scientific reports

Check for updates

OPEN

A hybrid approach for lithium-ion battery remaining useful life prediction using signal decomposition and machine learning

Yibiao Fan[1,4], Zhishan Lin[2,4], Fan Wang[3,4] & Jianpeng Zhang[3,4]✉

Lithium-ion batteries are widely used in many fields, and accurate prediction of their remaining useful life (RUL) was crucial for effective battery management and safety assurance. In order to solve the problem of reduced RUL prediction accuracy caused by the local capacity regeneration phenomenon during battery capacity degradation, this paper proposed a novel RUL prediction method, which combined complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) technique with an innovative hybrid prediction strategy that integrated the support vector regression (SVR) and the long short-term memory (LSTM) networks. First, CEEMDAN was used to decompose the battery capacity data into high-frequency and low-frequency components, thereby reducing the impact of capacity regeneration. Subsequently, the SVR model predicted the low-frequency component that characterized the main degradation trend, and the high-frequency component that contained capacity regeneration features was predicted using an LSTM network optimized by the sparrow search algorithm (SSA). Finally, the final RUL prediction was obtained by combining the predictions of the two models. Experimental results on NASA public datasets showed that the proposed hybrid method significantly outperformed existing methods: the RMSE of the methods proposed in this paper were all less than 0.0086 Ah, the MAE were all less than 0.0060 Ah, the R^2 values were all higher than 0.96, and the RUL prediction errors were controlled within one cycle. This method gave full play to the complementary advantages of SVR and LSTM and provided an accurate and reliable solution for RUL prediction of lithium-ion batteries.

Keywords Li-ion batteries, Remaining useful life prediction, Long short-term memory, Support vector regression, CEEMDAN, Sparrow search algorithm

With the advent of the new energy era, lithium-ion batteries have gained extensive applications across fields such as electric vehicles, aerospace, military communications, and distributed energy storage systems, owing to their lightweight design, high energy density, low self-discharge rate, large capacity, and long cycle life[1,2]. As usage time increases, lithium-ion batteries experience a slowdown in chemical reactions and degradation of electrode materials, leading to a reduction in capacity and an increase in internal resistance, which may result in faults or safety incidents[3]. Therefore, accurate prediction of the RUL of lithium batteries is of great significance for efficient battery management and avoidance of safety accidents[4].

Currently, two main approaches are employed for predicting lithium-ion battery lifespan: model-based analysis methods and data-driven prediction methods[5]. Model-based analysis methods integrate the fundamental properties of battery materials, operating environmental conditions, detailed physicochemical changes, and the key mechanisms involved in the aging process[6]. Khodadadi Sadabadi et al. proposed an electrochemical model approach based on an enhanced single particle model for predicting the RUL of composite electrode lithium-ion batteries[7]. Liu et al. proposed a particle filter framework based on an electrochemical model, incorporating

[1]School of Physics and Mechanical and Electrical Engineering, Longyan University, Longyan 361000, China. [2]Fujian Antong Electric Co., Ltd, Longyan 361000, China. [3]School of Aerospace Engineering, Xiamen University, Xiamen 361102, China. [4]Yibiao Fan, Zhishan Lin, Fan Wang and Jianpeng Zhang contributed equally to this work. ✉email: 34520241151602@stu.xmu.edu.cn

battery aging mechanisms to predict the RUL of lithium-ion batteries[8]. The accuracy of predictions made by model-based analysis methods often depends on testing under laboratory conditions, which typically involve complex calculation processes and challenging parameter estimation.

Data-driven methods directly analyze collected data using various signal processing and analytical techniques to uncover hidden informationand link it to battery degradation trends, thereby establishing predictive models[9]. Currently, commonly used data-driven methods can be categorized into three types: artificial intelligence methods, statistical methods, and signal processing methods. Artificial intelligence methods are popular because of their ability to model complex nonlinear relationships in data. Especially machine learning algorithms such as support vector machine (SVM), random forest (RF), and deep learning (DL) techniques, can provide highly accurate and reliable predictions of battery RUL by learning from historical data, making it one of the most commonly used and promising tools in battery health management[10]. Yao et al. proposed a joint state of charge (SOC) and state of health (SOH) online estimation method based on multi-task learning and cross-stitch networks. The method improves the estimation accuracy by sharing information among different tasks and better captures the dynamic features of battery state changes through multi-scale data processing[11]. Kang et al. proposed a novel RUL prediction method combining fuzzy evaluation and gaussian process regression (GPR), which effectively improves the prediction accuracy and expresses the uncertainty of the prediction results[12]. Lin et al.proposed an adaptive adjustable hybrid radial basis function (RBF) network for SOH estimation. The network structure is adaptively adjusted by Brownian motion modeling and particle filter, so that the model achieves high SOH prediction accuracy[13]. Yao et al. proposed a multi-scale data-driven framework utilizing temporal convolutional networks and cross-scale self-attention for SOC estimation, which shows the importance of capturing and fusing features from different time scales[14]. Their work provides valuable insights into how multi-scale analysis can improve battery state estimation accuracy. Zhao et al. extracted the health indicators of batteries from the charging curves and employed LSTM neural networks to predict the trend of these indicators as the number of cycles increased, the predicted health indicators were then input into a GPR aging model to forecast the battery's health status[15]. Chen et al. selected the equal voltage drop discharge time as a health indicator using the gray relational analysis method and developed a battery life prediction model based on a convolutional neural network (CNN)-LSTM[16]. The LSTM model has inherent advantages in handling time-series data; however, the capacity regeneration phenomenon and noise signal that may occur during the degradation of lithium batteries may interfere with the performance of the network model.

The application of signal decomposition technology to lithium battery capacity data can well reduce the influence of noise interference and capacity regeneration phenomenon[17]. Cheng et al. combined empirical mode decomposition (EMD) method and back propagation LSTM to develop battery RUL prediction, and processed battery data through EMD method to reduce the impact of capacity regeneration and other conditions[18]. Li et al. used the EMD algorithm to decompose the recorded battery capacity and cycle count data into multiple sub-layers and employed the LSTM network to predict the high-frequency sub-layers and the Elman network to predict the low-frequency sub-layers, yielding highly accurate prediction results[19]. Yang et al. proposed a RUL prediction method for lithium-ion batteries based on gray wolf optimization (GWO) and SVR combined with ensemble empirical mode decomposition (EEMD), aiming to improve the prediction accuracy of the SVR model by decomposing the battery capacity sequence through EEMD[20]. Although the SVR model is superior in dealing with nonlinear relationships, it still has limitations in capturing the high-frequency components of time series data[21]. EMD and EEMD methods are prone to modal aliasing during signal decomposition and are sensitive to noise in the data, which may lead to unstable decomposition results[22]. The CEEMDAN algorithm effectively mitigates mode mixing by adding white noise and performing iterative decomposition. By averaging multiple noise perturbations, it reduces noise interference, resulting in more stable intrinsic mode functions (IMF) components[23].

In summary, CEMMDAN decomposes the battery capacity into data of different frequencies, which effectively reduces the influence of noise and capacity regeneration on the prediction accuracy. SVR model has good performance in dealing with nonlinear mapping, but its ability to deal with high frequency signals is weak. LSTM network has memory units and nonlinear activation functions, and is good at processing time series data with long-term dependence and nonlinear relationship, especially for high-frequency data. Therefore, this paper proposes a new lithium battery RUL prediction method, combined with CEEMDAN, SVR and LSTM. CEEMDAN is used to decompose the capacity data into multiple IMFs and a residual term. Each IMF represents a different frequency component in the original data, and the residual is the long-term trend part of the data. SVR is used to capture the low-frequency residual part of the data, and LSTM is used to capture the non-linearity and long-term and short-term dependencies of the high-frequency components. In order to prevent the influence of manual adjustment of LSTM hyperparameters on prediction accuracy, sparrow search algorithm (SSA) is used for hyperparameter optimization. The proposed method is experimentally verified on the public data set of NASA Ames prediction center of excellence (PCOE). The results show that the prediction accuracy of the proposed method is higher than that of other mainstream methods, which verifies the effectiveness of the proposed method in accurately predicting battery RUL.

The remainder of this paper is organized as follows. Section "Relevant theoretical basis" presents the theory behind CEEMDAN, SSA,LSTM and SVR. Section "Experimental design" describes the dataset used and the whole process of training and prediction using the proposed method. Section "Experimental results and analysis" presents the experimental results and error analysis. Finally, Section "Conclusions" concludes the paper.

Relevant theoretical basis

Definition of state of health

SOH is used to assess the comparison between the current state of the battery and its initial state, reflecting the extent of performance degradation. Currently, there are numerous definitions of battery SOH in the literature, with the most widely used approach defining SOH based on the degree of capacity degradation[24]. It is defined as:

$$\mathrm{SOH}(t) = \frac{C_t}{C_0} \times 100\% \tag{1}$$

where C_0 represents the battery's capacity at the time of manufacture, and C_t denotes the battery's capacity at time *t*. Generally, the end of the battery's life is considered to occur when its capacity reaches 70% of its initial capacity. The RUL of a battery is defined as the number of charge-discharge cycles it can complete before its capacity declines to 70% of its initial value[25].

Complete ensemble empirical mode decomposition with adaptive noise algorithm

The capacity sequence data of lithium batteries exhibit non-stationary and nonlinear characteristics. By decomposing the data with CEEMDAN, the non-stationarity of the capacity sequence data can be mitigated, enabling the capture of local features within the sequence[26]. The CEEMDAN algorithm is an improved version of the EMD algorithm. It adds white noise to the original data, making the decomposition more robust, effectively mitigating mode mixing issues and enhancing the stability of the decomposition results. Additionally, it offers faster computational speed[27]. The primary steps of the CEEMDAN algorithm are as follows:

(1) Add *n* white noise sequences to the signal $x(t)$ to be decomposed, generating *n* new sequences, as follows:

$$x_i(t) = x(t) + \alpha\omega_i(t), i = 1, 2, \cdots, n \tag{2}$$

where $x_i(t)$ represents the sequence after adding the *i*-th white noise, α is the signal-to-noise ratio, and $\omega_i(t)$ is the white noise sequence following a normal distribution.

