Discover Energy

Research

Remaining useful life prediction of lithium-ion batteries combined with SVD-SDAE and support vector quantile regression

Lin Sun[1] · Xiaojie Huang[2] · Jing Liu[2] · Jing Song[2]

Received: 3 December 2023 / Accepted: 7 March 2024
Published online: 27 March 2024

Abstract
Lithium-ion batteries are important energy storage materials, and the prediction of their remaining useful life has practical importance. Since traditional feature extraction methods depend on parameter settings and have poor adaptability, singular value decomposition was used to extract 15 health indicators from the degradation data of lithium-ion batteries. To eliminate redundancy among the extracted health indicators, Spearman correlation analysis was subsequently used to determine the most appropriate health indicators. On this basis, the selected health indicators were processed by the denoising stack autoencoder, and a fusion health indicator was obtained. Finally, the support vector quantile regression model was used to predict the battery capacity interval by the fusion health indicator. The National Aeronautics and Space Administration battery dataset and Massachusetts Institute of Technology battery dataset were used to verify the validity and generalizability of our proposed model, and our proposed model was compared with the existing four classical prediction models. The experimental results showed that our proposed prediction model had higher prediction accuracy and better robustness than the other models and could effectively improve the prediction effect of the remaining useful life of batteries. The mean value of the root mean square error of the predicted results using our proposed model remained within 1.3%, and the mean value of the coefficient of determination was above 0.97.

1 Introduction

Due to their advantages of high operating voltage, high power density, and long lifespan, lithium-ion batteries have become staple energy storage devices for many applications, including electric vehicles, smart grids, and energy storage systems [1, 2]. However, as the frequency of charging and discharging increases, the battery will gradually degrade or even become deactivated through different mechanisms [3]. Therefore, the study of the degradation of batteries is highly important, and reliable degradation models need to be established to accurately predict their remaining useful life (P).

In recent years, data-driven methods have drawn much attention in the field of lithium-ion battery RUL prediction due to the ability to directly predict battery degradation trends using historical monitoring data without relying on complex physical models [4]. Most current studies define capacity or internal resistance as a health indicator (HI) [5, 6]. However, the detection of capacity or internal resistance requires complex experimental conditions and instrumentation [7], and realizing online collection is difficult. Therefore, the extraction of HIs, which can be easily calculated and can quantify the degradation state of a battery, has great practical value. Khaleghi et al. [8] set the upper and lower voltage boundaries and specified a fixed sampling rate to extract HIs from partial charging voltage profiles and used them to estimate the health state of a battery. In addition, current variations during short discharge intervals [9] and time differences between fixed voltage intervals [10] can

✉ Lin Sun, hulongzn2002@163.com | [1]Basic Science Department, Wenhua College, Wuhan, China. [2]College of Information and Communication, National University of Defense Technology, Changsha, China.

also be used as features to reflect the health state of Li-ion batteries. Although these methods show excellent performance in extracting HIs, they require the setting of parameters such as fixed time or voltage, which not only relies on human experience, but also has the potential to complicate feature extraction [8–10]. Liu et al. [11] set different parameters to extract the same HIs and perform RUL prediction on battery datasets from the National Aeronautics and Space Administration (NASA) and CALCE, respectively, and the results indicated that the requirements of parameters for feature extraction could be different for different Li-ion batteries. Pang et al. [12] extracted the peak values of more sensitive IC curves and the under-peak area as the HIs and predicted RUL on the lithium-ion battery dataset provided by NASA. Pan et al. [13] extracted IC curves with two recognizable peaks on the battery dataset provided by CALCE. This feature extraction method was clearly not applicable to different Li-ion battery datasets. Although the feature extraction methods proposed in references [11–13] can be used to extract HIs, they may not be applicable to other lithium-ion battery datasets due to differences in battery parameters or experimental environments. As mentioned above, traditional feature extraction methods usually require parameter setting and are poorly adaptable; therefore, the first objective of this work is to find a feature extraction method that reduces or avoids parameter setting and is widely applicable to different lithium-ion battery datasets.

Based on the above analysis, a method for predicting the RUL interval of lithium-ion batteries is proposed based on adaptive health indicator extraction and support vector quantile regression. The main improvements of our proposed method are as follows:

First, singular value decomposition (SVD) [14] is used to extract the most important singular vector from the original measurement data as a health indicator of lithium-ion batteries. The original measurement data from the various types of lithium-ion batteries are distributed differently and may contain noise. The raw data are directly used as a health indicator of lithium-ion batteries; this process affects the prediction accuracy of the remaining useful life and leads to a lack of generality of the prediction model. Since SVD does not involve parameter setting in the decomposition process, the calculation process is simple, and the extracted features are very efficient. Therefore, in our proposed model, SVD is used to extract features from the original measurement data of lithium-ion batteries, and the extracted maximum singular vector is used as the degradation feature of the batteries.

Next, Spearman correlation analysis was used to screen the HIs extracted by SVD, and the screened HIs were fused by a stacked denoising autoencoder (SDAE) [15] to construct a fusion health indicator with stronger representation ability. A certain degree of redundancy exists between the HI extracted by SVD, and when the HI dimension is too high, the model can be more complex, affecting the accuracy of the RUL prediction. Therefore, Spearman correlation analysis was used to analyze each HI, and the HI that was highly correlated with the battery capacity sequence was retained. The retained HIs were fused with the SDAE to obtain a fusion health indicator with a stronger correlation.

Finally, the support vector quantile regression model (SVQR)[16] was used to perform interval prediction on the RUL of lithium-ion batteries. In the process of actual use, due to the complex working environment and large load fluctuations of lithium-ion batteries, the degradation mode of lithium batteries is not fixed; thus, the results from the RUL point estimation are uncertain. In this study, a support vector quantile regression model is used to predict the RUL interval of lithium batteries, and this model produces more generalizable prediction results. The experimental results show that our proposed SVQR model has high interval estimation accuracy and can be applied to different types of data.

The effectiveness of our proposed model was validated using lithium-ion battery datasets provided by the National Aeronautics and Space Administration research center [17] and the Massachusetts Institute of Technology (MIT) [18]. Our proposed method was compared with four existing classical prediction methods. The experimental results showed that the prediction accuracy of our proposed method was better than that of the existing four classical prediction methods. The remainder of this paper is arranged as follows. Section 1 briefly introduces the basic theory of stack denoising autoencoders. In Sect. 2, an adaptive health indicator extraction method based on SVD, Spearman correlation analysis and the SDAE is introduced. Section 3 introduces the RUL interval prediction method based on support vector quantile regression. Section 4 describes the experiment and analysis, and Sect. 5 provides the conclusion and future prospects.

2 Stack denoising autoencoder

A stacked denoising autoencoder, as the name suggests, is constructed by stacking multiple denoising autoencoders. Based on the autoencoder, the denoising autoencoder forces the network model to learn denoising and restore the input data by randomly shielding some training input data.

The traditional denoising autoencoder consists of a layer of encoder and a layer of decoder. If unsupervised pretraining is performed on each single hidden layer denoising autoencoder unit and then stacked, the output of the previous

encoder is used as the input of the next autoencoder; finally, reverse tuning training is performed, and an SDAE with a multilayer structure is obtained. Compared with traditional denoising autoencoders, a stacked denoising autoencoder has a deeper network structure and stronger feature learning ability and generalization ability [19].

