end{cases} \tag{10}$$

Factoring (9) yields:

$$\begin{aligned} S_{i,j} = {} & \boldsymbol{E}_{x_i}W_q\boldsymbol{W}_k^{\text{T}}\boldsymbol{E}_{x_j}^{\text{T}} + \boldsymbol{E}_{x_i}W_q\boldsymbol{W}_k^{\text{T}}\boldsymbol{P}_j^{\text{T}} + \\ & P_iW_q\boldsymbol{W}_k^{\text{T}}\boldsymbol{E}_{x_j}^{\text{T}} + P_iW_q\boldsymbol{W}_k^{\text{T}}\boldsymbol{P}_j^{\text{T}} \end{aligned} \tag{11}$$

It has been discovered experimentally that the relative positional information gets lost when an unknown linear transformation is applied. Therefore, to address

this issue, relative positional encoding is introduced. This allows the model to adapt to sequence lengths that have not been encountered before. Consequently, the transformation of (11) becomes:

$$S_{i,j} = \boldsymbol{E}_{x_i} W_q \boldsymbol{W}_{k,E}^{\mathrm{T}} \boldsymbol{E}_{x_j}^{\mathrm{T}} + \boldsymbol{E}_{x_i} W_q \boldsymbol{W}_{k,R}^{\mathrm{T}} \boldsymbol{R}_{i-j}^{\mathrm{T}} + u\boldsymbol{W}_{k,E}^{\mathrm{T}} \boldsymbol{E}_{x_j}^{\mathrm{T}} + v\boldsymbol{W}_{k,R}^{\mathrm{T}} \boldsymbol{R}_{i-j}^{\mathrm{T}} \tag{12}$$

where $\boldsymbol{P}_j^{\mathrm{T}}$ is replaced by $\boldsymbol{R}_{i-j}^{\mathrm{T}}$ representing the relative position; $P_i W_q$ is replaced by trainable parameters $u \in R^d$ in the third term and $v \in R^d$ in the fourth term, which means the attentive bias should remain unchanged, also, the attention bias should remain unchanged after the improvement; R^d represents a d-dimensional real number; finally, the two weight matrices $\boldsymbol{W}_{k,E}^{\mathrm{T}}$ and $\boldsymbol{W}_{k,R}^{\mathrm{T}}$ are separated to generate content-based and position-based key vectors, respectively.

The structure of informer is shown in Fig. 3.

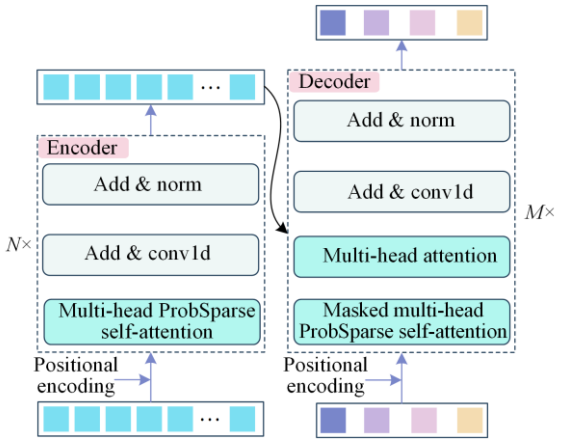


Fig. 3. Structural diagram of informer.

E. Overview of 1D CNN-informer Structure

Time series can be viewed as a set of observations that occur within specific time intervals. Additionally, each randomly observed variable has a unique distribution. Local properties and global representations are two complementary and crucial aspects. Currently, numerous studies have extensively examined techniques for handling time series data. In each sample, local features are non-stationary and exhibit stochastic processes over time. On the other hand, global features encode time-invariant characteristics over long distances. Based on this, we propose a concurrent neural network model called 1D CNN-informer, in which different dimensions of CNNs are used to process different data for improved performance. As 1D CNN performs exceptionally well in handling one-dimensional time series data while experimental data from supercapacitors consist of one-dimensional time series, using 1D CNN to extract and learn relevant features from input data is a suitable choice. Through its convolutional operations, it collects local variables in a hierarchical manner while preserving all local clues as feature maps. Informer, on the other hand, aggregates global attributes from compressed enhanced embeddings using its self-attention mechanism [60], [61]. Therefore, the proposed network framework fully exploits both local and global characteristics. The overall architecture diagram and the algorithm are shown in Fig. 4 and Table I, respectively, while the numerical changes of each layer in the 1D CNN-informer model is shown in Table II.

TABLE I
1D CNN-INFORMER ALGORITHM

Algorithm1. 1D CNN-informer algorithm

$\hat{y}_{ij} = \mathrm{CNN\_Informer}([x_{ij,t}, x_{ij,t+1}, \cdots, x_{ij,t+n+1}] \mid y_{ij})$

where $\hat{y}_{ij}$ is the predicted capacity of the jth supercapacitor in ith cycle; $x_{ij,t}$ is the real value at sampling time t and its length is m; n is the length of the sliding window.

Input: $\boldsymbol{X} = \begin{bmatrix} X_{1j,t} & X_{1,j,t+1} & \cdots & X_{1,j,t+n+1} & y_{1j} \\ X_{2j,t} & X_{2,j,t+1} & \cdots & X_{2,j,t+n+1} & y_{2j} \\ \vdots & & & & \vdots \\ X_{mj,t} & X_{m,j,t+1} & \cdots & X_{m,j,t+n+1} & y_{2j} \end{bmatrix}$

Output: Capacity (F)
Step 1: Set filters number and filter size
Step 2: function TRAIN (Y)
Step 3: for number of training iterations do
Step 4: model = send Y to proposed LSTM, CNN-transformer, CNN-informer for training
Step 5: calculate loss function
Step 6: end for
Step 8: return model
Step 9: end function
Step 10: split $\boldsymbol{X}$ into training (70%) and testing datasets (30%)
Step 11: calculate average estimation loss value to determine various hyperparameters
Step 12: save the best model by weight optimization method and optimizers is Adam
Step 13: load the optimal weight of the best model
Step 14: $\hat{y}$ = model.predict (testing)
Step 15: calculate RUL
Step 16: evaluate and compare the estimation results by MAE,MAPE, RMSE, R^2

TABLE II
THE NUMERICAL CHANGES OF EACH LAYER IN THE 1D CNN-INFORMER MODEL

Layer type	Activation layer	Output shape
Input		(none, 640, 7)
Conv1D	ReLU	(none, 640, 32)
Max-pooling-1D		(none, 80, 32)
Multi-Head Prob-Sparse Self-attention		(none, 80, 32)
Add_1		(none, 80, 32)
Conv1D		(none, 80, 32)
Layer-normalization_1		(none, 80, 32)
Dense_1		(none, 80, 32)
Dropout_1	GELU	(none, 80, 32)
Add_2		(none, 80, 32)
Conv1D		(none, 80, 32)
Layer-normalization_2		(none, 80, 32)
Flatten		(none, 2560)
Dropout_2		(none, 2560)
Dense_2		(none, 1)
Dropout_2		(none, 1)

Given the unique complementarity between the 1D CNN and informer feature styles, in 1D CNN-informer, local features are gradually extracted from the 1D CNN to enrich local details and then the Informer enhances the model's perception ability to the global level. Position encoding is used to describe the position of data points in a sequence, and assigns a unique representation between 0 and 1 to each position. The encoder is used to generate encodings for the correlated parts of the input information, and the input of the last encoder layer is used as the input for the decoder. The decoder also receives the output from the previous time step of the decoder to generate the output sequence.

In the CNN layer, ReLU is used as the activation function, and the output is operated by the max pooling layer with 32 filters, kernel size of 8, strides of 1, and a pooling window size of 8. In the informer, dropout is introduced, with a rate of 0.1, to prevent overfitting. The first added layer combines the output information of the max pooling layer and the attention layer, while the second added layer combines the dropped and normalized data from the past. Subsequently, the dimension is reduced through the flattening layer. Nonlinear transformations in the dense layer extract correlations between previously extracted features using the Gaussian error linear unit (GELU) as the activation function, ultimately mapping them to the output. Finally, the Dropout_2 layer connects to the prediction results through a dense layer.