(2) Decompose the signal $x_i(t)$ using EMD to obtain the corresponding $IMF_i(t)$. By calculating the mean of $IMF_i(t)$, the first mode component $IMF_1(t)$ of the CEEMDAN decomposition is obtained. Then, subtract $IMF_1(t)$ from $x(t)$ to obtain the first residual component $r_1(t)$ as follows:

$$IMF_1(t) = \frac{1}{n}\sum_{i=1}^{n} IMF_i(t) \tag{3}$$

$$r_1(t) = x(t) - IMF_1(t) \tag{4}$$

(3) Add the white noise $\omega_i(t)$ to $r_1(t)$ to obtain $r_i(t)$.Perform EMD decomposition on $r_i(t)$ and iteratively apply the aforementioned steps until the termination criteria for EMD are no longer satisfied, yielding *k* components. The final residual component is denoted as *R*(*t*), and the original signal *x*(*t*) is expressed as:

$$x(t) = R(t) + \sum_{k=1}^{K} IMF_k(t) \tag{5}$$

Sparrow search algorithm

The sparrow search algorithm (SSA) is a swarm intelligence optimization technique recently proposed by Xue et al.[28]. Inspired by the foraging and anti-predation behaviors of sparrow flocks, SSA is designed to meet the needs of global optimization in complex search spaces. This algorithm has been widely applied for parameter tuning in complex systems, demonstrating efficient global search capabilities and fast convergence.

SSA divides the sparrow population into three roles: discoverers, followers, and scouts. Discoverers are responsible for exploring new solution areas across the global search space and guiding the population toward potential optimal solutions. Followers closely track the discoverers, thereby enhancing the accuracy of local searches, while scouts monitor potential threats to prevent the population from becoming trapped in local optima. The algorithm continuously updates the positions of these three types of individuals and evaluates the fitness values of all individuals in each iteration to approach the optimal solution[29]. The main iterative update steps are as follows:

Step 1: Population Initialization

Randomly generate the sparrow population, calculate the proportions of discoverers and followers, determine the number of scouts, and define the maximum number of iterations.

Step 2: Fitness Calculation

Calculate the fitness value for each sparrow, rank the individuals from best to worst based on their fitness values, and select the highest-performing individual as the reference point.

Step 3: Update Discoverer Positions

Based on the alarm value R_2 and safety threshold ST, determine if a predator is present. If a predator is detected($R_2 \geq ST$), the discoverers will lead the population to a new area. If no predator is detected ($R_2 < ST$), the discoverers will perform a wide-range search. As shown in the following equations:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t} \cdot \exp\left(\frac{-i}{\alpha \cdot iter_{\max}}\right) & if R_2 < ST \\ X_{i,j}^{t} + Q \cdot L & if R_2 \geq ST \end{cases} \tag{6}$$

where, t denotes the current iteration, and $j = 1, 2, \ldots, d$. $X_{i,j}^{t}$ represents the value of the j-th dimension for the i-th sparrow at iteration t. The parameter $\mathrm{iter}_{\max}$ is a constant that specifies the maximum number of iterations, and $\alpha \in (0, 1]$ is a randomly generated number. $R2$ and ST, with values in $[0, 1]$ and $[0.5, 1.0]$ respectively. Q is a normally distributed random variable, while L is a $1 \times d$ matrix where each element is set to 1.

Step 4: Update Follower Positions

The followers update their positions according to the locations of the discoverers. Followers with lower fitness place greater emphasis on the current worst position, while those with higher fitness move towards the best position discovered by the discoverers. The position update formulas for the followers are provided as follows:

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{X_{worst}^{t} - X_{i,j}^{t}}{i^2}\right) & if(i > n/2) \\ X_{p}^{t+1} + \left|X_{i,j}^{t} - X_{p}^{t+1}\right| \cdot A^{+} \cdot \mathrm{L} & \text{otherwise} \end{cases} \tag{7}$$

where X_{P} denotes the optimal position found by the discoverer, while X_{worst} represents the current global worst position. A is a $1 \times d$ matrix in which each element is randomly set to either 1 or -1, and A^{+} is defined as $A^{+} = A^{T}(AA^{T})^{-1}$. When $i > n/2$, this indicates that the i-th follower, with lower fitness, is more likely to be in a "starving" state.

Step 5: Update Scout Positions

When scouts are located at the periphery, they encounter potential danger. To mitigate this threat, they exhibit antipredatory behavior by rapidly relocating to a safer area. When scouts are near the center, they explore their surroundings by adjusting their position relative to the global worst position, thus improving their adaptability. As shown in the following equations:

$$X_{i,j}^{t+1} = \begin{cases} X_{\mathrm{best}}^{t} + \beta \cdot |X_{i,j}^{t} - X_{\mathrm{best}}^{t}| & \text{if} f_i > f_g \\ X_{i,j}^{t} + K \cdot \left(\frac{|X_{i,j}^{t} - X_{\mathrm{worst}}^{t}|}{(f_i - f_w) + \varepsilon}\right) & \text{if} f_i = f_g \end{cases} \tag{8}$$

where X_{best}^{t} denotes the current global best position, and X_{worst}^{t} represents the current global worst position. The parameter f_i indicates the fitness value of the *i*-th sparrow, with f_g and f_w representing the best and worst fitness values in the population, respectively. β and B are random variables that control the step size of position updates. ϵ is a small constant added to avoid division by zero. When $f_i > f_g$, the scout moves toward the global best position to avoid danger. When $f_i = f_g$, the scout adjusts its position relative to the global worst position to enhance search diversity.

Step 6: After updating the positions for each role, recalculate the fitness value of each sparrow, assess the overall state of the population, and update the current best and worst solutions.

Step 7: If the specified convergence criterion is satisfied, output the current best solution; otherwise, return to Step 2 and proceed with the iteration.

In this paper, SSA is employed to optimize the hyperparameters of the LSTM network. The global search capability of SSA helps mitigate the risk of local optima during hyperparameter tuning, while the role-based division of tasks facilitates efficient exploration of the complex parameter space to identify the optimal configuration.

Long short-term memory networks

The LSTM network was designed as a specialized type of recurrent neural network (RNN) to address the issues of gradient vanishing and gradient explosion encountered when processing long sequence data. The LSTM network addresses time series data by introducing memory cells and a gating mechanism. The memory cells function to retain long-term dependency information, while the gating mechanism, composed of the forget gate, input gate, and output gate, regulates the retention, updating, and final output of information[30]. Figure 1 shows the structural unit of the LSTM network. The core components of LSTM include:

(1) *Forget Gate*: Determines which information from the previous time step's cell state, c_{t-1}, should be discarded. The calculation formula is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{9}$$

where f_t is the forget gate activation at time *t*, σ is the sigmoid activation function, used to map values to the range (0, 1) . W_f and b_f are the weight matrix and bias for the forget gate, h_{t-1} is the hidden state from the previous time step, x_t is the input at the current time step.

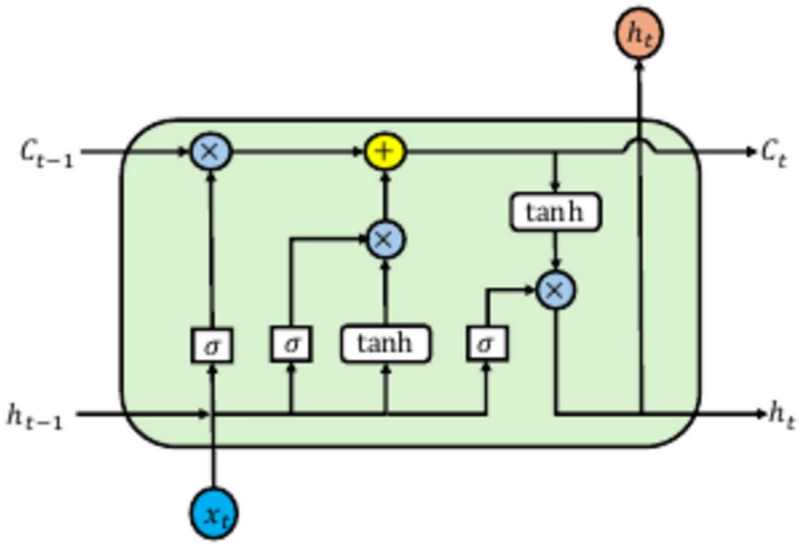

Fig. 1. LSTM network structure unit.

(2) *Input Gate*: Determines which parts of new information should be updated in the cell state. The calculation formula is as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{10}$$

where i_t denotes the input gate activation at time *t*, and W_i and b_i represent the weight matrix and bias associated with the input gate.

(3) *Candidate Cell State*: Calculates the new candidate values to be added to the cell state. The calculation formula is as follows:

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{11}$$

where $\tilde{c}_t$ represents the candidate cell state at time *t*, and W_c and b_c denote the weight matrix and bias for generating the candidate state. The hyperbolic tangent function $\tanh$ ensures non-linearity and bounds the candidate values within a fixed range, aiding in the regulation of cell state updates.

(4) *Cell State Update*: Updates the cell state by combining the previous cell state and new candidate values. The calculation formula is as follows:

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \tag{12}$$

where c_t represents the updated cell state at time *t*.