The training of the stacked denoising autoencoder involves making the output signal Z of each denoising autoencoder infinitely close to the input signal X and defining the error function as shown in Eq. (1):

$$L_D = ||\mathbf{X} - g(f(\mathbf{X}_1))|| \tag{1}$$

Unsupervised learning for pretraining is used to determine the parameters $\boldsymbol{\theta} = (\mathbf{W}, \mathbf{b})$ of each denoising autoencoder. Then, supervised learning training is carried out on the entire stacked denoising autoencoder by the back propagation (BP) algorithm to fine-tune the global parameters $\boldsymbol{\theta} = (\mathbf{W}, \mathbf{b})$. Finally, the model is ensured to have good discriminant performance and effectively extract multidimensional features of the input signals.

3 Adaptive extraction of the health indicators via SVD and the SDAE

3.1 Experimental data

The RUL of a lithium-ion battery is defined as the number of charge and discharge cycles the battery goes through when its capacity drops to 70–80% of its initial value. Therefore, predicting the RUL of lithium-ion batteries essentially involves predicting the capacity value of the battery after each round of charge and discharge.

The selected battery degradation dataset 1 is from the NASA Research Center [18]. The dataset contains a collection of degradation state data for four 18,650-type Li-ion batteries in the same experimental environment. The nominal capacity of the four batteries is 2.2 Ah, and the nominal voltage is 3.7 V. In the charging stage, the batteries are charged in constant current (CC) mode at 1.5 A until the battery voltage reaches 4.2 V and then continue to be charged in constant voltage (CV) mode until the charging current drops to 20 mA. In the discharging stage, the battery is discharged with 2A CC until the voltages of the B0005, B0006, B0007 and B0018 batteries dropped to 2.7 V, 2.5 V, 2.2 V and 2.5 V, respectively. Figure 1 shows the actual decay curve of the battery capacity.

3.2 HI extraction based on SVD

In this section, the singular value decomposition [20] method is initially introduced and employed for feature extraction. SVD is a matrix decomposition method that allows complex matrices to be represented as multiplications of smaller and simpler submatrices.

Assuming that matrix A is a matrix of size $p \times q$, the SVD of matrix A is defined as shown in Eq. (2):

$$A = U \sum V^T \tag{2}$$

The matrix U has a rank r. U is an orthogonal matrix of size $p \times q$. The column vectors of U are called left singular vectors. $\sum$ is a diagonal matrix of size $p \times q$ with all elements being 0 except those on the main diagonal. Each element on the main diagonal of $\sum$ is a singular value of A, which is denoted as σi ($i = 1,\ldots,r$). V is an orthogonal matrix of size $q \times q$.

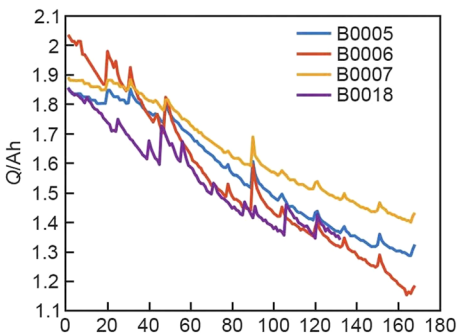

Fig. 1 Capacity degradation curves of four batteries

The column vectors of V are right singular vectors. Multiplying A by its transposed matrix A^T yields Eq. (3), and similarly, Eq. (4) is associated with AA^T,

$$AA^T = V\sum U^T U\sum V^T = V\sum^2 V^T \tag{3}$$

$$AA^T = U\sum VV^T\sum U^T = U\sum^2 U^T \tag{4}$$

The eigenvectors of AA^T and AA^T form the matrices V and U in the SVD, respectively. Because the eigenvalues of AA^T and AA^T are $\lambda_1,\ldots,\lambda_r$, as shown in Eq. (5), the singular values of matrix A are denoted as $\sigma_i = \sqrt{\lambda_i}$ $(i=1,\ldots,r)$, as follows:

$$AA^T = A^T A = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_r \end{bmatrix} \tag{5}$$

The matrix singular value is closely related to the vital information hidden in the matrix, where the significance of the information is positively correlated with the singular value. SVD does not involve parameter setting in the matrix decomposition process, which allows simple calculations and simplifies the feature extraction process such that it can avoid dependence on human experience. Therefore, SVD is selected to extract HIs that characterize the degradation state of the battery.

Normally, the parameters, including the voltage, current and temperature, of Li-ion batteries are measurable; thus, the respective singular values are first extracted as 12 HIs from 12 sets of measurable data by SVD (see Table 1 for a detailed description). Moreover, since the degradation information contained in the extracted objects has an important impact on the accuracy of the extracted HIs in quantifying the degradation state of the battery, three additional sets of feature extraction objects are calculated with the matrices dQ/dV, dV/dQ, and dT/dV. A lithium-ion battery is degraded in a dynamic and nonlinear process with various degradation modes, such as mechanical degradation of electrode materials, loss of lithium inventory, and loss of active materials. All of these degradation modes can further interact with each other. dQ/dV, dV/dQ, and dT/dV [21, 22] are often used to analyze the above related degradation modes and thus can be used to extract the features reflecting the degradation state of the batteries, from which HI13, HI14, and HI15 are derived, respectively. dQ/dV, dV/dQ and dT/dV are calculated as follows:

$$\frac{dQ}{dV} = \frac{I\times dt}{dV} = I\times\frac{dt}{dV} \tag{6}$$

Table 1 Spearman correlation coefficients of the 15 potential HIs and capacity of batteries

Feature	HIs	Spearman correlation coefficients			
		B0005	B0006	B0007	B0015
Discharge temperature	HI–1	– 0.7658	– 0.8011	– 0.5607	– 0.2438
Discharge time	HI–2	0.9974	0.9978	0.9983	0.9978
Load current	HI–3	0.9873	0.9986	0.9949	0.9886
Load voltage	HI–4	0.9434	0.9964	0.9050	0.9924
Discharge voltage	HI–5	0.9536	0.9507	0.9407	0.9887
Discharge current	HI–6	– 0.8957	– 0.9967	– 0.7943	– 0.9806
Charging temperature	HI–7	0.9536	0.1558	0.0415	0.1255
Charging interval	HI–8	– 0.8957	– 0.2442	– 0.2396	0.5385
Charging voltage	HI–9	– 0.2388	– 0.9729	– 0.9495	– 0.8211
Charging current	HI–10	– 0.9495	0.9799	0.9570	0.8599
Charging load current	HI–11	0.9908	0.9907	0.9913	0.9775
Charge and discharge time	HI–12	– 0.3560	– 0.3352	– 0.3674	0.9349
dQ/dV	HI–13	0.9968	0.9991	0.9961	0.9957
dV/dQ	HI–14	– 0.5064	– 0.7388	– 0.3494	0.6377
dT/dV	HI–15	0.9844	0.9875	0.9741	0.9936

$$\frac{dV}{dQ} = \frac{dV}{I \times dt} = \frac{1}{I} \times \frac{dV}{dt} \tag{7}$$

$$\frac{dT}{dV} = \frac{dT}{dt} / \frac{dV}{dt} \tag{8}$$

Here, I, V, t and T refer to the current, voltage, time and temperature during discharge, respectively.

For example, in the derivation of HI1 from the discharging voltage of the B0005 battery, the process of using SVD to derive HI from the battery data is as follows: first, the voltage data of all discharging cycles are regarded as matrices and constructed into the set of matrices $\{A(i)\}_{i=1}^{c}$, where c is the number of total discharging cycles. Second, the voltage data of all cycle periods, i.e., matrix $A(i)$, are placed into the SVD one by one. Since each SVD is processed for a single cycle, according to the principle of the matrix for SVD, a singular value σ_i, which represents its characteristic information, can be obtained in every discharging cycle. Finally, the singular value σ_i, which is extracted from all discharging cycles, can constitute HI1, denoted as $\{\sigma_i\}_{i=1}^{c}$.