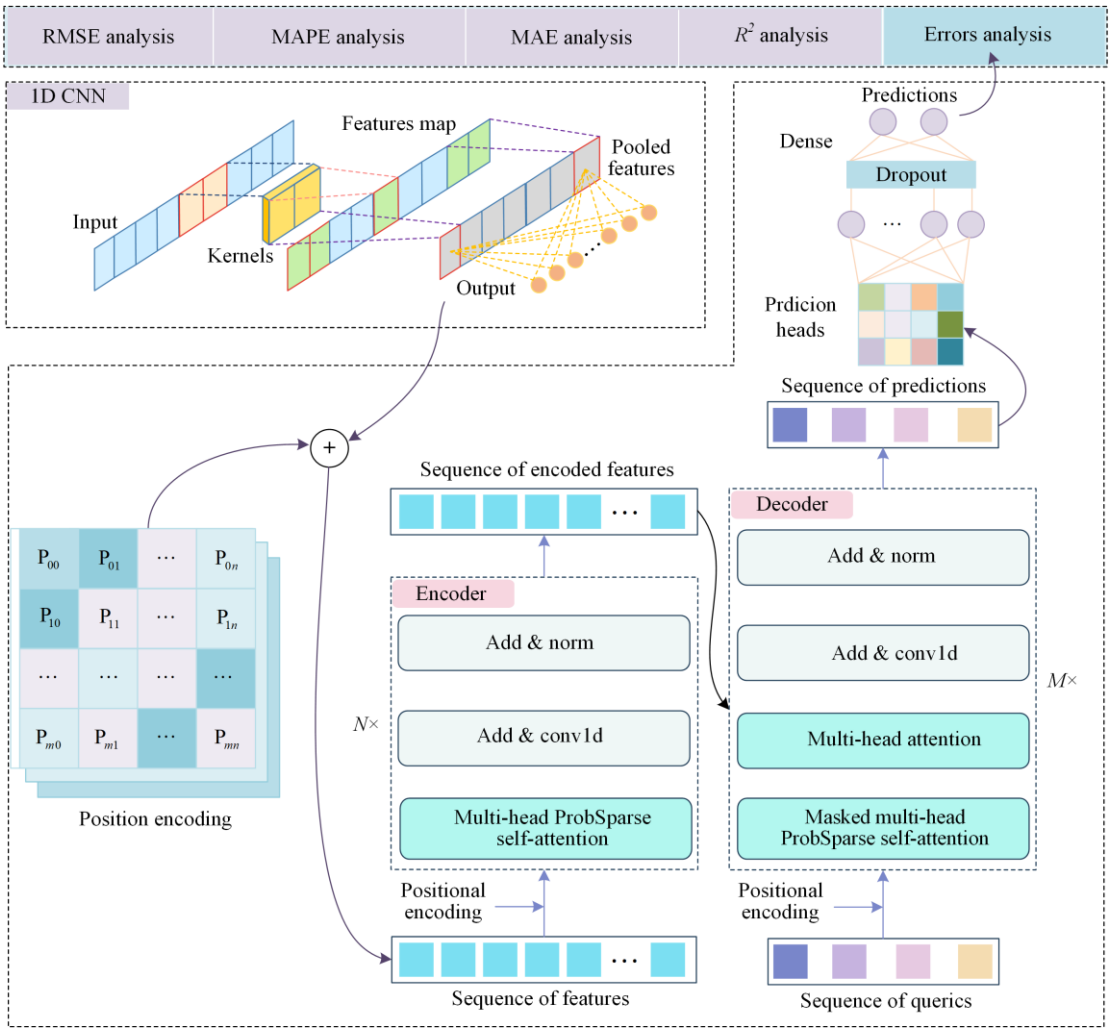


Fig. 4. Overview of 1D CNN-informer structure.

1D CNN enriches local details by progressively extracting local features, while Informer enhances the global perception ability of the model. Position encoding is used to describe the position of data points in the sequence. Since the method of position encoding has a significant impact on the predictive ability of the Informer, the introduction of a relative position encoding algorithm allows the model to better learn the sequence relationship between data, effectively reducing prediction variations. The output of the last encoder layer serves as the input to the decoder in informer, and generates the output sequence by receiving the input from the encoder. To address the issue of inconsistent data capabilities and features between 1D CNN and informer,

the data formats in both structures are unified. Additionally, coupling 1D CNN and informer variables through cross-entropy loss enhances the global perception ability of local variables and the local details of global attributes. Finally, the results of the decoder block are processed and the final prediction is achieved by coordinating the flattening, dropout, and dense layers. The dropout layer effectively prevents overfitting by randomly ignoring a set of neurons, while the output of each neuron in the previous layer is passed to the neurons in the dense layer. Therefore, the combination of CNN and informer networks allows the model to fully utilize a variety of feature information in the input data, leading to comprehensive representation and accurate prediction of the RUL of supercapacitors.

III. Aging Test of Supercapacitors

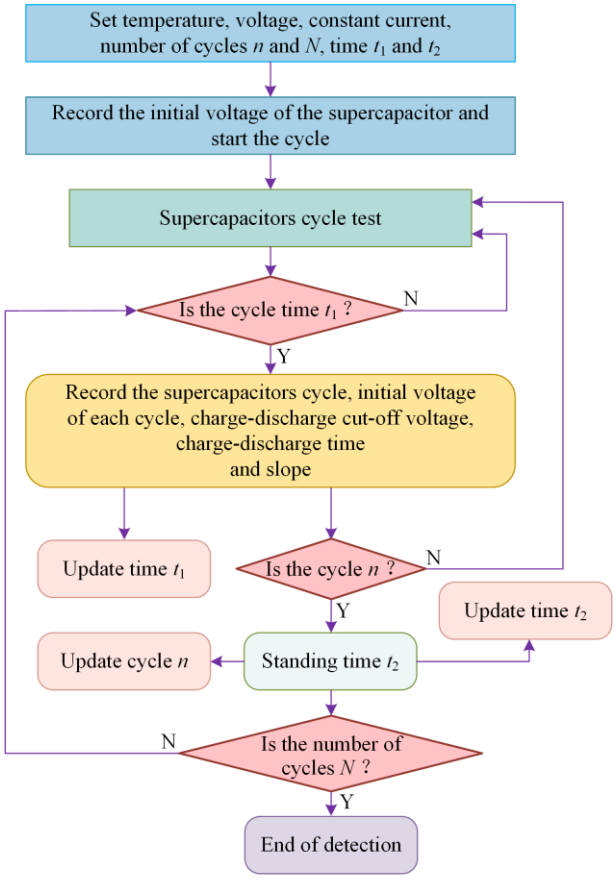
A. Aging Testing Hardware System

The supercapacitor aging state testing system consists of three main components: the testing system, the host computer, and a high-low temperature chamber. Because of the impact of factors such as temperature, voltage, and frequency on supercapacitors, to ensure consistency in parameter measurements, the capacitances of the supercapacitors are measured in the same environment.

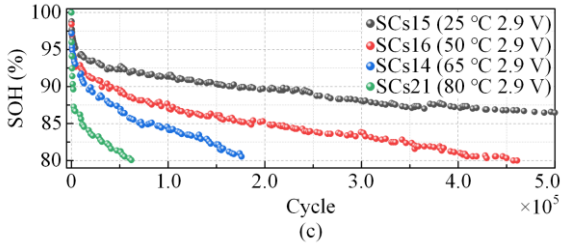
The testing system used is Chroma 17011, which has built-in multiple test modes. It can perform constant current to constant voltage (CC-CV), constant current (CC), and constant power (CP) charging and discharging tests for supercapacitors. It can also perform capacitance testing and DC resistance testing. Each test mode switches steps based on the set time, voltage, current, or power conditions. The collected data includes feedback test steps, status, voltage, current, capacitance, etc. There are multiple flexible sampling methods to choose from, based on time, voltage, current, or capacitance conditions.

The CPU model of the host computer is Intel i7-11700, which is used for data storage and processing. The high-low temperature chamber is used to provide specific temperature environments for the aging state testing of supercapacitors. In this experiment, the TIGSTOR supercapacitor TIG-1W160P010R01 is selected. The capacitance of the supercapacitor is 10 F, the rated voltage is 2.7 V, and the operating temperature range is −40 ℃ to 70 ℃. The supercapacitor is charged with a constant current of 3 A until the voltage reaches 2.7 V and then continues to charge in constant voltage mode. Multiple charge-discharge tests are conducted at different temperatures and voltages, with the discharge depth controlled at 50%. The tests are conducted in the same environment, without considering the influence of vibration, and repeated for hundreds of thousands of cycles. The testing setup is shown in Fig. 5(a), and the cycling test conditions for the supercapacitor are illustrated in Table III.

(a)



(b)



(c)

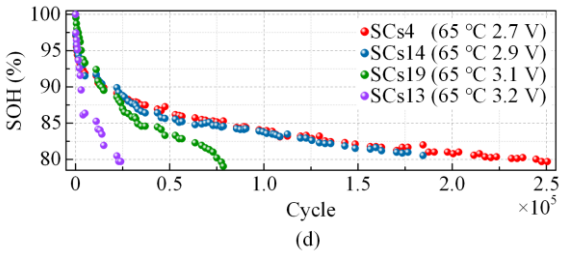


Fig. 5. Aging test and characteristic analysis of supercapacitors. (a) Supercapacitor aging test system (supercapacitor charging/discharging with CC-CV charging protocol at different temperatures and voltages). (b) Schematic diagram of the experimental procedure. (c) Capacity degradation trajectory of supercapacitor with increasing number of cycles at 25 °C, 50 °C, 65 °C, and 80 °C. (d) Capacity degradation trajectory of supercapacitor at 2.7 V, 2.9 V, 3.1 V, and 3.2 V capacitance degradation trend of supercapacitor with the increase of cycle number.