(5) *Output Gate*: Determines the output from the cell, which will be used as the hidden state. The calculation formula is as follows:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{13}$$

where o_t denotes the output gate activation at time *t*, and W_o and b_o are the weight matrix and bias for the output gate.

(6) *Hidden State*: Combines the output gate activation and the updated cell state to produce the hidden state. The calculation formula is as follows:

$$h_t = o_t \cdot \tanh(c_t) \tag{14}$$

where h_t represents the hidden state at time *t*.

Support vector regression

SVR, extended from SVM, is a powerful regression technique that aims to find a function with at most ε deviation from the target values while maintaining maximum margin[31].

(1) *Basic Principle*: Given a training dataset $(x_i, y_i)_{i=1}^N$, SVR seeks to find a linear function in the feature space through the following optimization problem:

$$f(x) = \mathrm{w}^T \phi(x) + b \tag{15}$$

where $\phi(x)$ maps the input space to a high-dimensional feature space, w is the weight vector, and b is the bias term.

(2) *Optimization Problem*: The primal form of SVR can be expressed as:

$$\min_{\mathrm{w},b,\xi,\xi^*} \frac{1}{2}\|\mathrm{w}\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{16}$$

subject to:

$$\begin{aligned} y_i - (\mathrm{w}^T\phi(x_i) + b) &\le \varepsilon + \xi_i \\ (\mathrm{w}^T\phi(x_i) + b) - y_i &\le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* &\ge 0 \end{aligned} \tag{17}$$

where C is the penalty parameter, ε is the tube size, and ξ_i, ξ_i^* are slack variables.

(3) *Final Solution*: Using the kernel trick, the regression function takes the form:

$$f(x) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K(x_i, x) + b \tag{18}$$

where α_i, α_i^* are Lagrange multipliers and $K(x_i, x)$ is the kernel function.

Experimental design

Introduction of the experimental dataset

In this paper, the original dataset on lithium battery cyclic aging was obtained from the NASA Ames Research Center. Four batteries, numbered B0005, B0006, B0007, and B0018, were selected to validate the effectiveness of the proposed method. All tests on the batteries were conducted at a constant ambient temperature of 24 °C. The specific charge-discharge protocol is as follows: In the charging phase, a constant current (CC) of 1.5 A was initially applied until the battery voltage reached 4.2 V, after which it switched to a constant voltage (CV) mode until the charging current decayed to 20 mA, marking the end of charging. The discharging process involved a constant current (CC) of 2 A until the battery voltage dropped to the cutoff voltages specific to each battery: 2.7 V for battery B0005, 2.5 V for batteries B0006 and B0018, and 2.2 V for battery B0007, completing the discharge process, and Table 1 also lists other key parameters including discharge current (DC), discharge voltage (DV), and battery failure threshold (TS). In this paper, experimental validation was conducted using capacity data preprocessed from the original dataset. The end of life (EOL) of a battery was defined as when the capacity dropped to 70% of the rated capacity (1.4 Ah). However, for battery B0007, which does not reach 1.4 Ah, the EOL was set at 72% of the rated capacity (1.45 Ah). Figure 2 shows the capacity degradation trend of the NASA lithium battery.

Results of CEEMDAN decomposition

In this paper, the CEEMDAN algorithm was directly applied to decompose lithium-ion battery capacity data, resulting in multiple IMF. Compared to the decomposition of lithium-ion battery current and voltage data, this approach offers faster processing speed, higher accuracy, and reduced computational complexity. The selection of the mode number k significantly impacts the decomposition performance. When k is too small, essential

Battery number	CV (V)	DV (V)	CC (A)	DC (A)	TS (AH)
B0005	4.2	2.7	1.5	2.0	1.4
B0006	4.2	2.5	1.5	2.0	1.4
B0007	4.2	2.2	1.5	2.0	1.4
B0018	4.2	2.5	1.5	2.0	1.4

Table 1. Experimental parameters of battery degradation.

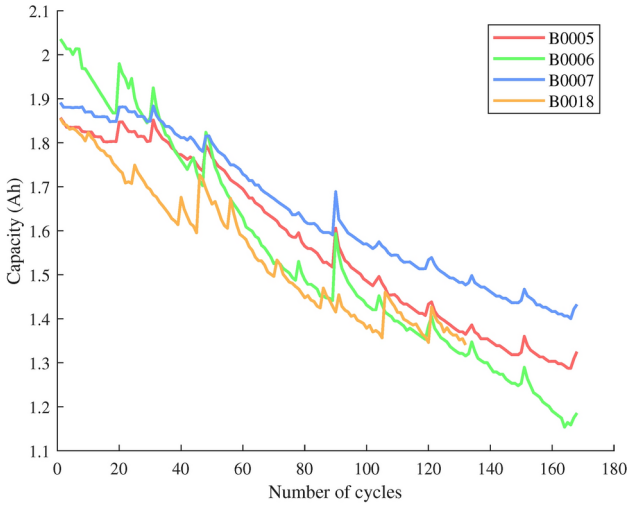

Fig. 2. NASA lithium battery capacity degradation trend.

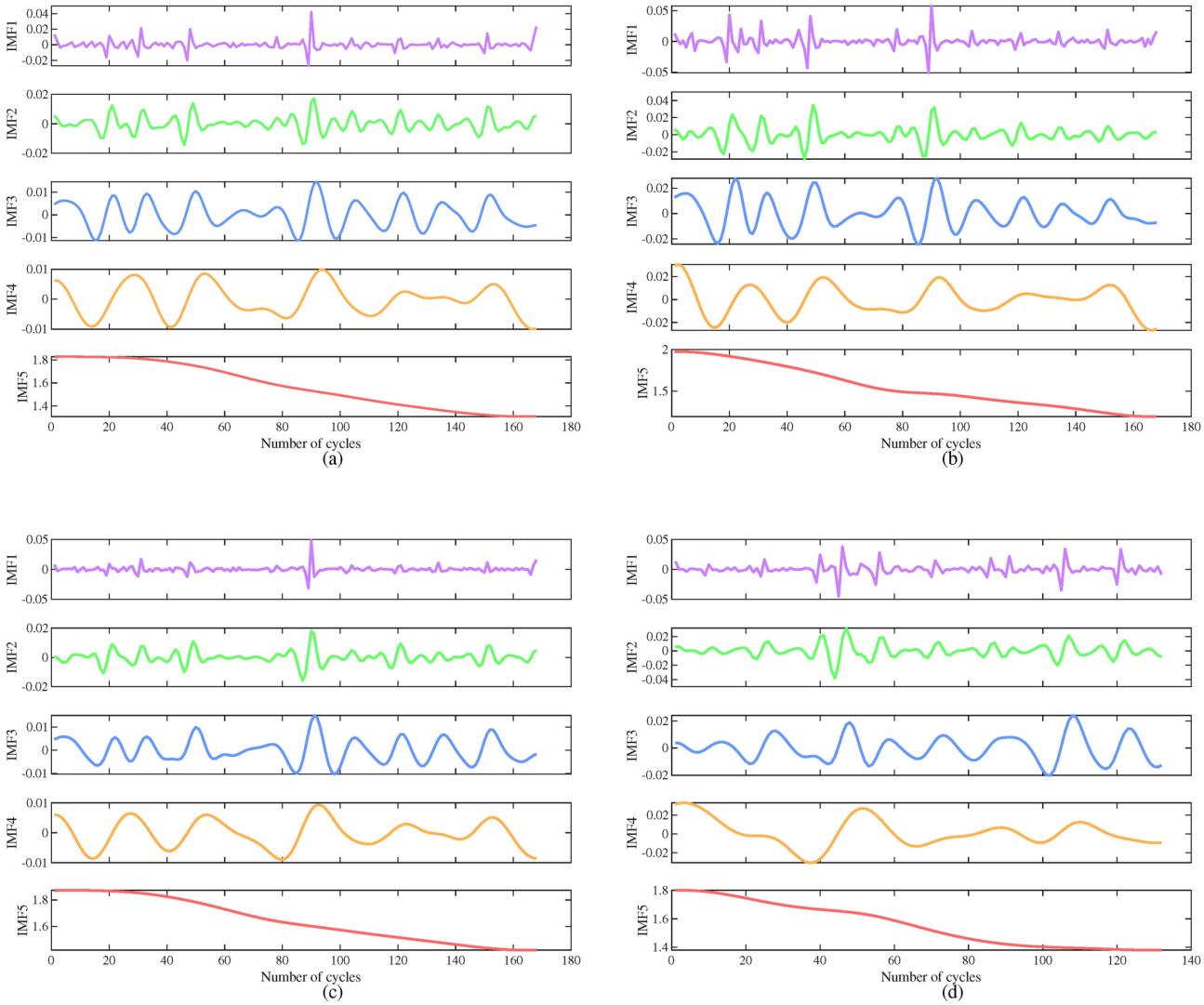

Fig. 3. IMF obtained from the CEEMDAN decomposition: (**a**) B0005 IMF component, (**b**) B0006 IMF component, (**c**) B0007 IMF component and (**d**) B0018 IMF component.

information in the original signal may be lost, whereas a larger k can lead to frequency mixing. Various k values were tested, and by examining the corresponding center frequencies, it was observed that when $k = 5$, the center frequencies were well-dispersed. Therefore, k was set to 5 for this paper. Figure 3 shows the IMF obtained from the decomposition of NASA lithium battery capacity data using the CEEMDAN algorithm. In the figure, each IMF reveals distinct characteristics of the lithium-ion battery data. High-frequency IMFs (IMF1 to IMF3) primarily capture short-term variations and rapid fluctuations, although resembling random noise, it can still provide insights into subtle changes in battery performance. IMF4 highlights the capacity regeneration phenomenon observed in certain cycles, suggesting a partial recovery pattern. IMF5 represents the primary trend, reflecting the overall capacity degradation over time. This decomposition enables a clear analysis, distinguishing between transient fluctuations, periodic features, and the main degradation trend.