3.3 Feature evaluation and selection by Spearman correlation analysis

Feature integration is a key step in constructing HIs, and quantitative evaluation and feature selection on the potential HIs extracted need to be conducted before integration to obtain those with stronger correlation to the capacity. Considering that the data used in this work do not conform to a normal distribution, Spearman's correlation coefficient is selected to measure the correlation of HIs. The correlation analysis between the fifteen potential HIs and capacity is shown in Table 1. Based on the absolute values of the correlation coefficients in Table 1, the HIs with very strong correlations (1 to 0.9) are selected for the subsequent feature integration. The nine His of HI2, HI3, HI4, HI5, HI6, HI10, HI11, HI13 and HI15 are selected for integration with B0005, B0006, B0007 and B0018.

3.4 HI integration based on the SDAE

SDAE is an unsupervised neural network that can be used for dimensionality reduction and feature extraction. SDAE is a stack of multiple denoising autoencoders and overcomes the disadvantage that a single layer of encoders cannot efficiently extract complex features. The highly correlated HIs selected in Sect. 2.3 are processed using SDAE to reduce dimensionality and noise. The specific structure of the SDAE is shown in Fig. 2. The two SDAEs are stacked into a stacked autoencoder for two-order parameter fusion, and the structure is shown in Fig. 3. In the constructed SDAE, the number of nodes in the output layer is set to 1 such that the fusion result is a series, and the output result is the second-order fusion HI.

In the Fig. 2, $\boldsymbol{X}$ and $\hat{\boldsymbol{X}}$ are the inputs and outputs of the network, respectively; $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$ are the connection weights; and $\mathbf{b}_1$, $\mathbf{b}_2$ and $\mathbf{b}_3$ are the bias values. The hidden layer node output $\mathbf{h}_1$, $\mathbf{h}_2$ and the output layer node $\hat{\mathbf{X}}$ are represented by the following equations [15]:

$$\mathbf{h}_1 = f_1(\mathbf{W}_1\mathbf{X} + \mathbf{b}_1) \tag{9}$$

$$\mathbf{h}_2 = f_2(\mathbf{W}_2\mathbf{h}_1 + \mathbf{b}_2) \tag{10}$$

$$\hat{\mathbf{X}} = f_3(\mathbf{W}_3\mathbf{h}_2 + \mathbf{b}_3) \tag{11}$$

where $f_i(\cdot), i = 1, 2, 3$ represents the node excitation functions and SDAE mainly maps the n-dimensional input sample $\mathbf{X}$ to 1-dimensional $\hat{\mathbf{X}}$ through encoding excitation functions.

To further analyze the accuracy of the integrated HI in quantifying the degradation state of the battery, quantitative and qualitative analyses are performed on it. The results of the quantitative analysis are given in Table 2, and to test the significance of the results, a test is conducted using the p value method. The table shows that the Spearman correlation coefficients of the integrated HI for all four batteries are greater than 0.99, and the p value is also much less than 0.01, indicating that a significant correlation exists between the integrated HI and capacity. The results of the qualitative analysis are shown in Fig. 4. The degradation trends from the four batteries' integrated HI curves and the original capacity curves are basically the same. Taken together, the results from the quantitative and qualitative analyses indicate that the integrated HI can effectively represent the degradation state of the batteries.

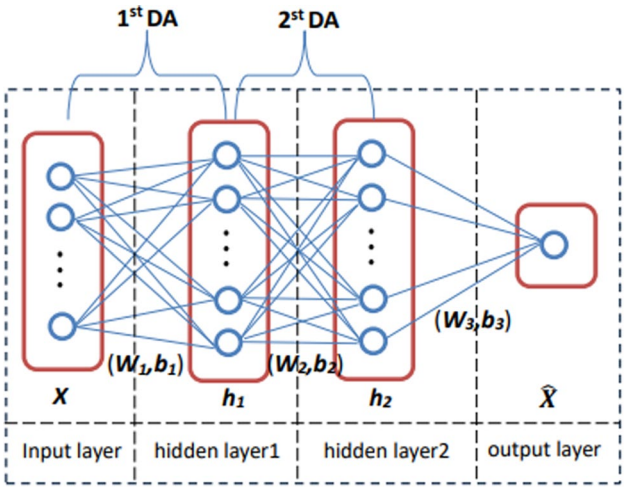

Fig. 2 Second-order feature fusion based on SDAE

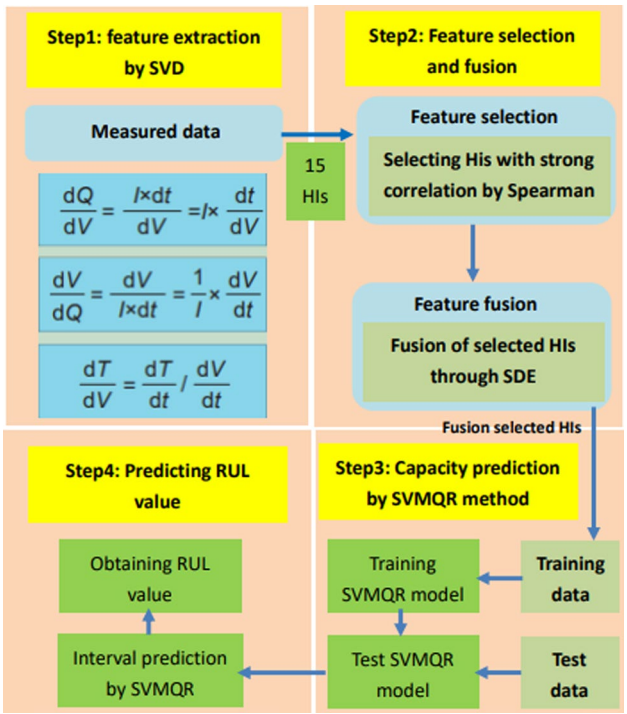

Fig. 3 Flowchart of the RUL prediction for Li-ion batteries based on SVD-SDAE-SVQR

Table 2 Spearman correlation analysis of the fused HI and capacity of four batteries

Spearman CC	B0005	B0006	B0007	B0018
HI	0.9974	0.9991	0.9989	0.9937
p value	1.92E−205	6.07E−218	7.94E−209	2.27E−113

4 RUL interval prediction based on support vector quantile regression

4.1 Quantile regression

Quantile regression (QR) [22] extends the traditional mean regression to quantile regression, can be used to describe the global characteristics of response variables in more detail and is not affected by outliers; moreover, the results are more robust. Importantly, QR does not require regression residuals to satisfy the basic assumptions of a normal distribution, random independence, and uniform mean regression methods. This approach is highly valuable for practical engineering problems such as RUL estimation of lithium batteries because the actual operation of lithium batteries is affected by variable working conditions, operating environments and other factors; thus, the monitoring data has difficulty meeting the normal distribution hypothesis. Given training set $T = (x_i, y_i), i = 1, 2, \cdots, n$, the QR model can be constructed according to the following process:

$$q_\tau(\mathbf{x}_i) = \mathbf{x'}_i\boldsymbol{\beta}(\tau), \quad i = 1, 2, \cdots, n; \tau = \tau_1, \tau_2, \cdots, \tau_M \tag{12}$$

where $\mathbf{x}_i = [x_{i1}, x_{i2}, \cdots, x_{id}]$ is an independent variable with d components, x_{ij} is the j th component of $\mathbf{x_i}$, $\mathbf{y_i}$ is the dependent variable corresponding to $\mathbf{x_i}$, and $q_\tau(\mathbf{x_i})$ is the τ-quantile of the dependent variable $\mathbf{y_i}$ under the independent variable $\mathbf{x_i}$. For any quantile $0 < \tau < 1$, $\boldsymbol{\beta}(\tau) = [\beta_1(\tau), \cdots, \beta_d(\tau)]'$ is the regression coefficient under the quantile τ. The estimate of $\boldsymbol{\beta}$ is obtained by minimizing the test function $\rho_\tau(\sigma)$, as follows:

$$\hat{\boldsymbol{\beta}}(\tau) = \arg\min \sum_{i=1}^{n} \rho(\tau)(y_i - x_i'\boldsymbol{\beta}(\tau)) \tag{13}$$

where n is the number of samples and the test function $\rho_\tau(\sigma)$ is a piecewise linear function. The definition of $\rho_\tau(\sigma)$ is as follows:

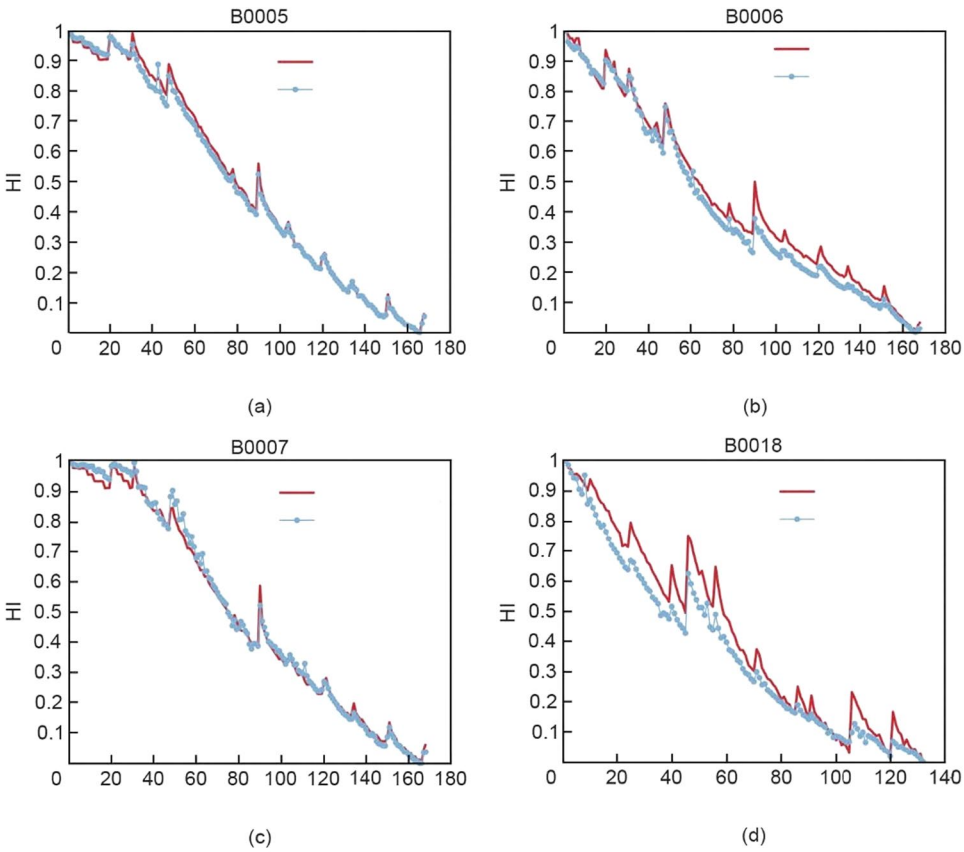

Fig. 4 Fusion HI qualitative analysis: **a** B0005; **b** B0006; **c** B0007; **d** B0018

$$\rho_\tau(\sigma) = \sigma(\tau - I(\sigma)) \tag{14}$$

where $I(\sigma)$ is an indicator function and $I(\sigma)$ is defined as follows:

$$I(\sigma) = \begin{cases} 0, \sigma \geq 0 \\ 1, \sigma < 0 \end{cases} \tag{15}$$

4.2 Support vector quantile regression

Although the QR method can describe the global characteristics of response variables in more detail without the need for the described object to follow a specific distribution and based on the definition of QR, QR is essentially a regression analysis for linear problems. In practical engineering problems, the relationship between the independent variable and dependent variable is often nonlinear, and determining the explicit function form is difficult.

SVR [23, 24] can implicitly map low-dimensional nonlinear data to linear regression problems in high-dimensional feature spaces through kernel function techniques. Based on the principle of structural risk minimization, SVR has unique advantages in solving small sample sets and nonlinear problems; thus, it is desirable to combine SVR and QR to fully exploit their advantages.

$$\min \ \frac{1}{2}|\mathbf{w}|^2 + \lambda \sum_{i=1}^{n} \boldsymbol{\rho}_\tau(\mathbf{y}_i - \mathbf{w}^T\phi(\mathbf{x}_i) - \mathbf{b}) \tag{16}$$

where λ is the penalty parameter, n is the number of samples, and $\phi(\mathbf{x})$ is a nonlinear mapping function that can be solved by implicitly mapping the d-dimensional input vector $\mathbf{x}$ to a high-dimensional feature space. Similar to the process of solving for SVR, by introducing relaxation variables ξ_i, ξ_i^*, Eq. (19) can be converted to solve the following optimization problem:

$$\min_{\mathbf{w},\mathbf{b},\xi,\xi^*} \ \frac{1}{2}|\mathbf{w}|^2 + \lambda \sum_{i=1}^{n} (\tau\xi_i + (1-\tau)\xi_i^*) \tag{17}$$

$$s.t \begin{cases} y_i - \mathbf{w}^T\phi(\mathbf{x}_i) - b \leq \xi_i \\ \mathbf{w}^T\phi(\mathbf{x}_i) + b - y_i \leq \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{18}$$

The semiparametric method [25] can be adopted to solve the optimization problem of Eq. (18) by constructing the Lagrange function, as shown in Eq. (19):

$$\begin{cases} \mathbf{w}_\tau = \sum_{\tau=1}^{T} (a_t - a_t^*)\phi(\mathbf{x}_t) \\ (b_\tau, \boldsymbol{\beta}_\tau)^T = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T(\mathbf{y} - \mathbf{K}(\mathbf{a} - \mathbf{a}^*)) \\ Q_{yt}(\tau|\mathbf{u}_t, \mathbf{x}_t; \boldsymbol{\theta}) = b_\tau + \boldsymbol{\beta}_\tau^T\mathbf{u}_t + \mathbf{K}_t(\mathbf{a} - \mathbf{a}^*) \end{cases} \tag{19}$$

where a_t, a_t^* is the optimal Lagrange multiplier obtained by optimization, $\mathbf{U} = (1, u_t^T)$ is a design matrix, $\mathbf{y}$ is a vector composed of support vectors y_t, $t \in I_{SV}$, I_{SV} is the set of subscripts of support vectors, $I_{SV} = \{t = 1, 2, \cdots, T | 0 < a_t < \tau\lambda, 0 < a_t^* < (1-\tau)\lambda\}$, $K(\mathbf{x}_s, \mathbf{x}_t) = \phi(\mathbf{x}_s)^T\phi(\mathbf{x}_t)$ is the kernel function, $\mathbf{K}$ is the kernel matrix, and K_t is the t-th row of $\mathbf{K}$. The kernel function and penalty parameter λ have a great impact on the performance of SVQR. Many studies have shown that the Gaussian kernel function effectively performs in most cases [26]; therefore, the Gaussian kernel function is adopted, and its form is as follows:

$$K(\mathbf{x}_s, \mathbf{x}_t) = \exp\left(-\frac{|\mathbf{x}_s - \mathbf{x}_t|^2}{2\sigma_K^2}\right) \tag{20}$$