TABLE III
CYCLIC TEST CONDITIONS OF SUPERCAPACITORS

The number of the supercapacitor	Voltage (V)	Temperature (°C)
SCs1	3.2	25
SCs2	3.2	50
SCs3	3.2	65
SCs4	2.7	65
SCs5	3.2	80
SCs6	2.7	80
SCs7	2.9	50
SCs8	3.1	25
SCs9	3.1	50
SCs10	3.1	65
SCs11	2.9	80
SCs12	3.1	80
SCs13	2.9	25
SCs14	2.9	65
SCs15	2.7	25
SCs16	2.7	50

B. Analysis of Aging Characteristics

In this experiment, the supercapacitor is cyclically charged and discharged in different working conditions using the CC-CV charging protocol, and the experimental flow is shown in Fig. 5(b). The aging trend of the supercapacitor's capacitance in different working conditions is analyzed, as shown in Figs. 5(c) and (d). The analysis reveals that the aging rate of supercapacitor is closely related to temperature and voltage, and accelerates with the increase of temperature, beccause the movement of ions in the electrolyte and the various reactions occurring on the electrode surface are affected by temperature. Too high a temperature also affects the normal operation of the supercapacitor.

Although there are differences in supercapacitor aging in different operating conditions, it can be seen that the degradation trend of the supercapacitor's capacitance is similar at different voltages and temperatures. In other words, the aging of the supercapacitor follows a similar pattern of change. Therefore, the method proposed in this paper is applicable for predicting the RUL of supercapacitors in different operating conditions.

IV. RESULTS AND DISCUSSION

In this section, the results of LSTM, 1D CNN-Transformer, and 1D CNN-informer models for predicting the RUL of supercapacitors are presented. The initial parameters of the 1D CNN network include filters of 32, a kernel size of 8, strides of 1, the activation function of "relu", pool size of 8 in the MaxPooling1D layer, and dropout rate of 0.05. The initial learning rate is set to 0.0001, and the learning rate decay factor is set to 0.8. The training time is measured using the time.time function from the Python TIME LIBRARY.

The offline data of the supercapacitor is divided into a training set (70%) and a test set (30%). To demonstrate the generalizability of the models, different datasets in various operating conditions are randomly selected for validation and error analysis. The performance of the models is evaluated using metrics such as root mean square error (RMSE), mean absolute percentage error (MAPE) in percentage, mean absolute error (MAE), and R-squared value (R^2).

A. The Evaluation Criteria

RMSE is used to represent the deviation from the mean. The magnitude of the RMSE value represents the average difference between the predicted and true values, and the smaller the value, the more accurate and more stable the model. The value of RMSE (R_{MSE}) is calculated as:

$$R_{\mathrm{MSE}} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2} \tag{13}$$

where y_n represents the true value and $\hat{y}_n$ represents the predicted value.

MAE can suppress the problem of errors canceling each other, and is the average of the absolute errors between the predicted and true values. It can better reflect the actual situation of the predicted value errors, and the smaller its value, the higher the model accuracy. The value of MAE (M_{AE}) is:

$$M_{\mathrm{AE}} = \frac{1}{N}\sum_{n=1}^{N}\left|y_n - \hat{y}_n\right| \tag{14}$$

The R^2 coefficient of determination is a visual representation of the correlation between the model and the true value, and the higher the value, the more accurate the model. $R^2 = 1$ means that the model correctly predicts the true value and all observations fall on the regression line. $R^2 = 0$ means that the model predicts the true value with poor accuracy, while $R^2 < 0$ means that there is a lack of linear correlation among the data. It is widely used in the field of model prediction evaluation because it does not have a magnitude problem and can also compensate for the defect of not being able to predict because of the presence of several 0 values in the data. The calculation formula is:

$$R^2 = 1 - \frac{\sum_{n=1}^{N}\left(y_n - \hat{y}_n\right)^2}{\sum_{n=1}^{N}\left(y_n - \bar{y}_n\right)^2} \tag{15}$$

where $\bar{y}_n$ represents the arithmetic mean of the dependent variable in the original dataset.

MAPE is from 0 to positive infinity. When it is close to 0, it means that the model's prediction ability is better, while greater than 100% means that the prediction accuracy is poor. The value of MAPE (M_{APE}) is:

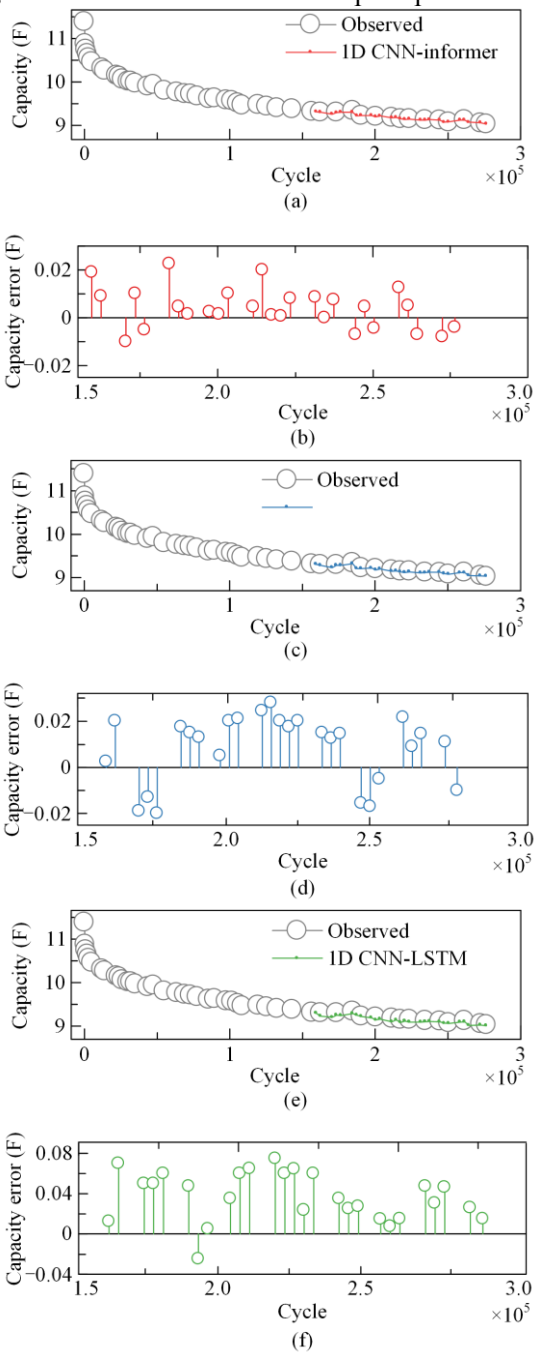
$$M_{\text{APE}} = \frac{1}{N}\sum_{n=1}^{N}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{16}$$

In this paper, R_{MSE}, M_{AE}, M_{APE} and R^2 are used as performance metrics for estimation and prediction accuracy.

B. Experimental Results and Analysis

The LSTM, 1D CNN-LSTM, 1D CNN-transformer, and 1D CNN-informer models are trained and used for capacity prediction using the supercapacitor datasets in different operating conditions. As previously described, the first 70% of the supercapacitor dataset is used for offline training of the models, while the remaining dataset is used for online testing. SCs4 and SCs7 are randomly selected as test data for validating the models. To comprehensively evaluate the robustness and effectiveness of the models, the other three methods, LSTM, 1D CNN-LSTM, and 1D CNN-transformer, predict the RUL of supercapacitors using the same offline training strategy, allowing for a thorough assessment of the model's robustness and effectiveness. The predicted results for SCs4 are shown in Figs. 6 (a)–(h), and the specific performance metrics for the four methods are listed in Table IV. Compared to the other three methods, the 1D CNN-informer model exhibits predictions that are closer to the observed values, indicating higher prediction accuracy. To further evaluate the models, Figs. 6 (i) and (g) display the box plots of prediction errors and the distribution of errors, respectively. Through comprehensive analysis, it can be concluded that the proposed model has smaller prediction errors and better predictive performance. The predicted results for SCs7 are shown in Figs. 7 (a)–(h), and the specific performance metrics for the four methods are listed in Table V.

In different operating conditions, the proposed model also exhibits higher prediction accuracy and is closer to the observed values. To further evaluate the estimation performance of the model, Figs. 7 (i) and (g) respectively present box plots of prediction errors and error distributions. As seen, the proposed model has smaller prediction errors and better predictive performance. The RMSE of the 1D CNN-informer model are 0.0208 and 0.0224 in the two different operating conditions, indicating higher prediction accuracy than the 1D CNN-transformer. In addition, the proposed model demonstrates more pronounced advantages than the other two models. For the LSTM model, its predictive performance is poorer in both operating conditions. This is mainly because 1D CNN is better at capturing local features, which is crucial for large-scale supercapacitor datasets. Therefore the introduction of 1D CNN effectively improves the predictive performance of the model. Additionally, since informer is optimized based on transformer and has parallel computing capabilities compared to LSTM, it has the advantage of improving model training efficiency. Also, LSTM has certain limitations in handling long-time series and needs to address gradient issues, leading to its limitations in predicting large-scale data such as those for supercapacitors.