Experimental procedure

In this paper, we preprocess the original data set and extract the battery capacity as the direct health index (HI) for battery RUL prediction. The change of capacity is directly related to the SOH of the battery, which can clearly and stably reflect the degradation process of the battery. Compared to indirect HIs (such as constant-voltage charging time, constant-voltage charging energy, discharge temperature peak, etc.), capacity is more stable under varying usage conditions, with lower data noise, leading to improved stability and accuracy of the predictive model.

The CEEMDAN algorithm is used to decompose the capacity data to obtain five IMF components with different frequencies. Among them, IMF1-IMF4 are high-frequency components, which reflect short-term and rapid fluctuations in capacity data and also contain information about capacity regeneration.IMF5 is a low-frequency component, which reflects the long-term declining trend of capacity data.

SVR shows a unique advantage in dealing with low-frequency components, which mainly stems from the natural match between its algorithmic properties and the low-frequency signal characteristics[32]. The ε -insensitive loss function of SVR can effectively filter out the influence of high-frequency noise, and at the same time accurately capture the main trends of the signal. This feature makes SVR particularly suitable for dealing with low-frequency gradient features in the system, which has been validated in the study of minimum frequency prediction in power systems, and shows higher prediction accuracy and better generalization ability compared with the traditional methods[33]. Based on the characteristics of LSTM networks, they have a natural advantage in processing high-frequency data. As Song et al pointed out, LSTM is able to effectively capture the fast fluctuations and nonlinear features in high-frequency data through its unique memory cells and gating mechanism[34]. Rundo's study further verified that LSTM is particularly suitable for processing high-frequency data containing frequent jumps and non-smooth nature, which is mainly attributed to its strong nonlinear modeling ability and adaptive learning of time-series data properties[35]. These studies show that LSTM can naturally adapt to the characteristics of high-frequency data due to the advantages of its network structure. These theoretical analyses show that the intrinsic properties of SVR are naturally compatible with low-frequency signal prediction, while the architectural advantages of LSTM make it particularly suitable for capturing high-frequency components, which provides a solid theoretical basis for our hybrid model design. Considering the small amount of battery capacity data and the relatively smooth curve, the SVR model uses RBF (Radial Basis Function) as the kernel function, with a penalty coefficient $C = 10$, which is used to control the model's tolerance to the training error, and a loss function parameter $\varepsilon = 0.005$, which defines the width of the insensitive region of the SVR. The parameter settings take full account of the characteristics of the dataset while maintaining the predictive accuracy of the model and avoiding the overfitting problem. Taking the B0005 battery as an example, the first 80 data points of the low-frequency component IMF5 are selected to train the model, and the last 88 data points are used as the test set. The prediction results of the SVR model are shown in Fig. 4. The prediction

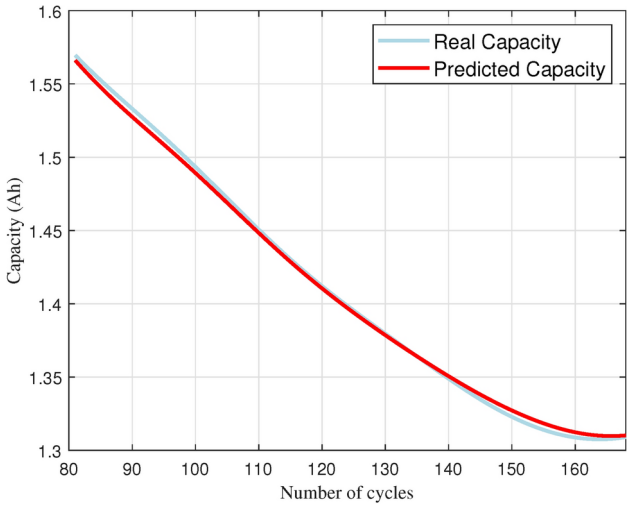

Fig. 4. SVR predictions for IMF5 component of B0005 battery.

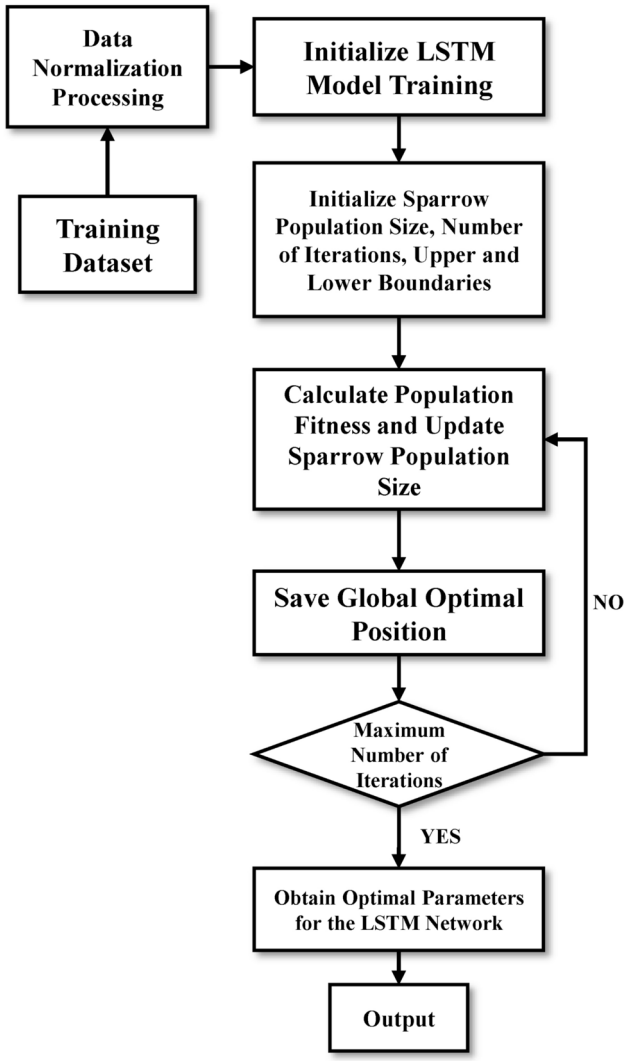

Fig. 5. SSA optimized LSTM network hyperparameter flowchart.

Number of populations	Number of iterations	Warning threshold	Dim
30	5	0.8	4

Table 2. SSA parameter setting.

Hidden units	Learning rate	Maximum epochs	Dropout rate
[300, 500]	[0.001, 0.002]	[400, 800]	[0.4, 0.6]

Table 3. LSTM hyperparameter search range.

accuracy of the whole test set is very high, indicating that the SVR model is very effective for the accurate prediction of low frequency components.

LSTM network parameters directly affect the prediction accuracy of the model and the generalization ability of unknown capacity data. However, existing parameter selection approaches often lack effective personalized solutions, and manual tuning can be time-consuming. This paper employs the SSA algorithm to optimize these primary hyperparameters of the network architecture. The process of using SSA to optimize LSTM network hyperparameters is shown in Fig. 5 The SSA parameters adopted are shown in Table 2, and the range of LSTM hyperparameter optimization is provided in Table 3. The decision to apply the SSA optimization exclusively to the LSTM, rather than to both the LSTM and SVR, was made because of the different characteristics of

these models. SVR is used to simulate the low-frequency component, IMF5, which exhibits a smoothed decay trend. The SVR employs an RBF kernel, with a C-value of 10 and ε-value of 0.005, which provides a robust performance with such smoothed time-series data. The SVR has a higher sensitivity to parameter variations than the LSTM and SVR. SVR contains fewer hyperparameters and is less sensitive to parameter variations when modeling smoothing trends. In contrast, LSTM networks involve multiple interconnected layers with numerous parameters, which has a critical impact on their ability to capture complex time dependencies in the high-frequency components. Our experimental results show that excellent predictive performance and computational efficiency can be achieved by optimizing the LSTM parameters through SSA while keeping the SVR parameters fixed.

In the process of using the LSTM network to predict the high-frequency components, an iterative method is used, where the first n capacity data points were used as input features, and the $(n+1)$-th data point served as the target label. By using a sliding window technique, the subsequent actual capacity value was employed as a new feature to continue predicting forward. In this paper,n was set to 3. In order to reduce the complexity of the iterative process of the algorithm, a single-layer LSTM was configured, followed by a Dropout layer, a Dense layer, and a final activation layer to output the predicted battery capacity. This setup takes into account the small size of the dataset and the characteristic of capacity regeneration. Taking B0005 battery as an example, the prediction results of high frequency components by using SSA optimized LSTM network are shown in Fig. 6. The LSTM network can capture the fluctuation of the capacity regeneration component, showing a good prediction result for the high frequency component.