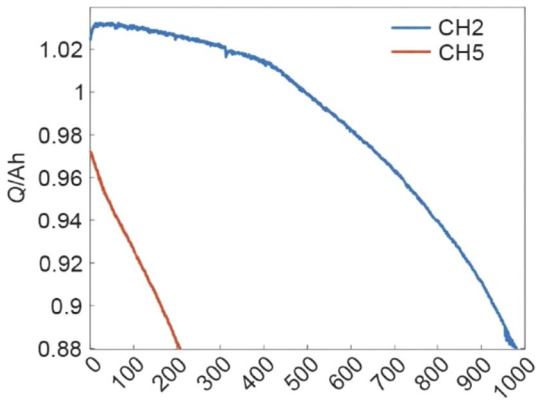

Fig. 5 Capacity degradation curves of CH2 and CH5

$\theta = (\lambda, \sigma_K^2)^T$ is the hyperparameter that needs to be determined before training. In this study, the generalized approximate cross-validation (GACV) criterion [24] of Eq. (21) is adopted to obtain the optimal hyperparameter combination:

$$\mathrm{GACV}(\boldsymbol{\theta}) = \frac{\sum_{t=1}^{n} \rho_\tau(y_t - Q_{y_t}(\tau|\mathbf{u}_t, \mathbf{x}_t;\theta))}{n - trace(\mathbf{H})} \tag{21}$$

where θ is the hyperparameter set and $\mathbf{H}$ can be obtained by

$$\left(Q_{y_1}(\tau|u_1, x_1), \cdots, Q_{y_n}(\tau|u_n, x_n)\right)^T = \mathbf{H}\mathbf{y} \tag{22}$$

4.3 Modeling process based on the SVD-SDAE-SVQR

In this section, a new SVD-SDAE-SVQR lithium-ion battery RUL prediction method is proposed by combining SVD, SDAE and SVQR. The specific process, as shown in Fig. 5, can be divided into four steps.

Step 1: SVD feature extraction. First, SVD is performed on the measurement data of the battery, and their singular values are taken as 12 potential HIs. Second, the dQ/dV, dV/dQ and dT/dV matrices are calculated based on parameters such as electricity, time, voltage and temperature, from which HI13, HI14 and HI15 are extracted using SVD.

Step 2: Feature selection and feature fusion. Spearman correlation analysis is used to select highly correlated HIs from the 15 potential HIs in Step 1 for feature fusion. The selected HIs are fused into an HI by SDAE processing.

Step 3: SVQR capacity prediction. With the fusion feature $\{\mathrm{HI}(i)\}_{i=1}^{c}$ as the input and the battery capacity $\{\mathrm{C}(i)\}_{i=1}^{c}$ as the output, a GPR-based capacity prediction model is constructed, where c is the total cycle period. The training set is denoted as $\{\boldsymbol{x}, \boldsymbol{y}\}$, where $\boldsymbol{x} = [\mathrm{HI}(1:k)^{\mathrm{T}}]$ and $\boldsymbol{y} = [\mathrm{C}(1:k)|$ and the number of cycles corresponding to the prediction starting point is denoted as $k+1$. The test set is expressed as $\{\boldsymbol{x}^*, \boldsymbol{y}^*\}$, where $\boldsymbol{x}^* = [\mathrm{HI}(k+1:c)^{\mathrm{T}}]$ and $\boldsymbol{y}^* = [\mathrm{C}(k+1:c)^{\mathrm{T}}]$. The prediction part uses $\boldsymbol{x}^*$ as input, and the predicted capacity value $\boldsymbol{y}^*$ is obtained.

Step 4: Calculation of the RUL prediction value. The remaining life (N_{RUL}) is the difference between the total number of charge and discharge cycles (N_{EOL}) when the actual battery capacity drops to the threshold and the current number of charge and discharge cycles (N_{ECL}). The calculation formula is shown in Eq. (23):

$$N_{RUL} = N_{EOL} - N_{ECL} \tag{23}$$

4.4 Evaluation indices of prediction performance

Three evaluation indices, the root mean square error (RMSE), coefficient of determination (R^2) and RUL absolute error (AE), were used to evaluate the prediction performance of our proposed method. The calculation formulas for the RMSE, R^2 and AE are calculated as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2},\ R^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2},\ AE = |RUL^T - RUL^P| \tag{24}$$

where N is the number of cycles of the battery; y_i and $\hat{y}_i$ represent the actual capacity and predicted capacity of the lithium-ion battery, respectively, during the i th cycle; and $\overline{y}$ is the mean of the actual capacity. RUL^T and RUL^P represent the actual RUL and predicted RuL, respectively, when the battery capacity reaches the failure threshold.

5 Experimental results and analysis

5.1 Li-ion battery RUL prediction results

To analyze the RUL prediction performance of the model, experiments with different prediction starting points are designed; the starting points of prediction are 60 and 80. Assuming that the capacity data length of the lithium-ion battery is N, in the experiment with a starting point of 60, the first 59 capacity data points are selected as the training set, and the remaining capacity data are selected as the testing set. Specifically, the training set is $[Q(1), \cdots, Q(59)]$, and the testing set is $[Q(60), \cdots, Q(N)]$. Similarly, in the experiment with a starting point of 80, the first 79 capacity data points ($[Q(1), \cdots, Q(79)]$) are selected as the training set, and the remaining capacity data point ($[Q(80), \cdots, Q(N)]$) are selected as the testing set. The Gaussian function is used as the kernel function of support vector regression, and the optimal hyperparameter combination for support vector regression is determined by tenfold cross-validation.

In the experiments, the capacity fault threshold of batteries B0005, B0006, and B0018 is 1.38 Ah. Because the B0007 battery does not reach the fault threshold, no fault threshold line is provided.

Figure 6 shows the prediction results and their uncertainty expression for the four batteries at prediction starting points 60 and 80. Overall, the predicted curve is close to the real capacity degradation curve, the prediction result is less affected by the prediction starting point, and the prediction curves of the four batteries at prediction starting point 80 are closer to the real capacity degradation curve than those at prediction starting point 60.

In addition, Table 3 shows the evaluation results from the prediction performance of the four batteries at different prediction starting points. As shown in the table, the prediction error at prediction starting point 80 is lower than that at prediction starting point 60, and the RMSE of the four batteries is less than 0.0301; thus, these results indicate that the prediction performance of our proposed model is good. The R^2 values of the four batteries are all greater than 0.93, indicating a high degree of fitting between the predicted curve and the actual curve. The AE values of batteries B0005, B0006 and B0018 gradually decrease as the prediction starting point is reduced. When the prediction starting point is 60, the AE values of the three batteries are all less than 6, and the error values are low. In addition, as the prediction starting point is set to 80, the AE value gradually decreases and is less than 2.

5.2 Adaptability of SVD to different battery datasets

To verify the adaptability of the SVD feature extraction method to different lithium-ion battery datasets, the battery datasets provided by MIT [27] are selected for the experiments in this section. The battery parameters and experimental environment are different from those of the battery datasets provided by NASA. The battery datasets used for the test are first described in detail. Then, the fusion HI are still extracted by combining SVD and the SDAE. Finally, the SVQR is used for capacity prediction.