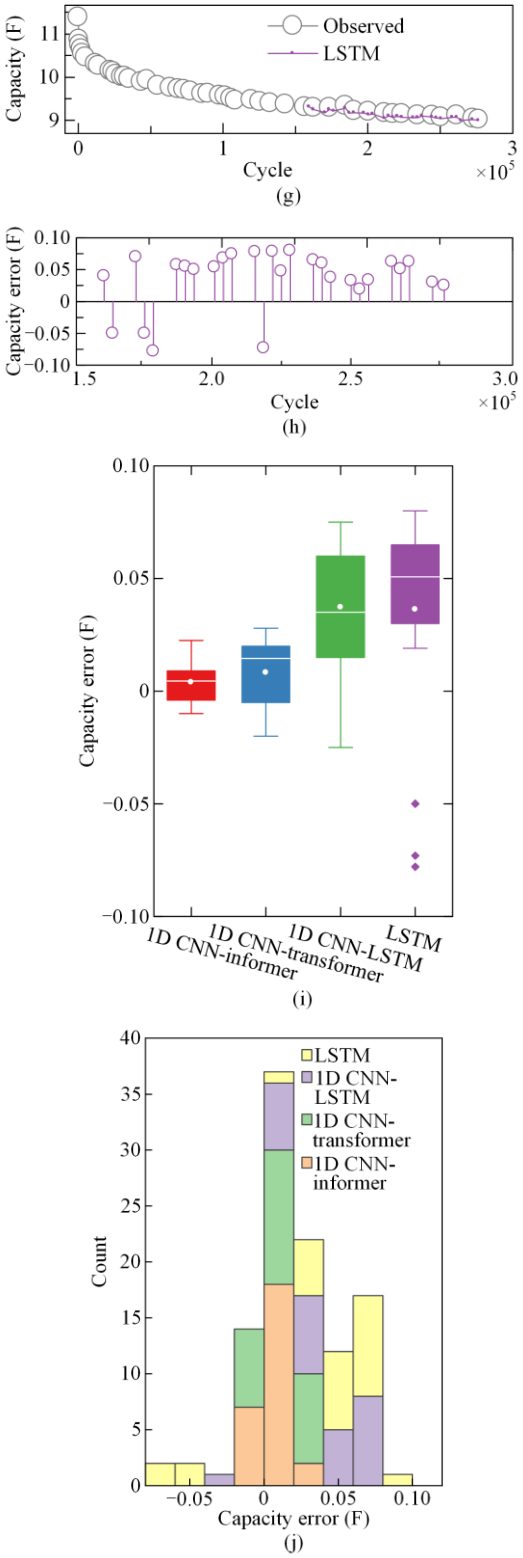
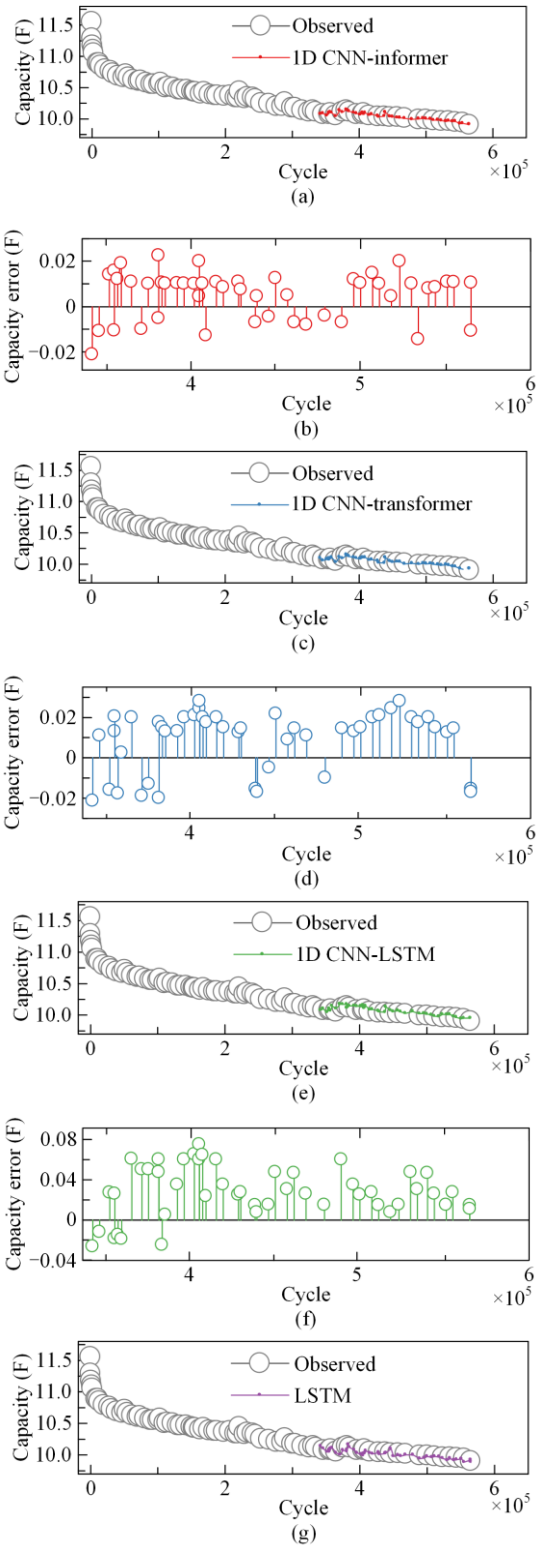
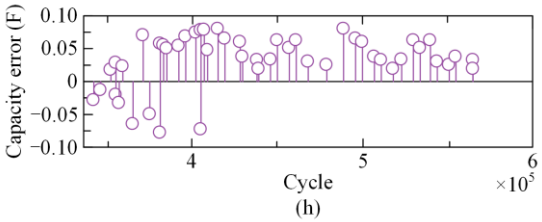


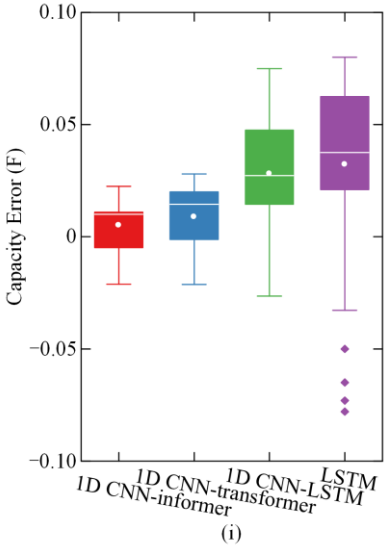
Fig. 6. RUL prediction results and error analysis of SCs4 supercapacitor (2.7 V, 3 A, 65 ℃) with different models. (a) RUL prediction using the 1D CNN-informer model. (b) Prediction error of the 1D CNN-informer. (c) RUL prediction using the 1D CNN-transformer model. (d) Prediction error of the 1D CNN-transformer. (e) RUL prediction using the 1D CNN-LSTM model. (f) Prediction error of the 1D CNN-LSTM. (g) RUL prediction using the LSTM model. (h) Prediction error of the LSTM. (i) Box plot analysis of prediction errors for the four models. (j) Distribution of prediction errors for the four models.

TABLE IV
RUL PREDICTION ACCURACY FOR SCs4 SUPERCAPACITOR

Methods	R_{MSE}	M_{APE} (%)	M_{AE}	R^2
1D CNN-informer	0.0208	0.1878	0.0207	0.9958
1D CNN-transformer	0.0276	0.1987	0.0266	0.9503
1D CNN-LSTM	0.0328	0.2418	0.0314	0.9130
LSTM	0.0391	0.3422	0.0379	0.8841