The final prediction result is obtained by combining the predictions from SVR and LSTM models. The mathematical formulation for the integration can be expressed as:

$$C(t) = C_{IMF5}(t) + \sum_{i=1}^{4} C_{IMFi}(t) \tag{19}$$

where $C(t)$ represents the final predicted capacity at time t, $C_{IMF5}(t)$ is the SVR prediction for the low-frequency component (IMF5), and $C_{IMFi}(t)$ represents the LSTM predictions for high-frequency components (IMF1-IMF4). This integration strategy effectively leverages the complementary advantages of both models, as SVR captures the overall degradation trend through IMF5 while LSTM accurately models the capacity regeneration

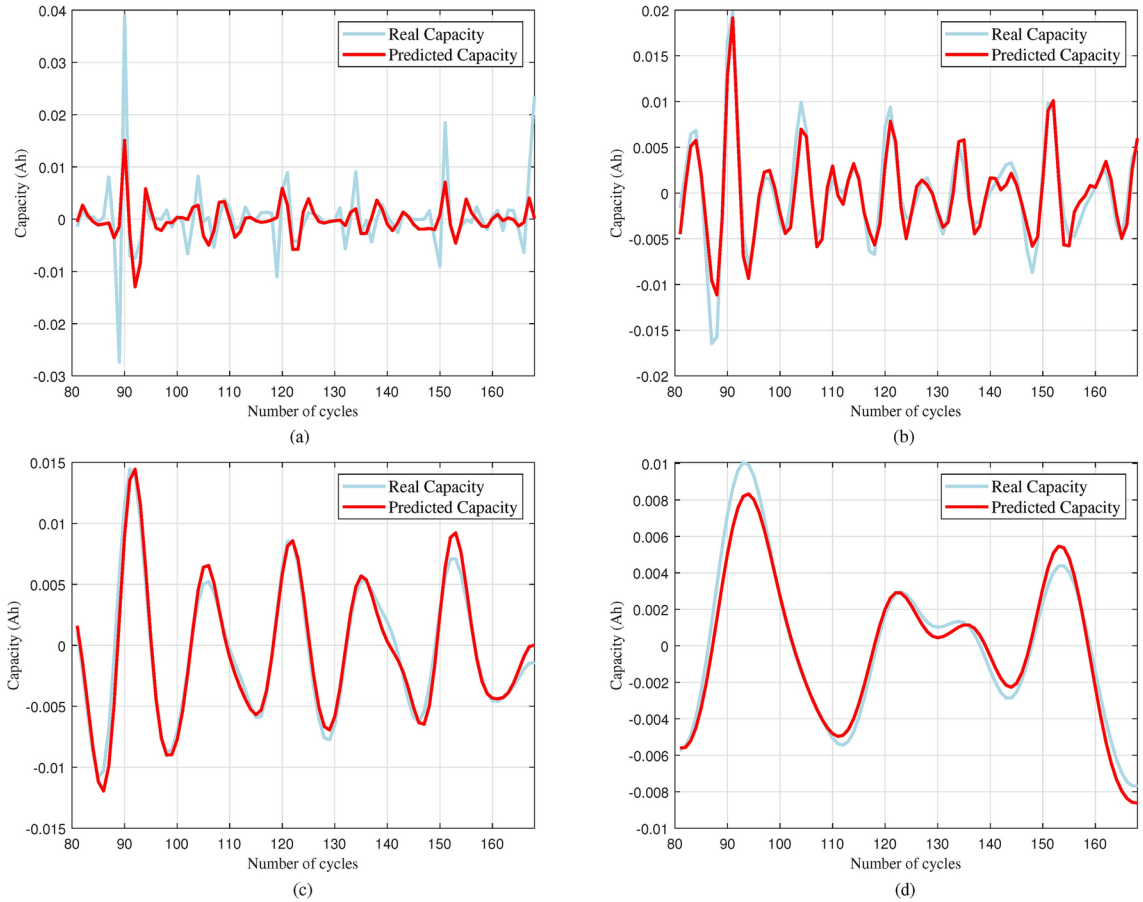

Fig. 6. LSTM network predictions for IMF1-IMF4 component of B0005 Battery: (**a**) IMF1, (**b**) IMF2, (**c**) IMF3 and (**d**) IMF4.

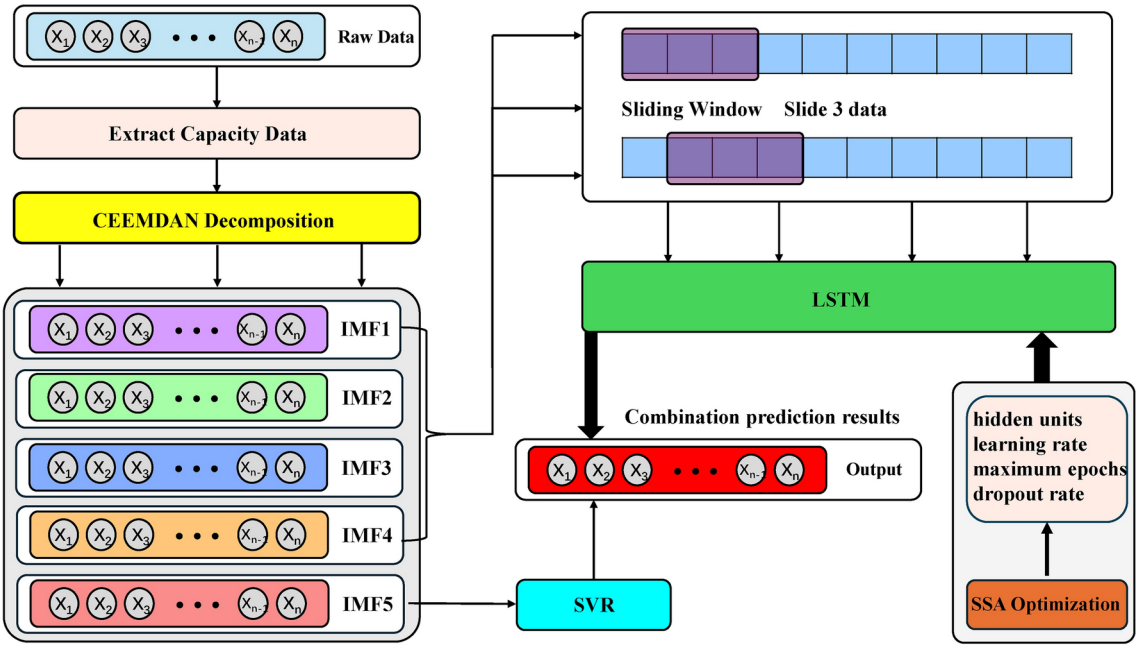

Fig. 7. Prediction process diagram.

and local fluctuations through IMF1-IMF4. The combination of these two models enables accurate tracking of both long-term degradation trends and short-term capacity fluctuations, thereby enhancing the overall accuracy of RUL prediction. The prediction results were subsequently analyzed using evaluation metrics. The schematic diagram of the whole experimental process is shown in Fig. 7.

Evaluation metrics

To evaluate the performance of the proposed method, the root mean square error (RMSE), mean absolute error (MAE), mean relative error (MRE), and the coefficient of determination R^2 were selected as evaluation metrics. The closer the RMSE, MAE, or MRE are to zero, the higher the accuracy of the proposed method. The closer R^2 is to 1, the more closely the predicted capacity values match the actual capacity values, indicating a better model fit. The specific calculation formulas are as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i}^{n}(\hat{y}_i - y_i)^2} \tag{20}$$

$$MAE = \frac{1}{n}\sum_{i}^{n}|\hat{y}_i - y_i| \tag{21}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{22}$$

$$MRE = \frac{1}{n}\sum_{i}^{n}\frac{|\hat{y}_i - y_i|}{y_i} \times 100\% \tag{23}$$

where $\hat{y}$ represents the predicted capacity value, y represents the actual capacity value, and i denotes the cycle number.

Additionally, the error between the predicted and actual cycle count when the lithium-ion battery capacity declines to the failure threshold is defined as follows:

$$E_r = |P_{RUL} - R_{RUL}| \tag{24}$$

$$PE_r = \frac{|P_{RUL} - R_{RUL}|}{R_{RUL}} \times 100\% \tag{25}$$

where P_{RUL} and R_{RUL} represent the predicted and actual values of RUL. E_r and PE_r denote the absolute error and relative error between the two values.

Experimental results and analysis

Comparison of battery RUL prediction results of different methods

In order to prove the effectiveness of the SVR-LSTM hybrid method proposed in this paper in accurately predicting the RUL direction of the battery, the method is compared with the non-hybrid model. The method proposed in this paper is denoted as M1, the CEEMDAN-SVR method is denoted as M2, the CEEMDAN-SSA-LSTM method is denoted as M3, the single SVR method is denoted as M4 and the single LSTM network prediction method is denoted as M5. The above five methods are applied to B0005, B0006, B0007 and B0018 battery data for experimental verification. The first 80 data points of the data set are selected to train the model, and then the last 88 data are predicted. Because B0018 battery has only 132 data points, the first 60 data points are selected as the training set. The results predicted by the five methods are shown in Fig. 8.

Observation of Fig. 8 shows that the predicted capacity curves of the hybrid SVR and LSTM model proposed in this paper are closer to the actual capacity curves as compared to the other methods. The SVR model captures the overall attenuation trend of battery capacity, but the fluctuation of the prediction curve is insufficient, indicating that it has limitations in describing the phenomenon of battery capacity regeneration. This observation is consistent with the inherent characteristics of SVR as it is mainly good at modeling linear and smooth time series data. The LSTM model shows better ability in capturing regenerative phenomena and local fluctuations in battery capacity. However, it deviates slightly from the actual degradation trend. In contrast, the proposed hybrid approach takes advantage of the complementary strengths of the two models: the low-frequency component representing the dominant degradation trend (IMF5) is efficiently modeled by SVR, while the SSA-optimized LSTM accurately captures the high-frequency component characterizing capacity regeneration and local variations (IMF1-IFM4) This decomposition-based ensemble strategy significantly enhances the model 's ability to predict global trends and local fluctuations in battery capacity degradation.

In order to quantitatively assess the prediction performance, four evaluation metrics, RMSE, MAE, MRE and R^2, were used for a comprehensive comparison, as shown in Fig. 9a-d, and the specific values are shown in Table 4. The proposed hybrid method M1 shows excellent performance in all metrics and test cases, with significant improvements over other models. As shown in Fig. 9a, the red bar representing the M1 RMSE value is significantly shorter than the other colored bars for all batteries tested, ranging from 0.0066 to 0.0086 Ah, which is an average reduction of 66.41% compared to the traditional SVR model M5. For the B0018 battery with large fluctuations in capacity data, the RMSE of M1 is 0.0086 Ah, which is 79.43% lower than the RMSE of M5 of 0.0418 Ah, indicating that the proposed method effectively captures the capacity degradation characteristics.

