J. Meas. Sci. Instrum., 2024, 15(3): 397-407. http://jmsi.publish.founderss.cn/
DOI: 10.62756/jmsi.1674-8042.2024041 jmsi@nuc.edu.cn

State of health prediction for lithium-ion batteries based on ensemble Gaussian process regression

HUI Zhouli, WANG Ruijie, FENG Nana, YANG Ming*

School of Mathematics, North University of China, Taiyuan 030051, China

***Corresponding author:** YANG Ming (hgsnje@nuc.edu.cn)

Received: May 16, 2024 **Revised:** September 2, 2024 **Accepted:** September 10, 2024

Abstract: The performance of lithium-ion batteries (LIBs) gradually declines over time, making it critical to predict the battery's state of health (SOH) in real-time. This paper presents a model that incorporates health indicators and ensemble Gaussian process regression (EGPR) to predict the SOH of LIBs. Firstly, the degradation process of an LIB is analyzed through indirect health indicators (HIs) derived from voltage and temperature during discharge. Next, the parameters in the EGPR model are optimized using the gannet optimization algorithm (GOA), and the EGPR is employed to estimate the SOH of LIBs. Finally, the proposed model is tested under various experimental scenarios and compared with other machine learning models. The effectiveness of EGPR model is demonstrated using the National Aeronautics and Space Administration (NASA) LIB. The root mean square error (RMSE) is maintained within 0.20%, and the mean absolute error (MAE) is below 0.16%, illustrating the proposed approach's excellent predictive accuracy and wide applicability.

Key words: lithium-ion batteryies (LIBs); ensemble Gaussian process regression (EGPR); state of health (SOH); health indicators (HIs); gannet optimization algorithm (GOA)

0 Introduction

Lithium-ion batteries (LIBs) are widely used in electric vehicles due to their high energy density, long cycle life, and low environmental impact[1]. When LIBs are used improperly or over time, their performance deteriorates, which reduces their capacity and power[2,3]. Battery management system (BMS) provides scientific assessment, risk warning, and periodic replacement of batteries to maintain stable operation[4,5]. State of health (SOH) is a crucial reference variable in a BMS[6], which refers to the ratio of battery's current capacity to its initial capacity during discharge [7].

Predicting SOH is challenging because the degradation of LIBs is a long-term, complex process influenced by a wide range of variables. Numerous researchers have proposed approaches for predicting the SOH of LIBs to address this issue. Over the past decades, SOH prediction methods can be classified into three fundamental categories. The direct measurement method is to indirectly estimate the SOH by measuring battery capacity[8], internal resistance[9], and electrochemical impedance spectroscopy (EIS) [10,11]. Love et al.[12] proposed single-point impedance diagnosis to monitor the SOH of Li-ion individual cells and 4S battery packs. Although this approach is straightforward, but it is difficult to implement and can only be used in a certain laboratory setting. The model-based methods involve estimating the battery's SOH using filter methods such as extended Kalman filter (EKF) [13] or particle filter (PF) [14]. For example, Yan et al. [15] suggested an EKF based on Leberger sampling and Qiu et al.[16] suggested an enhanced cuckoo and PF algorithm for the SOH estimation of LIBs.

Nowadays, data-driven methods have gradually become the mainstream trend of SOH prediction, which just needs to estimate SOH by studying exterior metrics, not the intrinsic chemistry of the battery. The critical points of data-driven approach lie in the selectrion of health indicators (HIs) [17-22] and the construction of machine learning model[23-33]. Sun et al. [20] used incremental capacity analysis (ICA) to investigate SOH-related HIs thoroughly. Xia et al.[21] reconstructed the voltage using a second-order RC model to produce the incremental capacity (IC) and differential voltage (DV) curves to identify aging signs of LIBs by filtering and reshaping degradation curves. Tian et al.[22] directly extracted HIs from the voltage, current, and temperature curves of the discharge using an easy, practical and promising method. Neural network (NN) methods[23,24] and statistical methods[25-28] are the two primary categories of machine learning methods. Wu et al.[29] employed long-short-term memory (LSTM) and Xia et al.[30] proposed a deep neural network (DNN) to estimate the SOH of LIBs. These NN methods usually need a significant amount of data for training and are computationally difficult. Small sample applications are better suited for statistical learning, which also has the benefits of quick calculation and simple training. To predict the SOH of LIBs, Wu et al. [31] presented a combined bat algorithm (BA) and support vector regression (SVR) algorithm. Cui et al.[32] devised a method for SOH estimates based on gated recurrent units (GRU) and Mawonou et al.[33] used the random forest (RF) algorithm to forecast battery aging.