5.2.1 MIT datasets

The test battery is a lithium iron phosphate (LFP)/graphite battery manufactured by an A123 system (APR18650 M1A), and it is placed in a forced convection temperature chamber at 30 ℃. During the charging stage, 72 charging

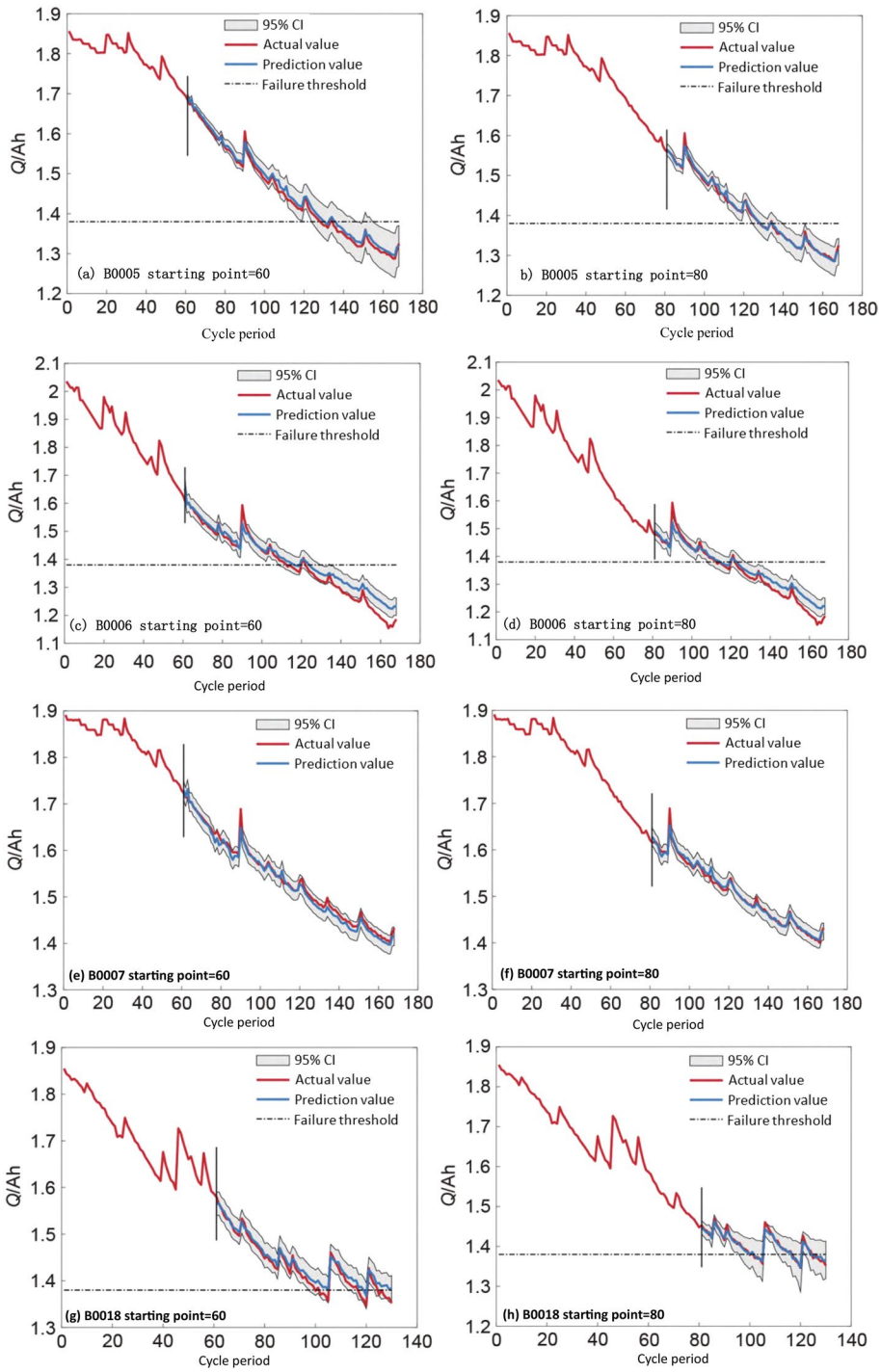

Fig. 6 Prediction results for the four batteries at different prediction starting points

Table 3 Prediction performance of four batteries at different prediction starting points

Evaluation index	Prediction starting point	B0005	B0006	B0007	B0018
RMSE	60	0.0116	0.0329	0.0199	0.0143
	80	0.0066	0.0280	0.0077	0.0080
R^2	60	0.9921	0.9399	0.9503	0.9533
	80	0.9982	0.9407	0.9894	0.9579
AE	60	2	4	–	6
	80	0	2	–	2

Table 4 Spearman correlation analysis between capacity and potential HIs of CH2 and CH5

Feature extraction objects	HIs	Spearman correlation coefficient	
		CH2	CH5
time	HI1	0.9378	0.9871
Charging capacity	HI2	0.9972	0.9986
electric current	HI3	0.9985	0.9979
voltage	HI4	0.9401	0.9841
temperature	HI5	0.9988	0.9893
dQ/dV	HI6	0.9971	0.9977
dV/dQ	HI7	− 0.9964	− 0.9921
dT/dV	HI8	− 0.2462	0.1982

Table 5 Spearman correlation analysis of the capacity and the fused HI

Spearman CC	CH2	CH5
HI	0.9978	0.9985
p value	$0.85*10^{-4}$	$0.42*10^{-4}$

modes are formed by a constant current fast charging combination with different rates, and the discharge process adopts a unified 4 C rate. The battery has a nominal capacity of 1.1 Ah and a nominal voltage of 3.3 V. Two sets of data, Channel 2 (CH2) and Channel 5 (CH5), in the second batch of data labelled as "2017-06-30," are selected for the experiment. The original capacity degradation trajectories of CH2 and CH5 are shown in Fig. 5.

5.2.2 Feature extraction and fusion

For CH2 and CH5, the same SVD is used to extract their respective singular values from the measurement data, dQ/dV, dV/dQ, and dT/dV, respectively, as HIs. dQ/dV is given in the dataset, and dV/dQ and dT/dV need to be calculated. When calculating dT/dV, because the same voltage value collected at some adjacent time points in the same cycle causes the dT/dV value to be invalid, the repeated voltage data and the temperature data corresponding to the corresponding measurement time are removed. Spearman correlation analysis is performed on the extracted potential HIs, and the analysis results are shown in Table 4. Except for HI8, the correlation coefficients between the remaining 7 potential HIs and capacity are all greater than 0.9 in absolute value, showing strong correlations. Therefore, according to the correlation coefficient, 7 HIs, excluding HI8, are selected and fused into one HI using the SDAE. To analyze the ability of the fusion HI to characterize the battery degradation state, the Spearman correlation coefficient and p value are calculated, and the specific results are provided in Table 5. As shown in Table 5, the Spearman correlation coefficients between the fused HI and tumor volume are greater than 0.99, and the p value is close to 0, indicating that the fused HI is significantly correlated with volume.

5.2.3 Capacity prediction

SVQR is used to model the prediction portion. By fusing the HI data as training data and capacity data as output data, a capacity prediction model based on the SVQR is constructed. Based on the data size of CH2 and CH5, the data from the first 500 and first 100 cycles for training were selected, respectively. The prediction results are shown in Fig. 7; the prediction curve approaches the real capacity degradation curve relatively accurately. The evaluation results of the prediction performance are provided in Table 6. As shown in the table, the RMSEs of CH2 and CH5 are both less than 0.008, and R^2 is greater than 0.94; thus, our proposed model has good prediction performance and the predicted curve is in good agreement with the actual curve.