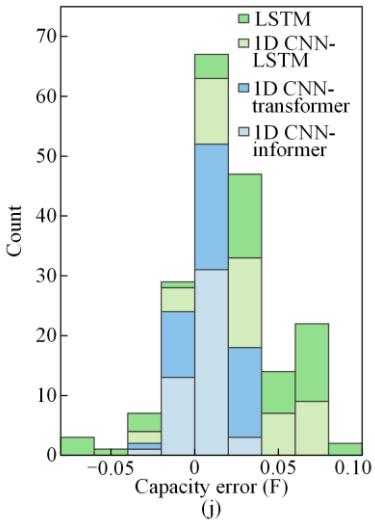


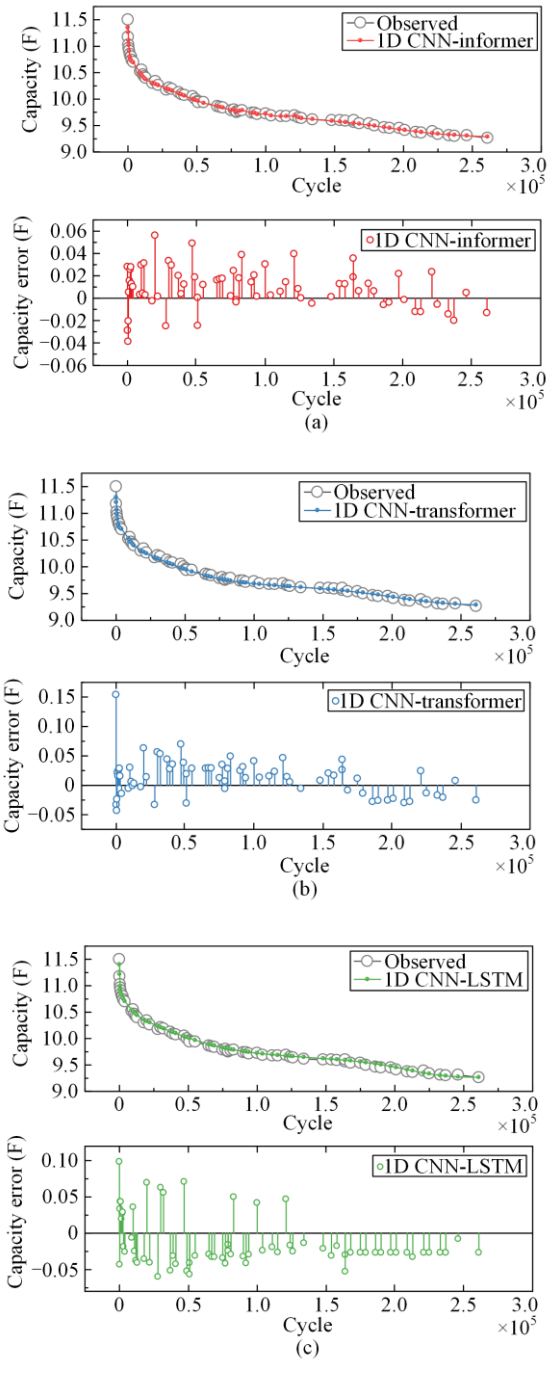
Fig. 7. RUL prediction results and error analysis of SCs7 supercapacitor (2.9 V, 3 A, 50 °C) with different models. (a) RUL prediction using the 1D CNN-informer model. (b) Prediction error of the 1D CNN-informer. (c) RUL prediction using the 1D CNN-transformer model. (d) Prediction error of the 1D CNN-transformer. (e) RUL prediction using the 1D CNN-LSTM model. (f) Prediction error of the 1D CNN-LSTM. (g) RUL prediction using the LSTM model. (h) Prediction error of the LSTM. (i) Box plot analysis of prediction errors for the four models. (j) Distribution of prediction errors for the four models.

In addition, experimental data of untrained SCs14 and SCs15 in different operating conditions are selected as test data to further verify the applicability and generalizability of the model. By calculating the errors and comparing them with the actual measured data, the superiority of the model can be validated.

TABLE V
PREDICTION ACCURACY OF RUL OF SCS7 SUPERCAPACITOR

Methods	R_{MSE}	M_{APE} (%)	M_{AE}	R^2
1D CNN-informer	0.0224	0.1882	0.0227	0.9948
1D CNN-transformer	0.0278	0.1997	0.0244	0.9513
1D CNN-LSTM	0.0324	0.2421	0.0328	0.9124
LSTM	0.0397	0.3423	0.0384	0.8821

Figures 8 (a)–(d) respectively display the prediction results of SCs14 at 2.9 V, 3 A, and 65 °C. The specific performance indicators of the four methods are presented in Table VI.



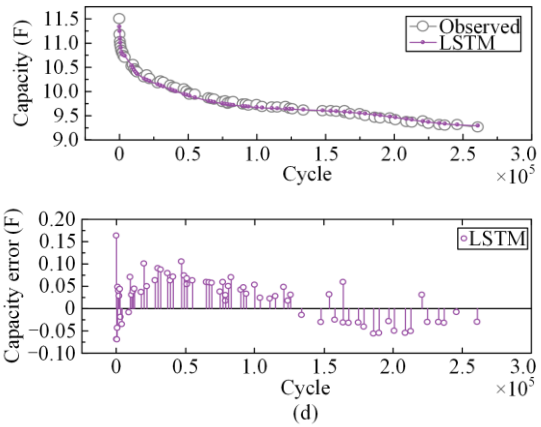
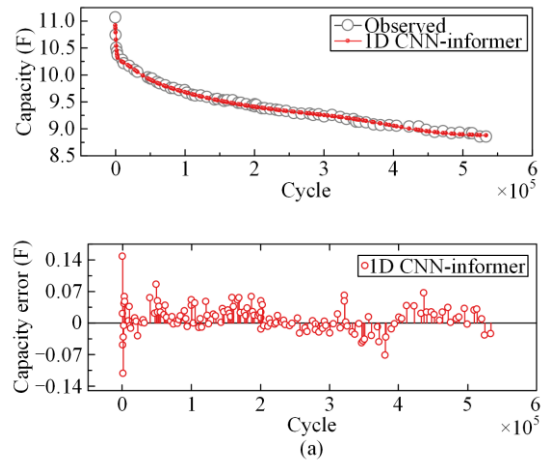


Fig.8. RUL results and error analysis of SCs14 supercapacitor (2.9 V, 3 A, 65 °C) with different models. (a) Prediction of RUL and prediction error using the 1D CNN-informer model. (b) Prediction of RUL and prediction error using the 1D CNN-transformer model. (c) Prediction of RUL and prediction error using the 1D CNN-LSTM model. (d) Prediction of RUL and prediction error using the LSTM model.

TABLE VI
RUL PREDICTION ACCURACY OF SCS14 SUPERCAPACITOR

Methods	R_{MSE}	M_{APE} (%)	M_{AE}	R^2
1D CNN-informer	0.0235	0.2822	0.0237	0.9878
1D CNN-transformer	0.0286	0.3077	0.0304	0.9551
1D CNN-LSTM	0.0399	0.4318	0.0421	0.8923
LSTM	0.0461	0.5132	0.0564	0.8803

Figures 9 (a)–(d) respectively display the prediction results of SCs15 at 2.7 V, 3 A, and 25 °C. The specific performance indicators of the four methods are presented in Table VII. The results indicate that because of the untrained data of the supercapacitor in these conditions, there is a relatively large initial prediction error. To visually demonstrate the prediction performances of different models in various operating conditions, the accuracy of RUL prediction for SCs14 and SCs15 supercapacitors is shown in Fig. 10.



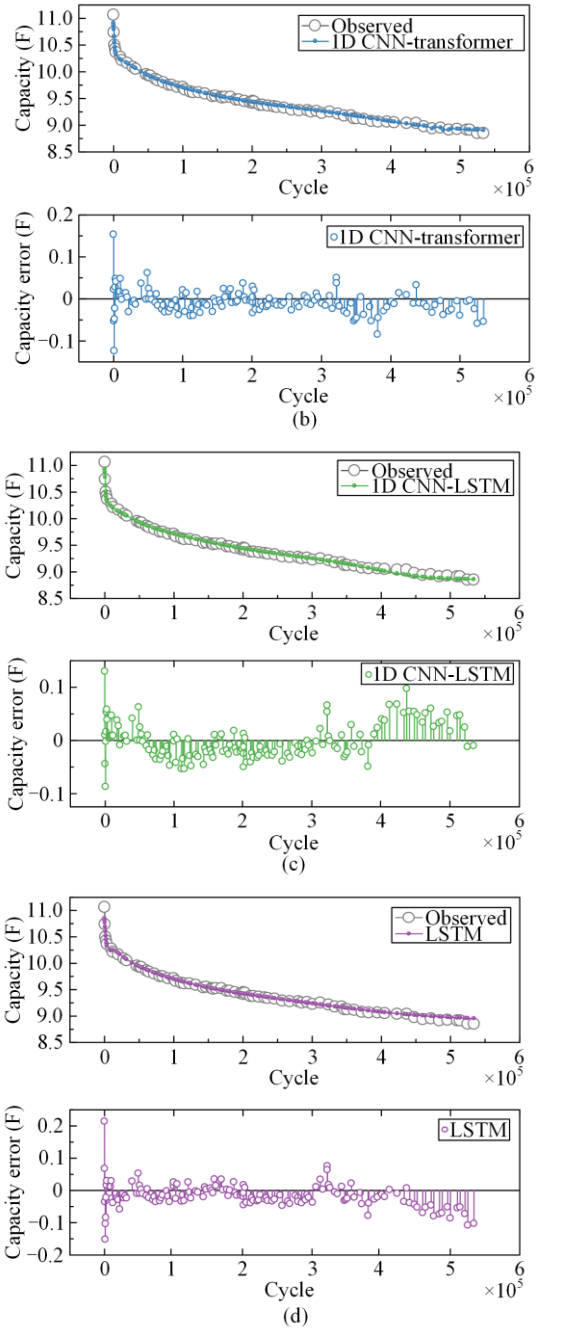


Fig. 9. RUL prediction results and error analysis of SCs15 supercapacitor (2.7 V, 3 A, 25 °C) with different models. (a) Prediction of RUL and prediction error using the 1D CNN-informer model. (b) Prediction of RUL and prediction error using the 1D CNN-transformer model. (c) Prediction of RUL and prediction error using the 1D CNN-LSTM model. (d) Prediction of RUL and prediction error using the LSTM model.