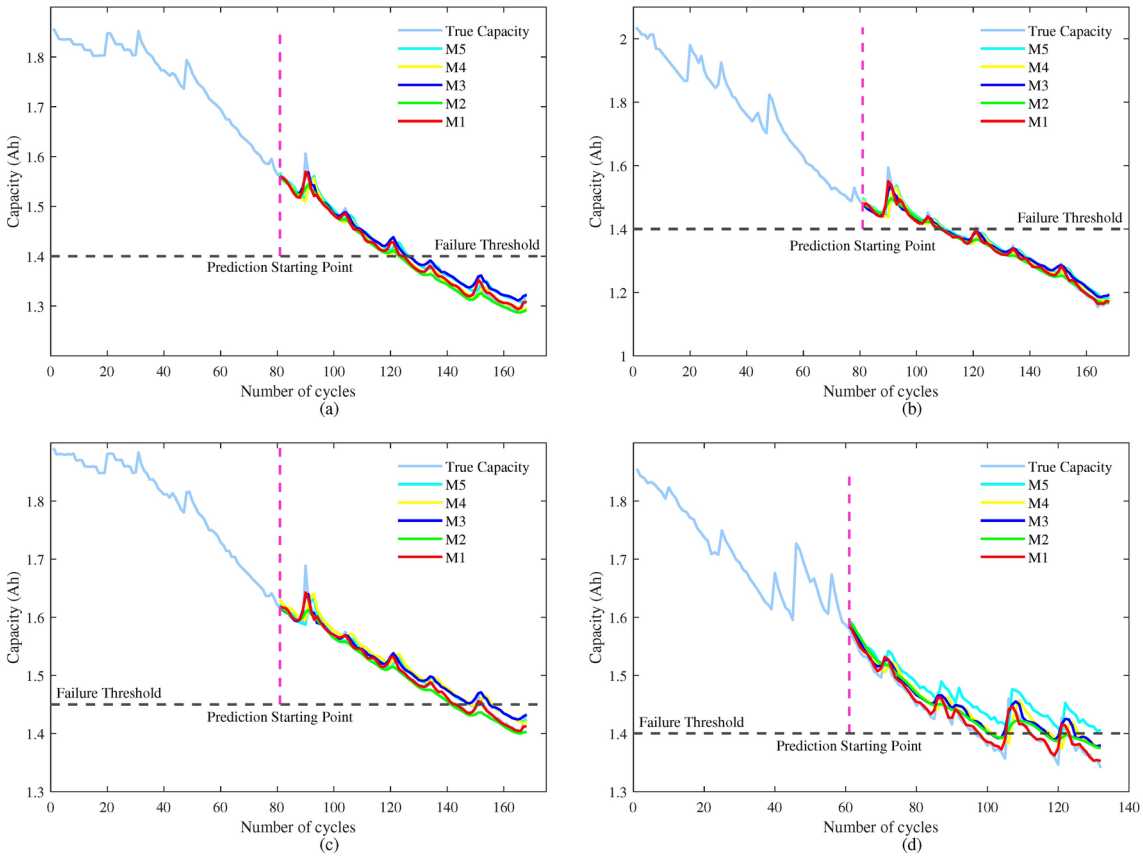

Fig. 8. Prediction results of different methods under 80 training data: (**a**) B0005, (**b**) B0006, (**c**) B0007 and (**d**) B0018.

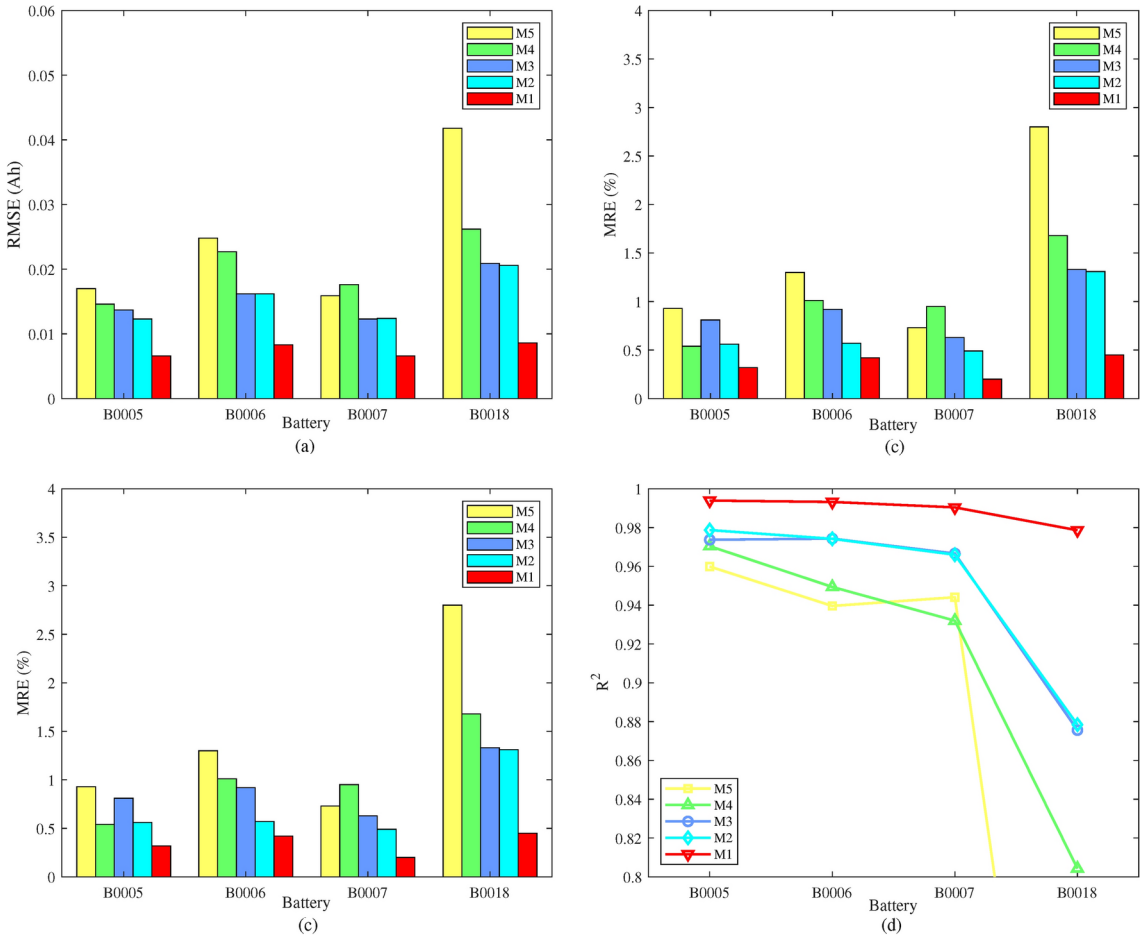

Fig. 9. Evaluation metrics for different prediction methods with 80 training data: (**a**) RMSE, (**b**) MAE, (**c**) MRE and (**d**) R^2.

Similarly, in the MAE comparison, M1 consistently shows the shortest bars and maintains a low error between 0.0031 and 0.0065 Ah, with an average decrease of 71.74% compared to M5. The MRE results shown in Fig. 9c, where M1 has significantly shorter red bars, exhibit excellent stability, with errors consistently below 0.45% and an average reduction of 72.45% compared to the other models. In the line graph of R^2 values, the curve for M1 clearly maintains the highest position across all test cases, with coefficients consistently above 0.97. This excellent performance is most evident in battery B0018, where M1 maintains a high R^2 value of 0.9786, while the curves for the other models show a clear downward trend, especially for M5 which drops significantly to 0.5012. In addition, the absolute error between the proposed method and the actual value of the battery RUL prediction is less than one cycle, and the relative error is also lower than other methods compared.

A comprehensive evaluation shows that the proposed hybrid approach successfully integrates the complementary advantages of SVR and LSTM models while effectively addressing their respective limitations. Comparison of the errors in the prediction results of various methods confirms the effectiveness of the proposed method in accurately predicting capacity and RUL.

To further validate the effectiveness and robustness of the proposed hybrid model, we extended our experiments by increasing the training dataset size. The previous analysis used 80 capacity cycles as training data, and now we want to check the performance of the model when training on 100 cycles (B0018 is 60 cycles). The results of the five methods of prediction are shown in Fig. 10, and the specific values of the four evaluation metrics are shown in Table 6. Analyzing Fig. 10, it can be seen that for the three batteries B0005, B0006 and B0007, the addition of training data makes the prediction curves smoother, especially at the early stage of prediction when the fluctuation of each method is significantly reduced. For battery B0018, although there are large fluctuations in its capacity decay curve, the M1 method still tracks this nonlinear change characteristic better, while the other methods show different degrees of divergence in the late prediction period. This indicates that adding training data further improves the prediction stability and robustness of the proposed method. From Table 6, it can be seen that the hybrid method (M1) proposed in this paper achieves the best performance in all evaluation metrics. For the B0005 battery, the RMSE of M1 is 0.0044 Ah, which is 71.24% lower than M5, and for the more fluctuating B0018 battery, the RMSE of M1 is 0.0086 Ah, which is 61.43 lower than M5. The R^2 values of the M1 method for all the tested batteries are all above 0.96, and the RUL prediction error is 0 or 1 cycle, which validates the proposed method's prediction accuracy and stability of the proposed method.