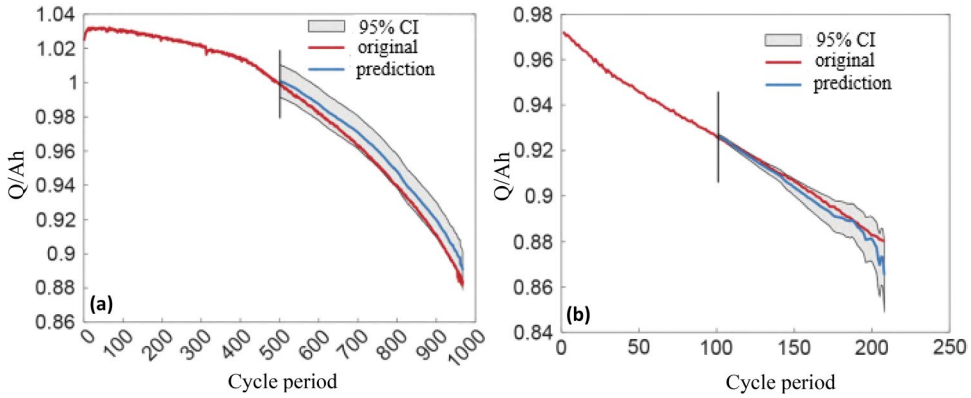

Fig. 7 Capacity prediction results for CH2 and CH5 **a** CH2 **b** CH5

Table 6 Prediction performances of CH2 and CH5

Evaluation index	CH2	CH5
RMSE	0.0068	0.0035
R^2	0.9614	0.9568

Table 7 Prediction performance of the three methods

Evaluation index	Method	B0005	B0006	B0007	B0018
RMSE	Proposed method	0.0055	0.0284	0.0082	0.0068
	SVD-PCA-SVQR	0.0177	0.0333	0.0125	0.0195
	SHI-SVQR	0.0272	0.0402	0.0139	0.0096
R^2	Proposed method	0.9961	0.9409	0.9903	0.9586
	SVD-PCA-SVQR	0.9570	0.8961	0.9747	0.9334
	SHI-SVQR	0.8996	0.8473	0.9590	0.9466
AE	Proposed method	1	2	–	1
	SVD-PCA-SVQR	3	4	–	3
	SHI-SVQR	9	6	–	7

In summary, through battery dataset experiments provided by NASA and MIT, SVD can be used to extract HIs and has good adaptability to different battery datasets while avoiding the artificial parameter setting. In addition, the fused HI has a strong correlation and better prediction performance and accuracy.

5.3 Comparative test analysis of different HIs

To more fully illustrate the effectiveness and superiority of using the SVD method to extract potential HIs from measurement data, dQ/dV, dV/dQ and dT/dV are used; highly correlated HIs are selected for feature fusion by using the SDAE and combined with the SVQR prediction model (SVD-SDAE-SVQR); and two comparison models, SVD-PCA-SVQR and SHI-SVQR, are designed. In the SVD-PCA-SVQR method, HIs with strong correlations are selected from the 15 potential HIs by Spearman correlation analysis, the selected HIs are fused into an HI indicator using PCA, and the fused HI indicator and SVQR model are used for RUL prediction. In the SHI-SVQR method, the discharge time of an equal voltage drop from 3.8 V to 3.5 V during the discharge phase is directly used as the HI, and the single HI and SVQR models (SHI-SVQR) are used for RUL prediction.

The prediction performances of the three models for the B0005, B0006, B0007, and B0018 batteries are shown in Table 7. For the SVD-SDAE-SVQR model, the RMSEs of batteries B0005, B0006, B0007 and B0018 are 0.0055, 0.0284, 0.0082 and 0.0068, respectively. The RMSEs of the four batteries are lower than those of the SVD-PCA-SVQR model and SHI-SVQR. The RMSE of the SVD-PCA-SVQR model is also lower than that of the SHI-SVQR model. For the SVD-SDAE-SVQR model, the AE values of the three batteries except B0007 are 1, 2 and 1, which are also lower than the AE values of SVD-PCA-SVQR and SHI-SVQR. Moreover, SVD-PCA-SVQR has a better prediction effect than the SHI-SVQR model. From the

RMSE and AE results, the SVD-SDAE-SVQR model has better prediction accuracy than those of the SVD-PCA-SVQR and SHI-SVQR models, and the prediction performance of the HIs extracted in this study after PCA fusion is also better than that of a single HI.

The R^2 values of the four batteries of the SVD-SDAE-SVQR model are 0.9961, 0.9409, 0.9903 and 0.9586, which are significantly greater than the R^2 values of the SVD-PCA-SVQR and HI-SVQR models. These findings fully indicate that the prediction results of the SVD-SDAE-SVQR method can better fit the original degradation curve than those of the SVD-PCA-SVQR and SHI-SVQR models. In addition, except for B0018, the R^2 values of the other three batteries of the SVD-PCA-SVQR model are also significantly greater than those of the SHI-SVQR model, indicating that the prediction results of HI after PCA fusion better fit the original curve than those of the SHI-SVQR model. The comparative experimental results show that the same HIs after SDAE fusion have better HI prediction performance than those after PCA fusion; thus, the SDAE is better than PCA in feature fusion. In addition, the prediction performance of the potential HIs extracted in this study after fusion is better than that of other HIs, indicating the effectiveness and superiority of HIs extracted by our proposed model.

5.4 Comparison with existing methods

To further validate the performance of our method, we compared it to four existing prediction methods on NASA datasets. The GPR model [28], GCO-LSTM model [29], ARRVM model [30], and ADE-MESN model [31] were selected and compared with our proposed prediction model. The 80th cycle was taken as the starting point of prediction, and *RMSE*, MAE and AE were selected as evaluation indices. The experimental results on the MIT dataset (CH2, CH5, CH9 and CH10) are shown in Table 8.

As shown in Table 8, compared with the GPR, GGO-LSTM, AR-RVM and ADE-MESN models, our proposed model has a lower RMSE and higher R^2 values. Compared with those of the other four methods, the RMSEs of our proposed method are reduced by 56.72%, 61.17%, 17.57% and 30.94%, and the R^2 are increased by 1.5012%, 1.9893%, 0.7806% and 0.8617%, respectively. Compared with the other four methods, our proposed prediction method has higher prediction accuracy.

6 Conclusion

In this paper, a RUL prediction method for lithium-ion batteries based on SVD-SDAE-SVQR is proposed. The main contributions are as follows:

First of all, in order to solve the shortcomings of traditional feature extraction methods, which depend on parameter setting and have poor adaptability to different lithium-ion battery data sets, SVD is used for feature extraction to solve the above problems.

Then, 15 potential HIs are extracted from the raw measurement data and calculated feature extraction objects containing more degradation information. Spearman correlation analysis and SDAE are used to deal with potential HIs to obtain a fused HI, considering that redundancy and insufficiency of HIs can affect RUL prediction.