TABLE VII
RUL PREDICTION ACCURACY OF SCS15 SUPERCAPACITOR

Methods	R_{MSE}	M_{APE} (%)	M_{AE}	R^2
1D CNN-informer	0.0248	0.2922	0.0257	0.9867
1D CNN-transformer	0.0296	0.3377	0.0304	0.9515
1D CNN-LSTM	0.0429	0.4518	0.0441	0.8913
LSTM	0.0504	0.5332	0.0594	0.8802

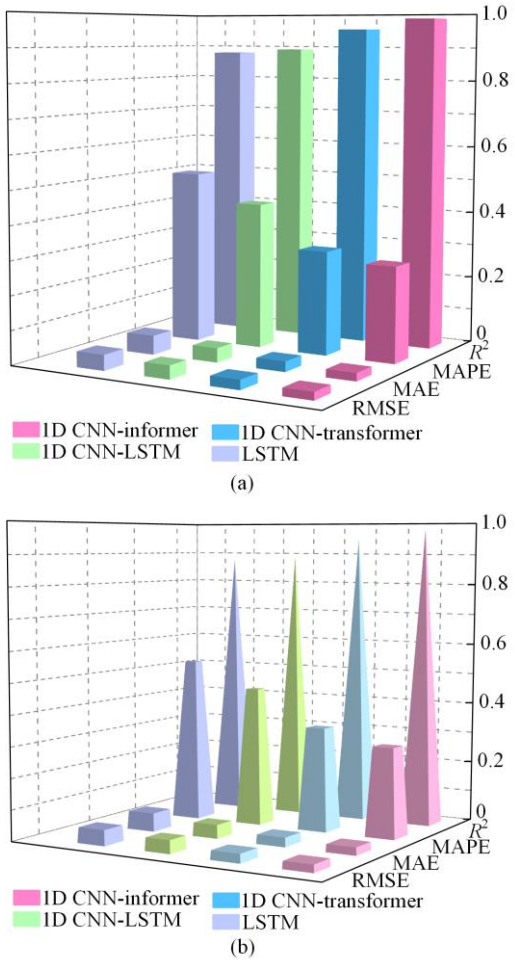


Fig. 10. Prediction error analysis. (a) Bar graphs of the RUL prediction error details for supercapacitors SCs14. (b) Bar graphs of the RUL prediction error details for supercapacitors SCs15.

The comparative analysis of the proposed method with those from recent publications is presented in Table VIII. Clearly, the RMSE values obtained through the comparative analysis demonstrate the superiority of the proposed approach over other RUL prediction algorithms. As the proposed method integrates the advantages of 1D CNN and Informer architectures to learn global and local features from multiple perspectives, it achieves comprehensive information perception. Consequently, the proposed model exhibits better prediction accuracy in different operating conditions because 1D CNN can extract relevant features from the data, enabling a more comprehensive capture of local informative details. Therefore it can also achieve higher prediction accuracy on untrained datasets. LSTM has disadvantages in parallel processing, whereas although the gradient issues of RNN have been partially resolved, there are still limitations. Some methods may perform well on small-scale datasets but have restrictions in terms of magnitude, while the gradient problems are not fully addressed. Hence there are certain limitations for large-scale data prediction for supercapacitors, resulting in compromised prediction accuracy and efficiency.

Since the positional encoding method has a significant impact on the predictive performance of the Informer model, a relative positional encoding algorithm is introduced into Informer to allow the model to better learn the sequence relationships between the data, effectively reducing prediction variations. The proposed model demonstrates competitiveness with other models, resulting in lower RMSE, MAPE (%), MAE values, and higher R^2. By extracting local feature information through convolutional mechanisms and capturing global representations using ProbSparse self-attention, the proposed method effectively balances local and global information. Compared to other methods, the proposed approach combines the strengths of both models, resulting in higher prediction accuracy and better stability.

TABLE VIII
COMPARISON OF RECENT PUBLICATIONS WITH OUR PROPOSED METHOD FOR PREDICTING RUL RESULTS FOR SUPERCAPACITORS

Method	References	Metric	Error
HHO-LSTM	[34]	RMSE	0.0301
TCN	[4]	RMSE	0.0270
HGA-LSTM	[16]	RMSE	0.0294
Mixers-BTCN	[62]	RMSE	0.0248
SSA-Elman	[63]	RMSE	0.0295
GPR	[64]	RMSE	0.0344
IPSO-BPNN	[65]	RMSE	0.0351
1D CNN-informer	Proposed in this paper	RMSE	0.0208

With the continuous development of electronic technology, in the future all of electrical and electronic architecture, the data processing and algorithmic computing functions of the supercapacitor management system (SMS) will be transferred to domain processors, which have more powerful computational capabilities [66], [67]. This will provide opportunities for the practical application of neural networks [68]–[71]. Therefore the comprehensive performance of the above methods is analyzed below, with evaluation indicators including FLOPs, parameters, training time, and storage size. FLOPs are used to measure computational complexity, and parameters indirectly affect the model's storage size. The storage size is calculated using the os.path.getsize function, while training time represents the training efficiency of the model, which is recorded using the time.time function. Table IX presents the above parameters for each model. As can be seen, Informer has a lower time complexity than Transformer, and the self-attention distillation mechanism can effectively reduce the temporal dimension of feature maps, thereby reducing memory consumption and improving prediction efficiency. Compared to LSTM, the proposed model does not have an advantage in storage size. However, in practical application, the accuracy and robustness of the model need to be considered, so a reasonable balance should be achieved based on spe-

cific application scenarios.

TABLE IX
COMPREHENSIVE EVALUATION OF THE PROPOSED METHOD

Methods	Performance index			
	FLOPs ($\times 10^6$)	Parameters ($\times 10^6$)	Training time (hour)	Storage size (kB)
1D CNN-informer	1.201	0.0248	0.196	2212
1D CNN-transformer	1.223	0.0261	0.223	2436
1D CNN-LSTM	0.631	0.0075	0.291	976
LSTM	0.684	0.0045	0.268	762

V. CONCLUSION

This study proposes a novel improved 1D CNN-Informer model to achieve accurate prediction of the RUL of supercapacitors. The model captures local details through convolutional mechanisms and obtains global representations using ProbSparse self-attention, effectively leveraging the advantages of both approaches. Therefore it can better capture both local and global information than other traditional prediction methods. Also, this strategy is well-suited for large datasets as in supercapacitors. Compared to other methods, the proposed 1D CNN-informer model achieves higher RUL prediction accuracy, with RMSE, MAE, and MAPE values all below 0.4% and R^2 values above 0.98. The experimental results demonstrate that the proposed model has higher accuracy and robustness than the others. Additionally, the model exhibits higher training and prediction efficiency, indicating further research and development potential. This work contributes to the development of intelligent energy management systems and provides theoretical support for the application of supercapacitor systems. While the proposed model has advantages in prediction accuracy and efficiency, it has higher complexity and requires more storage size than LSTM, resulting in a greater computational burden. With the continuous development of electronic technology, the derived domain processors will have more powerful computing capabilities, which will facilitate the practical application of neural networks. In addition, the proposed model has certain advantages in time series problems and can also be applied to similar fields.

This study also investigates the 1D CNN-LSTM model with attention mechanism and the Informer model without relative positional encoding algorithm to comprehensively demonstrate the performance of the proposed model. However, considering the inherent drawbacks of these two models, such as long training time and slow convergence speed, as well as the lack of significant improvement in prediction accuracy, their performance is not shown, while several representative models that can effectively compare the performance between models are presented.

In future work, it is worth further exploring how to reduce the computational burden, improve the convergence speed, and reduce the training time of the model while enhancing prediction accuracy and efficiency. Additionally, various factors that affect the rate of supercapacitor aging, including internal factors such as charge-discharge rates and external factors such as pressure and air humidity, will be further considered. Furthermore, the aging mechanism and RUL prediction of supercapacitors will be studied at the system level. Compared to lithium-ion batteries, publicly available datasets for supercapacitors are limited, and more research is needed to continue improving in this area.

ACKNOWLEDGMENT

Not applicable.

AUTHORS' CONTRIBUTIONS

Hao Zhang and Zhenxiao Yi: writing original draft, methodology, conceptualization, software, investigation, formal analysis, writing review and editing. Le Kang and Kai Wang: writing review and editing, conceptualization, methodology, supervision, project administration and funding acquisition. Yi Zhang and Kai Wang: writing review and editing. All authors read and approved the final manuscript.