Battery	Method	RMSE	MAE	MRE (%)	R^2	E_r	PE_r (%)
B5	M5	0.0170	0.0129	0.93	0.9599	2	4.55
	M4	0.0146	0.0077	0.54	0.9704	4	9.09
	M3	0.0137	0.0111	0.81	0.9737	2	4.55
	M2	0.0123	0.0079	0.56	0.9787	1	2.27
	M1	0.0066	0.0046	0.32	0.9939	0	0
B6	M5	0.0248	0.0174	1.30	0.9396	1	3.45
	M4	0.0227	0.0135	1.01	0.9494	0	0
	M3	0.0162	0.0120	0.92	0.9743	0	0
	M2	0.0162	0.0079	0.57	0.9742	1	3.45
	M1	0.0083	0.0057	0.42	0.9932	0	0
B7	M5	0.0159	0.0110	0.73	0.9441	5	7.94
	M4	0.0176	0.0143	0.95	0.9320	5	7.94
	M3	0.0123	0.0094	0.63	0.9666	5	7.94
	M2	0.0124	0.0074	0.49	0.9660	2	3.17
	M1	0.0066	0.0031	0.20	0.9904	0	0
B18	M5	0.0418	0.0395	2.80	0.5012	24	64.86
	M4	0.0262	0.0239	1.68	0.8043	5	13.51
	M3	0.0209	0.0188	1.33	0.8756	3	8.11
	M2	0.0206	0.0187	1.31	0.8784	4	10.81
	M1	0.0086	0.0065	0.45	0.9586	1	2.72

Table 4. Error comparison between this method and other methods under 80 training data.

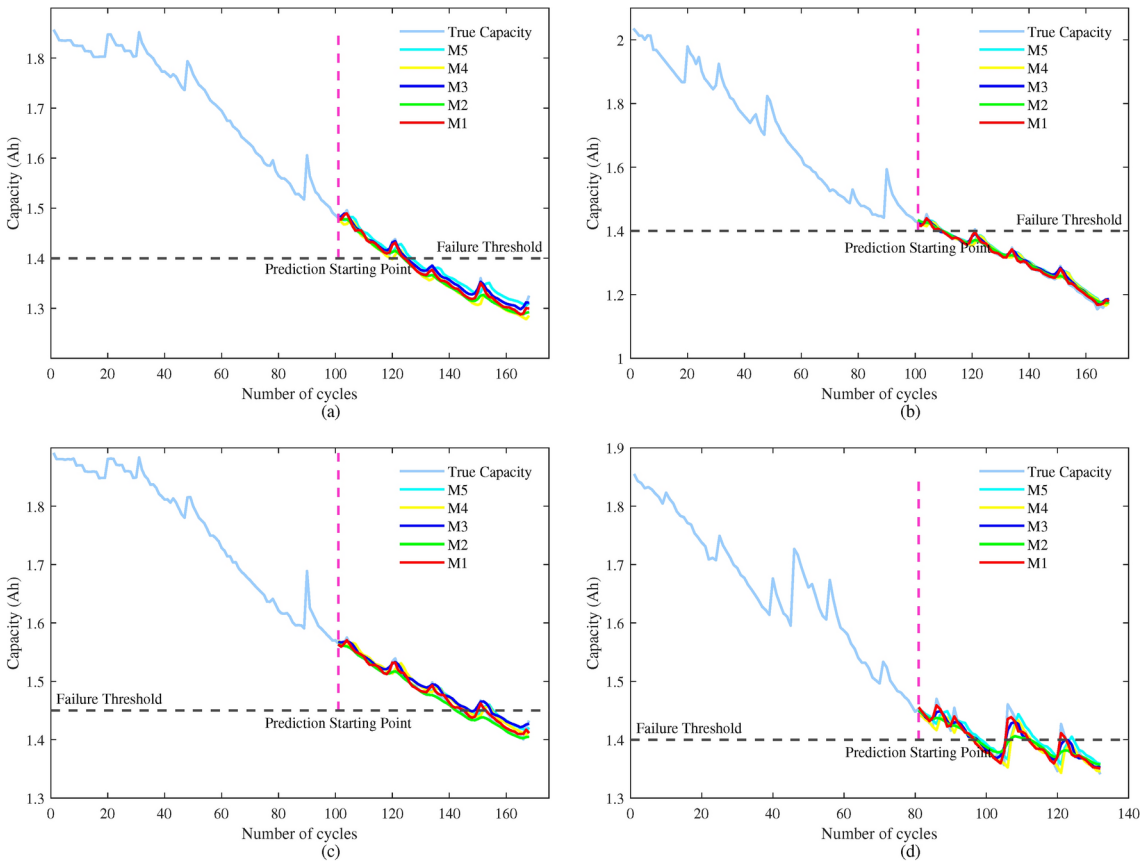

Fig. 10. Prediction results of different methods under 100 training data: (**a**) B0005, (**b**) B0006, (**c**) B0007 and (**d**) B0018.

Model	RMSE	MAE	R^2
IHSSA-LSTM-TCN[36]	0.0111	0.0816	–
PSO-LSTM[37]	0.0162	0.0095	0.9136
SVR-MC[38]	0.0063	–	0.9920
PF-SVR[39]	0.0146	0.1200	–
Differential Voltage-ELM[40]	0.0051	0.0043	–
PCA-CNN-BiLSTM[41]	0.0114	0.0090	–
Model of this paper	0.0043	0.0026	0.9941

Table 5. Comparison of prediction results of different methods based on B0005 battery.

Battery	Method	RMSE	MAE	MRE (%)	R^2	E_r	PE_r (%)
B5	M5	0.0153	0.0139	1.02	0.9303	3	12.50
	M4	0.0133	0.0086	0.63	0.9474	1	4.17
	M3	0.0071	0.0062	0.46	0.9851	1	4.17
	M2	0.0091	0.0060	0.44	0.9754	0	0
	M1	0.0044	0.0026	0.19	0.9942	0	0
B6	M5	0.0161	0.0137	1.06	0.9592	1	11.11
	M4	0.0152	0.0118	0.92	0.9638	0	0
	M3	0.0093	0.0078	0.61	0.9866	0	0
	M2	0.0106	0.0083	0.64	0.9823	1	11.11
	M1	0.0055	0.0044	0.35	0.9952	0	0
B7	M5	0.0116	0.0103	0.70	0.9428	4	9.30
	M4	0.0104	0.0090	0.61	0.9540	3	6.98
	M3	0.0103	0.0090	0.62	0.9554	4	9.30
	M2	0.0082	0.0052	0.35	0.9718	1	2.33
	M1	0.0045	0.0033	0.23	0.9916	1	2.33
B18	M5	0.0223	0.0154	1.09	0.5475	1	5.88
	M4	0.0233	0.0116	0.81	0.8182	1	5.88
	M3	0.0121	0.0075	0.53	0.8881	1	5.88
	M2	0.0179	0.0117	0.83	0.8799	1	5.88
	M1	0.0086	0.0060	0.43	0.9628	1	5.88

Table 6. Error comparison between this method and other methods under 100 training data.

Through comparative experiments, we verified the effectiveness of using SVR for low-frequency components and LSTM for high-frequency components. As shown in Tables 4 and 6, the hybrid method M1 consistently outperforms M2 (CEEMDAN-SVR) and M3 (CEEMDAN-SSA-LSTM). In the case of cell B0005 (100 training cycles), for example, the RMSE of M1 was 0.0044 Ah, which is significantly lower than M2's 0.0091 Ah (using SVR only) and M3's 0.0071 Ah (using LSTM only). The excellent performance of all test cells suggests that our hybrid strategy effectively utilizes the capabilities of SVR in capturing smooth trends and the advantages of LSTM in simulating complex temporal patterns.

Performance comparison with other existing methods

To further validate the effectiveness of the proposed method in RUL prediction, the results of this paper's method compared with other methods using the same NASA B0005 battery dataset are given in Table 5. As can be seen from the table, the method proposed in this paper achieves better performance in all evaluation metrics: the RMSE is 0.0044 Ah, the MAE is 0.0026 Ah, and the R^2 reaches 0.994 when using 100 training data. Compared with the existing methods, the prediction accuracy of the proposed method has been significantly improved, which verifies the CEEMDAN decomposition-based SVR-LSTM hybrid strategy is superior in battery RUL prediction.

Conclusions

In this paper, a RUL prediction method for lithium-ion batteries based on CEEMDAN decomposition and combining SVR and SSA-LSTM networks is proposed. The method firstly employs CEEMDAN to decompose the battery capacity sequence into high-frequency and low-frequency components, and then uses SVR and SSA-optimized LSTM networks to predict the different frequency components, respectively. The validity of the proposed method is verified using NASA public dataset and the main contributions are as follows:

(1) Aiming at the problem of prediction accuracy degradation caused by capacity regeneration during lithium-ion battery degradation, a CEEMDAN decomposition method is proposed to decompose the degradation curve into high-frequency and low-frequency IMF components. Among them, the high frequency component captures the capacity fluctuation and regeneration characteristics, and the low frequency component maintains the overall degradation trend of the battery. The decomposition strategy not only effectively alleviates the noise interference caused by capacity regeneration, but also improves the stability of the data, thereby enhancing the prediction accuracy of the model.
(2) A hybrid SVR-LSTM-based prediction strategy was designed. The low-frequency components characterizing the main degradation trends are modeled using the SVR model, and the LSTM network is employed to capture the nonlinear and long- and short-term dependent features in the high-frequency components. The two models complement each other's strengths and significantly improve the prediction of capacity degradation trends and local fluctuations. The experimental results show that the RMSE of the proposed method is reduced by 66.41% on average, and the R^2 value remains above 0.96.
(3) The SSA algorithm is introduced to optimize the LSTM network, and a comprehensive prediction model is proposed. This method adopts intelligent search strategy, which not only reduces the workload and inefficiency of manual parameter adjustment, but also effectively prevents the model from falling into local optimum in the training process, thus improving the overall prediction accuracy.