Finally, the fusion HI and SVQR were combined to predict RUL, and the uncertainty expression of the result was given. In the validation phase, experimental results on battery datasets provided by NASA and MIT demonstrate the effectiveness and adaptability of this feature extraction strategy. The superiority of the proposed HI was verified by comparing

Table 8 Prediction results from the different models

No	Method	RMSE	R^2	AE	No	Method	RMSE	R^2	AE
CH2	GPR	0.0256	0.9855	2	CH9	GPR	0.0205	0.9715	1
	GWO-LSTM	0.0289	0.9817	2		GWO-LSTM	0.0262	0.9681	2
	AR-RVM	0.0083	0.9914	1		AR-RVM	0.0166	0.9757	1
	SADE-MESN	0.0091	0.9885	0		SADE-MESN	0.0113	0.9809	1
	SVD-SDAE-SVQR	0.0066	0.9982	0		SVD-SDAE-SVQR	0.0075	0.9894	0
CH5	GPR	0.0474	0.9221	3	CH10	GPR	0.0225	0.9497	4
	GWO-LSTM	0.0493	0.9205	3		GWO-LSTM	0.0249	0.9401	4
	AR-RVM	0.0322	0.9358	2		AR-RVM	0.0106	0.9532	3
	SADE-MESN	0.0396	0.9323	2		SADE-MESN	0.0127	0.9513	3
	SVD-SDAE-SVQR	0.0280	0.9407	2		SVD-SDAE-SVQR	0.0081	0.9579	2

the HI after PCA fusion with other HI. In summary, the SVD feature extraction method has good adaptability, and the fusion HI obtained by Spearman correlation analysis and SDAE processing has good prediction performance and the proposed RUL prediction framework has the advantages of high accuracy and output probability.

The proposed method also has some limitations. For example, laboratory data of lithium-ion batteries may vary according to different experimental conditions and measuring instruments. Will these differences affect the feature extraction effect of SVD-SDAE? Can the proposed method be directly applied to different types battery data without changing the model parameters? Therefore, the proposed method needs to be further explored in combination with transfer learning, so that it can be more widely applied to the battery measurement data under relatively complex working conditions.

Author contributions L. Sun: Conceptualization, Methodology, Supervision, Writing—Review and Editing. X. Hang, J. Liu, J. Song: Software, Formal analysis, Validation, Visualization, Writing—Original Draft

Data availability The data that support the findings of this study are openly available at https:// ti.arc.nasa.gov/tech/dash/grou-ps/pcoe/prognostic-data-repository/.

Declarations

Competing interests The authors declare no competing interests.

References

1. Pargolettie AS, Cappelletti G, et al. Smart interfaces in Li-ion batteries: Near-future key challenges. Electrochimica Acta. 2022;415:140258.
2. Liang XC, Zhang M, Huang GJ. Review on lithium-ion battery modeling methods based on BMS. Energy Storage Sci Technol. 2020;9:1933–9.
3. Hu XS, Zhang K, Liu KL, et al. Advanced fault diagnosis for lithium-ion battery systems: a review of fault mechanisms, fault features, and diagnosis procedures. IEEE Ind Electron Mag. 2020;14:65–91.
4. Chen Y, Bai YF, He Y. Comparison of data-driven lithium battery state of health estimation methods. Energy Storage Sci Technol. 2019;8:1204–10.
5. Pang XQ, Zhao Z, Wen J, et al. An interval prediction approach based on fuzzy information granulation and linguistic description for remaining useful life of lithium-ion batteries. J Power Sources. 2022;542: 231750.
6. Chu A, Allam A, Cordoba A, et al. Stochastic capacity loss and remaining useful life models for lithium-ion batteries in plug-in hybrid electric vehicles. J Power Sources. 2020;478: 228991.
7. Zhou WL, Lu Q, Zheng YP. Review on the selection of health indicator for lithium—ion batteries. Machines. 2022;10:512.
8. Khaleghi HMS, Karimi D, et al. Developing an online data-driven approach for prognostics and health management of lithium-ion batteries. Appl Energy. 2022;308:118348.
9. Jia JF, Yuan SF, Shi YH, et al. Improved sparrow search algorithm optimization deep extreme learning machine for lithiumion battery state-of-health prediction. iScience. 2022;25:103988.
10. Li XD, Zhang XW. State of health estimation method for lithium-ion batteries based on principal component analysis and whale optimization algorithm-Elman model. Energy Storage Sci Technol. 2022;11:4010–21.
11. Liu J, Chen ZQ. Remaining useful life prediction of lithium-ion batteries based on health indicator and Gaussian process regression model. IEEE Access. 2019;7:39474–84.
12. Pang XQ, Liu XY, Jia JF, et al. A lithium-ion battery remaining useful life prediction method based on the incremental capacity analysis and Gaussian process regression. Microelectron Reliab. 2021;127: 114405.
13. Pan WJ, Luo XS, Zhu MT, et al. A health indicator extraction and optimization for capacity estimation of Li-ion battery using incremental capacity curves. Journal of Energy Storage. 2021;42: 103072.
14. Wall ME, Rechtsteiner A, Rocha LM. Singular value decomposition and principal component analysis A practical approach to microarray data analysis. Boston: Kluwer Academic Publishers; 2005.
15. Yang GY, Cheng Y, Xi CB, Liu L, Gan X. Combine harvester bearing fault-diagnosis method based on SDAE-RCmvMSE. Entropy. 2022;24:1139.
16. Ye YF, Shao YH, Li CN, Hua XY, Guo YR. Online support vector quantile regression for the dynamic time series with heavy-tailed noise. Appl Soft Comput. 2021;119: 107560.
17. Saha B, Goebel K. Battery data set[R/OL]. [2020–10–20]. https:// ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/.
18. Sevrson KA, Attia PM, Jin N, et al. Data-driven prediction of battery cycle life before capacity degradation. Nat Energy. 2019;4:383–91.

19. Ribeiro M, Lazzaretti AE, Lopes HS. A study of deep convolutional auto encoders for anomaly detection in videos. Pattern Recogn Lett. 2018;105:13–22.
20. Tang T, Yuan HM. The capacity prediction of Li-ion batteries based on a new feature extraction technique and an improved extreme learning machine algorithm. J Power Sources. 2021;514: 230572.
21. Fly A, Chen R. Rate dependency of incremental capacity analysis (dQ/dV) as a diagnostic tool for lithium-ion batteries. J Energy Storage. 2020;29: 101329.
22. Shibagaki T, Merla Y, Offer GJ. Tracking degradation in lithium iron phosphate batteries using differential thermal voltammetry. J Power Sources. 2018;374:188–95.
23. Yoshida T. Extreme value inference for quantile regression with varying coefficients. Commun Statist Theory Method. 2021;50:685–710.
24. Shim J, Hwang C, Seok K. Support vector quantile regression with varying coefficients. Comput Statistics. 2016;31:1015–30.
25. Shim J, Hwang C, Jeong S, et al. Semivarying coefficient least-squares support vector regression for analyzing high-dimensional gene-environmental data. J Appl Stat. 2018;45:1370–81.
26. Anand P, Rastogi RNK, Chandra S. A new asymmetric epsilon-insensitive pinball loss function based support vector quantile regression model. Appl Soft Comput. 2020;94: 106473.
27. Xi C, Weidong L, Xiaojun M, et al. Distributed high-dimensional regression under a quantile loss function. Mach Learn Res. 2020;21:182.
28. Clerici D, Mochera F, Soma A. Electrochemical-mechanical multi-scale model and validation with thickness change measurements in prismatic lithium-ion batteries. J Power Sourc. 2022;542: 231735.
29. Wang JW, Deng ZW, Yu T, et al. State of health estimation based on modified Gaussian process regression for lithium-ion batteries. J Energy Storage. 2022;51: 104512.
30. Feng HL, Yan HM. State of health estimation of large-cycle lithium-ion batteries based on error compensation of autoregressive model. J Energy Storage. 2022;52: 104869.
31. Ma Y, Shan C, Gao JW, et al. A novel method for state of health estimation of lithium-ion batteries based on improved LSTM and health indicators extraction. Energy. 2022;251: 123973.
32. Ji YF, Chen ZW, Shen Y, et al. An RUL prediction approach for lithium-ion battery based on SADE-MESN. Appl Soft Comput. 2021;104: 107195.