FUNDING

This research is funded by the Youth Fund of Shandong Province Natural Science Foundation (No. ZR2020QE212), the Key Projects of Shandong Province Natural Science Foundation (No. ZR2020KF020), the Guangdong Provincial Key Lab of Green Chemical Product Technology (No. GC 202111), the Zhejiang Province Natural Science Foundation (No. LY22E070007) and the National Natural Science Foundation of China (No. 52007170).

AVAILABILITY OF DATA AND MATERIALS

Please contact the corresponding author for data material request.

DECLARATIONS

Competing interests: 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 article.

AUTHORS' INFORMATION

Hao Zhang works at the College of Materials Science and Engineering, Xi'an University of Science and Technology, Xi'an 710054, China. His research inter-

ests include state assessment and life prediction of new energy storage devices, energy storage element.

Zhenxiao Yi is currently studying at the School of Electrical Engineering, Qingdao University, Shandong Province, China. His research interests include state assessment and life prediction of new energy storage devices, distributed microgrid and energy storage.

Le Kang works at the College of Materials Science and Engineering, Xi'an University of Science and Technology, Xi'an, Shaanxi Province, China. His research interests include state assessment and life prediction of new energy storage devices, energy Internet.

Yi Zhang works at the School of Energy Sciences and Engineering, Nanjing Tech University, Nanjing, Jiangsu Province, China. His research interests include state assessment and life prediction of new energy storage devices, energy storage element.

Kai Wang works at the School of Electrical Engineering, Qingdao University, Shandong Province, China. His research interests include state assessment and life prediction of new energy storage devices, energy storage element, storage and conversion of new energy.

REFERENCES

[1] L. Wang, T. Wang, and G. Huang *et al.*, "Softly collaborated voltage control in PV rich distribution systems with heterogeneous devices," *IEEE Transactions on Power Systems*, pp. 1-12, Dec. 2023.

[2] D. Jin, Z. Gu, and Z. Zhang, "Lithium battery health degree and residual life prediction algorithm," *Power System Protection and Control*, vol. 51, no. 1, pp. 122-130, Jan. 2023. (in Chinese)

[3] S. Jha, M. Yen, and Y.S. Salinas *et al.*, "Machine learning-assisted materials development and device management in batteries and supercapacitors: performance comparison and challenges," *Journal of Materials Chemistry A*, vol. 11, no. 8, pp. 3904-3936, Jan. 2023.

[4] C. Liu, D. Li, and L. Wang *et al.*, "Strong robustness and high accuracy in predicting remaining useful life of supercapacitors," *APL Materials*, vol. 10, no. 6, Jun. 2022.

[5] J. Ruan, Q. Song, and W. Yang, "The application of hybrid energy storage system with electrified continuously variable transmission in battery electric vehicle," *Energy*, vol. 183, pp. 315-330, Sept. 2019.

[6] V. Sawant, R. Deshmukh, and C. Awati, "Machine learning techniques for prediction of capacitance and remaining useful life of supercapacitors: a comprehensive review," *Journal of Energy Chemistry*, vol. 77, pp. 438-451, Feb. 2023.

[7] C. Liu, Q. Li, and K. Wang, "State-of-charge estimation and remaining useful life prediction of supercapacitors," *Renewable & Sustainable Energy Reviews*, vol. 150, Oct. 2021.

[8] Y. Xu, H. Zhang, and F. Yang *et al.*, "State of charge estimation of supercapacitors based on multi-innovation unscented Kalman filter under a wide temperature range," *International Journal of Energy Research*, vol. 46, no. 12, pp. 16716-16735, Jul. 2022.

[9] Z. Yi, K. Zhao, and J. Sun, *et al.*, "Prediction of the remaining useful life of supercapacitors," *Mathematical Problems in Engineering*, vol. 2022, May 2022.

[10] W. Wang, D. Yang, and Z. Huang *et al.*, "Electrodeless nanogenerator for dust recover," *Energy Technology*, vol. 10, no. 12, Oct. 2022.

[11] C. Liu, Y. Zhang, and J. Sun *et al.*, "Stacked bidirectional LSTM RNN to evaluate the remaining useful life of supercapacitor," *International Journal of Energy Research*, vol. 46, no. 3, pp. 3034-3043, Oct. 2021.

[12] Y. Zhou, Y. Huang, and J. Pang *et al.*, "Remaining useful life prediction for supercapacitor based on long short-term memory neural network," *Journal of Power Sources*, vol. 440, Nov. 2019.

[13] Y. Liu, L. Wang, and D. Li *et al.*, "State-of-health estimation of lithium-ion batteries based on electrochemical impedance spectroscopy: a review," *Protection and Control of Modern Power Systems*, vol. 8, no. 3, pp. 1-17, Jul. 2023.

[14] F. Naseri, S. Karimi, and E. Farjah *et al.*, "Supercapacitor management system: a comprehensive review of modeling, estimation, balancing, and protection techniques," *Renewable & Sustainable Energy Reviews*, vol. 155, Mar. 2022.

[15] Z. Yi, Z. Chen, and K. Yin *et al.*, "Sensing as the key to the safety and sustainability of new energy storage devices," *Protection and Control of Modern Power Systems,* vol. 8, no. 2, pp. 1-22, Apr. 2023.

[16] Y. Zhou, Y. Wang, and K. Wang *et al.*, "Hybrid genetic algorithm method for efficient and robust evaluation of remaining useful life of supercapacitors," *Applied Energy*, vol. 260, Feb. 2020.

[17] A. El Mejdoubi, H. Chaoui, and H. Gualous *et al.*, "Online parameter identification for supercapacitor state-of-health diagnosis for vehicular applications," *IEEE Transactions on Power Electronics*, vol. 32, no. 12, pp. 9355-9363, Jan. 2017.

[18] Q. Lin, Z. Xu, and C. M. Lin, "Battery-supercapacitor state-of-health estimation for hybrid energy storage system using a fuzzy brain emotional learning neural network," *International Journal of Fuzzy Systems*, vol. 24, no. 1, pp. 12-26, Aug. 2021.

[19] L. Wang, L. Xie, and Y. Yang *et al.*, "Distributed online voltage control with fast PV power fluctuations and imperfect communication," *IEEE Transactions on Smart Grid*, vol. 14, no. 5, pp. 3681-3695, Jan. 2023.

[20] Z. Cui, J. Dai, and J. Sun *et al.*, "Hybrid methods using neural network and kalman filter for the state of charge estimation of lithium-ion battery," *Mathematical Problems in Engineering*, vol. 2022, May 2022.

[21] F. Naseri, E. Farjah, and T. Ghanbari *et al.*, "Online parameter estimation for supercapacitor state-of-energy and state-of-health determination in vehicular applications," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 9, pp. 7963-7972, Sept. 2020.

[22] C. Zhang, C. Cao, and R. Chen *et al.*, "Three-leg quasi-Z-source inverter with input ripple suppression for

renewable energy application," *Energies*, vol. 16, no. 11, May 2023.

[23] X. Yu, Y. Shang, and L. Zheng *et al.*, "Application of nanogenerators in the field of acoustics," *Acs Applied Electronic Materials*, vol. 5, no. 9, pp. 5240-5248, Sept. 2023.

[24] W. Wang, D. Yang, and X. Yan *et al.*, "Triboelectric nanogenerators: the beginning of blue dream," *Frontiers of Chemical Science and Engineering*, vol. 17, no. 6, pp. 635-678, Jun. 2023.

[25] M. Zhang, D. Yang, and J. Du *et al.*, "A review of SOH prediction of Li-ion batteries based on data-driven algorithms," *Energies*, vol. 16, no. 7, pp. 3167, Apr. 2023.

[26] M. Zhang, W. Wang, and G. Xia *et al.*, "Self-Powered electronic skin for remote human-machine synchronization," *Acs Applied Electronic Materials*, vol. 5, no. 1, pp. 498-508, Jan. 2023.

[27] M. Zhang, Y. Liu, and D. Li *et al.*, "Electrochemical Impedance Spectroscopy: A new chapter in the fast and accurate estimation of the state of health for lithium-ion batteries," *Energies*, vol. 16, no. 4, pp. 1599, Feb. 2023.

[28] H. Zhang, J. Gao, and L. Kang *et al.*, "State of health estimation of lithium-ion batteries based on modified flower pollination algorithm-temporal convolutional network," *Energy*, vol. 283, Nov. 2023.

[29] X. Sun, Y. Zhang, and Y. Zhang *et al.*, "Summary of health-state estimation of lithium-ion batteries based on electrochemical impedance spectroscopy," *Energies*, vol. 16, no. 15, Aug. 2023.

[30] P. Ma, S. Cui, and M. Chen *et al.*, "Review of family-level short-term load forecasting and its application in household energy management system," *Energies*, vol. 16, no. 15, Aug. 2023.

[31] J. Chen, G. Qi, and K. Wang, "Synergizing machine learning and the aviation sector in lithium-ion battery applications: a review," *Energies*, vol. 16, no. 17, Sept. 2023.