The experimental results show that the proposed hybrid method has higher prediction accuracy and stability than the existing methods. On the B0005 battery data set, when 100 training data are used, the RMSE of the method is 0.0044 Ah, the MAE is 0.0026 Ah, and the R^2 reaches 0.994, which verifies the effectiveness of the method. This method provides a reliable solution for RUL prediction of lithium-ion batteries.

Data availibility

The data that support the findings of this study are available on request from the corresponding author, JP.Z, upon reasonable request.

Received: 19 January 2025; Accepted: 26 February 2025
Published online: 10 March 2025

References

1. Rao, Z., Wu, J., Li, G. & Wang, H. Voltage abnormity prediction method of lithium-ion energy storage power station using informer based on Bayesian optimization. *Sci. Rep.* **14**, 21404 (2024).
2. Li, Z., Shen, S., Ye, Y., Cai, Z. & Zhen, A. An interpretable online prediction method for remaining useful life of lithium-ion batteries. *Sci. Rep.* **14**, 12541 (2024).
3. Zhao, F.-M., Gao, D.-X., Cheng, Y.-M. & Yang, Q. Application of state of health estimation and remaining useful life prediction for lithium-ion batteries based on at-cnn-bilstm. *Sci. Rep.* **14**, 29026 (2024).
4. Iftikhar, M. et al. A deep learning approach to optimize remaining useful life prediction for li-ion batteries. *Sci. Rep.* **14**, 25838 (2024).
5. Liu, C., Wang, Y. & Chen, Z. Degradation model and cycle life prediction for lithium-ion battery used in hybrid energy storage system. *Energy* **166**, 796–806 (2019).
6. Li, B., Han, X., Zhang, W., Zeng, W. & Wu, J. Review of the remaining useful life prediction methods for lithium-ion batteries. *Energy Storage Sci. Technol.* **13**, 1266 (2024).
7. Sadabadi, K. K., Jin, X. & Rizzoni, G. Prediction of remaining useful life for a composite electrode lithium ion battery cell using an electrochemical model to estimate the state of health. *J. Power Sources* **481**, 228861 (2021).
8. Liu, Q., Zhang, J., Li, K. & Lv, C. The remaining useful life prediction by using electrochemical model in the particle filter framework for lithium-ion batteries. *IEEE Access* **8**, 126661–126670 (2020).
9. Hell, S. M. & Kim, C. D. Development of a data-driven method for online battery remaining-useful-life prediction. *Batteries* **8**, 192 (2022).
10. Hu, X., Xu, L., Lin, X. & Pecht, M. Battery lifetime prognostics. *Joule* **4**, 310–346 (2020).
11. Yao, J., Neupert, S. & Kowal, J. Cross-stitch networks for joint state of charge and state of health online estimation of lithium-ion batteries. *Batteries* **10**, 171 (2024).
12. Kang, W. et al. Research on remaining useful life prognostics based on fuzzy evaluation-gaussian process regression method. *IEEE Access* **8**, 71965–71973 (2020).
13. Lin, M., Zeng, X. & Wu, J. State of health estimation of lithium-ion battery based on an adaptive tunable hybrid radial basis function network. *J. Power Sources* **504**, 230063 (2021).
14. Yao, J. & Kowal, J. A multi-scale data-driven framework for online state of charge estimation of lithium-ion batteries with a novel public drive cycle dataset. *J. Energy Storage* **107**, 114888 (2025).
15. Zhao, J. et al. Method of predicting SOH and RUL of lithium-ion battery based on the combination of LSTM and GPR. *Sustainability* **14**, 11865 (2022).
16. Chen, C. & Chen, D. Research on indirect prediction of lithium battery RUL based on CNN-LSTM. *J. Power Sources* **45**, 589–594 (2021).
17. Yu, D., Cheng, J. & Yang, Y. Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings. *Mech. Syst. Signal Process.* **19**, 259–270 (2005).
18. Cheng, G., Wang, X. & He, Y. Remaining useful life and state of health prediction for lithium batteries based on empirical mode decomposition and a long and short memory neural network. *Energy* **232**, 121022 (2021).
19. Li, X., Zhang, L., Wang, Z. & Dong, P. Remaining useful life prediction for lithium-ion batteries based on a hybrid model combining the long short-term memory and elman neural networks. *J. Energy Storage* **21**, 510–518 (2019).
20. Yang, Z., Wang, Y. & Kong, C. Remaining useful life prediction of lithium-ion batteries based on support vector regression optimized and grey wolf optimizations. In *International Symposium on Artificial Intelligence and Robotics 2021*, 11884, 1188402 (SPIE, 2021).
21. Zhao, Q., Qin, X., Zhao, H. & Feng, W. A novel prediction method based on the support vector regression for the remaining useful life of lithium-ion batteries. *Microelectron. Reliab.* **85**, 99–108 (2018).
22. Deléchelle, E., Lemoine, J. & Niang, O. Empirical mode decomposition: An analytical approach for sifting process. *IEEE Signal Process. Lett.* **12**, 764–767 (2005).

23. Cao, J., Li, Z. & Li, J. Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A* **519**, 127–139 (2019).
24. Liu, L., Sun, W., Yue, C., Zhu, Y. & Xia, W. Remaining useful life estimation of lithium-ion batteries based on small sample models. *Energies* **17**, 4932 (2024).
25. Lee, C.-J., Kim, B.-K., Kwon, M.-K., Nam, K. & Kang, S.-W. Real-time prediction of capacity fade and remaining useful life of lithium-ion batteries based on charge/discharge characteristics. *Electronics* **10**, 846 (2021).
26. Lin, Y. et al. Forecasting the realized volatility of stock price index: A hybrid model integrating CEEMDAN and LSTM. *Expert Syst. Appl.* **206**, 117736 (2022).
27. Lv, K., Ma, Z., Bao, C. & Liu, G. Indirect prediction of lithium-ion battery RUL based on CEEMDAN and CNN-BIGRU. *Energies* **17**, 1704 (2024).
28. Xue, J. & Shen, B. A novel swarm intelligence optimization approach: Sparrow search algorithm. *Syst. Sci. Control Eng.* **8**, 22–34 (2020).
29. Gao, X. et al. Short-term wind power forecasting based on SSA-VMD-LSTM. *Energy Rep.* **9**, 335–344 (2023).
30. Ouyang, M. & Shen, P. Prediction of remaining useful life of lithium batteries based on WOA-VMD and LSTM. *Energies* **15**, 8918 (2022).
31. Dong, H., Jin, X., Lou, Y. & Wang, C. Lithium-ion battery state of health monitoring and remaining useful life prediction based on support vector regression-particle filter. *J. Power Sources* **271**, 114–123 (2014).
32. Xu, Z., Gao, Y. & Jin, Y. Application of an optimized SVR model of machine learning. *Int. J. Multimed. Ubiquitous Eng.* **9**, 67–80 (2014).
33. Bo, Q., Wang, X. & Liu, K. Minimum frequency prediction of power system after disturbance based on the v-support vector regression. In *2014 International Conference on Power System Technology*, 614–619 (IEEE, 2014).
34. Song, Y., Cai, C., Ma, D. & Li, C. Modelling and forecasting high-frequency data with jumps based on a hybrid nonparametric regression and lstm model. *Expert Syst. Appl.* **237**, 121527 (2024).
35. Rundo, F. Deep LSTM with reinforcement learning layer for financial trend prediction in FX high frequency trading systems. *Appl. Sci.* **9**, 4460 (2019).
36. Qiu, S., Zhang, B., Lv, Y., Zhang, J. & Zhang, C. A lithium-ion battery remaining useful life prediction model based on CEEMDAN data preprocessing and HSSA-LSTM-TCN. *World Electr. Veh. J.* **15**, 177 (2024).
37. Gong, Y. et al. State-of-health estimation of lithium-ion batteries based on improved long short-term memory algorithm. *J. Energy Storage* **53**, 105046 (2022).
38. Cui, J. et al. Remaining useful life prediction of aviation lithium-ion battery based on svr-mc. In *2022 34th Chinese Control and Decision Conference (CCDC)*, 505–510 (IEEE, 2022).
39. Wei, J., Dong, G. & Chen, Z. Remaining useful life prediction and state of health diagnosis for lithium-ion batteries using particle filter and support vector regression. *IEEE Trans. Industr. Electron.* **65**, 5634–5643 (2017).
40. LI, L. et al. Rul prediction of lithium-ion battery based on differential voltage and Elman neural network.. *Energy Storage Sci. Technol.* **10**, 2373 (2021).
41. Feng, J., Cai, F., Li, H., Huang, K. & Yin, H. A data-driven prediction model for the remaining useful life prediction of lithium-ion batteries. *Process Saf. Environ. Prot.* **180**, 601–615 (2023).

Acknowledgements

This research was funded by the Longyan-Xiamen University Institute of Industry and Education Integration (2024ZG011).

Author contributions

Yibiao Fan: Conceptualization of this study, Methodology, Software, Writing-original draft. Zhishan Lin: Conceptualization, Data curation, Writing-review & editing. Fan Wang: Modification for the final layout. Jianpeng Zhang: Conceptualization, Writing-original draft.

Funding

This research was funded by the Longyan-Xiamen University Institute of Industry and Education Integration (2024ZG011).

Declarations

Competing interests

The authors declare that there are no conflicts of interest.

Additional information

Correspondence and requests for materials should be addressed to J.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.