[32] H. Sun, D. Yang, and L. Wang *et al.*, "A method for estimating the aging state of lithium-ion batteries based on a multi-linear integrated model," *International Journal of Energy Research*, vol. 46, no. 15, pp. 24091-24104, Dec. 2022.

[33] Y. Guo, P. Yu, and C. Zhu *et al.*, "A state-of-health estimation method considering capacity recovery of lithium batteries," *International Journal of Energy Research*, vol. 46, no. 15, pp. 23730-23745, Dec. 2022.

[34] N. Ma, H. Yin, and K. Wang, "Prediction of the remaining useful life of supercapacitors at different temperatures based on improved long short-term memory," *Energies*, vol. 16, no. 14, Jul. 2023.

[35] A. El Mejdoubi, H. Chaoui, and J. Sabor *et al.*, "Remaining useful life prognosis of supercapacitors under temperature and voltage aging conditions," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4357-4367, May 2018.

[36] Y. Ma, Y. Chen, and X. Zhou *et al.*, "Remaining useful life prediction of lithium-ion battery based on Gauss-Hermite particle filter," *IEEE Transactions on Control Systems Technology*, vol. 27, no. 4, pp. 1788-1795, Jul. 2019.

[37] K. A. Severson, P. M. Attia, and N. Jin *et al.*, "Data-driven prediction of battery cycle life before capacity degradation," *Nature Energy*, vol. 4, no. 5, pp. 383-391, May 2019.

[38] C. Wang, R. Xiong, and J. Tian *et al.*, "Rapid ultracapacitor life prediction with a convolutional neural network," *Applied Energy,* vol. 305, Jan. 2022.

[39] Y. Zhou, Y. Huang, and J. Pang *et al.*, "Remaining useful life prediction for supercapacitor based on long short-term memory neural network," *Journal of Power Sources*, vol. 440, Nov. 2019.

[40] B. Yang, R. Liu, and E. Zio, "Remaining useful life prediction based on a double-convolutional neural network architecture," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9521-9530, Dec. 2019.

[41] D. Guo, Y. Xia, and X. Luo, "Scene classification of remote sensing images based on saliency dual attention residual network," *IEEE Access*, vol. 8, pp. 6344-6357, Jan. 2020.

[42] S. Wang, F. Wu, and P. T. Aninakwa *et al.*, "Improved singular filtering-Gaussian process regression-long short-term memory model for whole-life-cycle remaining capacity estimation of lithium-ion batteries adaptive to fast aging and multi-current variations," *Energy*, vol. 284, Dec. 2023.

[43] S. Wang, Y. Fan, and S. Jin *et al.*, "Improved anti-noise adaptive long short-term memory neural network modeling for the robust remaining useful life prediction of lithium-ion batteries," *Reliability Engineering & System Safety*, vol. 230, Feb. 2023.

[44] D. Li and L. Yang, "Remaining useful life prediction of lithium battery based on sequential cnn-lstm method," *Journal of Electrochemical Energy Conversion and Storage*, vol. 18, no. 4, Nov. 2021.

[45] P. Li, Z. Zhang, and Q. Xiong *et al.*, "State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network," *Journal of Power Sources*, vol. 459, May 2020.

[46] P. Li, Z. Zhang, and R. Grosu *et al.*, "An end-to-end neural network framework for state-of-health estimation and remaining useful life prediction of electric vehicle lithium batteries," *Renewable and Sustainable Energy Reviews*, vol. 156, Mar. 2022.

[47] O. Abedinia, N. Amjady, and H. Zareipour, "A new feature selection technique for load and price forecast of electrical power systems," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 62-74, Jan. 2017.

[48] K. Zhou, W. Wang, and T. Hu, "Time series forecasting and classification models based on recurrent with attention mechanism and generative adversarial networks," *Sensors*, vol. 20, no. 24, Dec. 2020.

[49] H. Zhou, Y. Zhang, and L. Yang, "Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism," *IEEE Access,* vol. 7, pp. 78063-78074, Jun. 2019.

[50] M. Aslam, M.S. Rehan, and F.R. Albogamy *et al.*, "Prognostication of failures using signal-to-noise ratio to determine partial discharges activities in power transformers," *IEEE Access,* vol. 10, pp. 87500-87507, Aug. 2022.

[51] A. Dixit, C. Ekanayake, and H. Ma *et al.*, "Thermal analysis of natural cooling type distribution transformer retrofilled with natural ester oil," *IEEE Transactions on*

Dielectrics and Electrical Insulation, vol. 29, no. 1, pp. 231-239, Feb. 2022.

[52] G. Luo, Y. Zhou, and X. Sun *et al.*, "Towards lightweight transformer via group-wise transformation for vision-and-language tasks," *IEEE Transactions on Image Processing*, vol. 31, pp. 3386-3398, Apr. 2022.

[53] Z. Yan, T. Tang, and Z. Yu *et al.*, "A circulating power suppression structure for three-winding-transformer-based converter," *IEEE Transactions on Power Electronics*, vol. 37, no. 8, pp. 8800-8804, Mar. 2022.

[54] H. Zhou, S. Zhang, and J. Peng *et al.*, "Informer: beyond efficient transformer for long sequence time-series forecasting," in the *Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)*, Washington, USA, Mar. 2021, pp. 11106-11115.

[55] J. Hu and W. Zheng, "Multistage attention network for multivariate time series prediction," *Neurocomputing*, vol. 383, pp. 122-137, Mar. 2020.

[56] T. Y. Kim and S. B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol. 182, pp. 72-81, Sept. 2019.

[57] J. Gu, Z. Wang, and J. Kuen *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognition,* vol. 77, pp. 354-377, May 2018.

[58] A. Vaswani, N. Shazeer, and N. Parmar *et al.*, "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, California, USA, Dec. 2017, pp. 6000-6010.

[59] Z. Ma, J. Song, and J. Zhang, "Energy consumption prediction of air-conditioning systems in buildings by selecting similar days based on combined weights," *Energy and Buildings*, vol. 151, pp. 157-166, Sept. 2017.

[60] C. Yildiz, H. Acikgoz, and D. Korkmaz *et al.*, "An improved residual-based convolutional neural network for very short-term wind power forecasting," *Energy Conversion and Management*, vol. 228, Jan. 2021.

[61] F. Zheng, Y. Xing, and J. Jiang *et al.*, "Influence of different open circuit voltage tests on state of charge online estimation for lithium-ion batteries," *Applied Energy*, vol. 183, pp. 513-525, Dec. 2016.

[62] J. Gao, D. Yang, and S. Wang *et al.*, "State of health estimation of lithium-ion batteries based on Mixers-bidirectional temporal convolutional neural network," *Journal of Energy Storage*, vol. 73, Dec. 2023.

[63] Y. Guo, D. Yang, and Y. Zhang *et al.*, "Online estimation of SOH for lithium-ion battery based on SSA-Elman neural network," *Protection and Control of Modern Power Systems*, vol. 7, no. 3, pp. 1-17, Jul. 2022.

[64] H. H. Goh, Z. Lan, and D. Zhang *et al.*, "Estimation of the state of health (SOH) of batteries using discrete curvature feature extraction," *Journal of Energy Storage*, vol. 50, Jun. 2022.

[65] Y. Ma, M. Yao, and Z. Tang, "State of health estimation and remaining useful life prediction for lithium-ion batteries by improved particle swarm optimization-back propagation neural network," *Journal of Energy Storage*, vol. 52, Aug. 2022.

[66] Y. Zhou, Z. Ma, and X. Shi *et al.*, "Multi-agent optimal scheduling for integrated energy system considering the global carbon emission constraint," *Energy,* vol. 288, Feb. 2023.

[67] Y. Wei, H. Sun, and T. Zhang *et al.*, (2024, Feb.), "Study of inductively coupled fuel cell dmppt converters," Electrical Engineering, [Online]. Available: https://doi.org/10.1007/s00202-024-02255-5

[68] X. Yu, T. Ai, and K. Wang, (2024, Feb.), "Application of nanogenerators in acoustics based on artificial intelligence and machine learning," *APL Materials*, [Online]. Available: https://doi.org/10.1063/5.0195399

[69] J. Sun, H. Sun, and J. Jiang, (2024, Feb.), "An improved modulation method for low common-mode current non-isolated series simultaneous power supply dual-input inverters for new energy generation applications," *Electrical Engineering*, [Online]. Available: https://doi.org/10.1007/s00202-024-02285-z.

[70] X. Yu, Y. Shang, and L. Zheng *et al.*, "Application of nanogenerator in the field of acoustics," *ACS Applied Electronic Materials*, vol. 5, pp. 5240-5248, Sept. 2023.

[71] Y. Zhou, Z. Ma, and J. Zhang *et al.*, "Data-driven stochastic energy management of multi energy system using deep reinforcement learning," *Energy*, vol. 261, Dec. 2022.