Gaussian process regression (GPR) has been used for SOH prediction over the past few years for its more effective management of high dimensional small-sample regression[34,35]. Gong et al.[36] developed a GPR model to estimate SOH based on three phases of energy characteristics. Feng et al.[37] adopted linear function as the mean function and combined kernel function as the covariance function to establish GPR model. Zheng et al.[38] designed a weighting strategy based on forecast uncertainty to integrate predictions from multiple GPR models and Liu et al.[39] modified the kernel function in GPR to an isotropic squared index kernel with an automatic correlation determination structure to predict SOH.

However, GPR models still have considerable untapped potential, and the learning capacity of individual GPR is limited. This study adopts a strategy of combining GPRs with various kernels for SOH prediction of LIBs, which is motivated by advantages of multi-model fusion. One of the main contributions of this research is the extraction of suitable HIs from the voltage temperature profile. Additionally, the GPR model incorporates the idea of ensemble learning, with the kernel function parameters optimized using a specific optimization method. The ensemble Gaussian process regression (EGPR) algorithm suggested in this research demonstrates superior performance compared to other machine learning algorithms, as evidenced by comparisons conducted under identical experimental conditions.

1 Battery aging data from NASA database

Appropriate battery capacity datasets are crucial for evaluating prognostic methods. For modeling and validation purposes, four groups of batteries, designated as B0005, B0006, B0007, and B0018, were selected from the NASA lithium battery datasets[40]. All the batteries examined are second-generation 18 650-sized lithium batteries, made from the same composite material, $LiCoO_2$. The experiments were conducted at room temperature (24 ℃) across three modes: constant current (CC) charging, constant voltage (CV) charging, and CC discharge. More details regarding the NASA battery datasets and their testing conditions can be found in Refs.[40,41]. The validity of utilizing this battery data for designing aging prognostic methods has been confirmed in numerous related studies[33,39,40]. Table 1 summarizes the key parameters and operational settings of the four datasets, while Fig. 1 illustrates their degradation curves in terms of SOH.

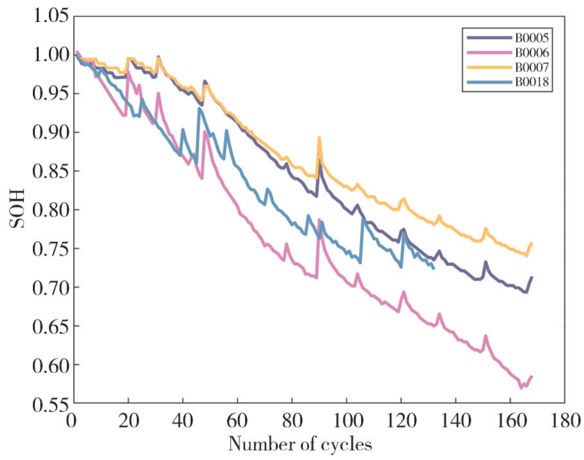


Fig. 1 SOH degradation curves of various selected LIBs

Table 1 Parameters of selected NASA LIBs

Battery No.	Cut-off voltage/V	Discharge cut-off voltage/V	Charging current/A	Discharging current/A	Temperature/℃
B0005	4.2	2.7	1.5	2	24
B0006	4.2	2.5	1.5	2	24
B0006	4.2	2.2	1.5	2	24
B0018	4.2	2.5	1.5	2	24

2 Methodology

2.1 EGPR

GPR is a machine learning model based on kernel functions and Bayesian theory. Without cross-validation, this model can still produce higher regularization outcomes[42]. Kernel functions are crucial in the GPR modeling process because they help formulate a prior hypotheses about the characteristics of prospective functions. The following four kernel functions[42] are utilized in this study, and they are

$$k_{\mathrm{M32}}=\sigma_1^2\left(1+\frac{\sqrt{3}\,r}{l_1}\right)\exp\left(-\frac{\sqrt{3}\,r}{l_1}\right), \tag{1}$$

$$k_{\mathrm{M52}}=\sigma_2^2\left(1+\frac{\sqrt{5}\,r}{l_2}+\frac{5r^2}{3l_2^2}\right)\exp\left(-\frac{\sqrt{5}\,r}{l_2}\right), \tag{2}$$

$$k_{\mathrm{SE}}=\sigma_3^2\exp\left(-\frac{r}{2l_3^2}\right), \tag{3}$$

$$k_{\mathrm{RQ}}=\sigma_4^2\left(1+\frac{r^2}{2\alpha l_4^2}\right)^{-\alpha}, \tag{4}$$

where $r=\sqrt{(x_i-x_j)^{\mathrm{T}}(x_i-x_j)}$ represents the distance between x_i and x_j; $\sigma_m^2(m=1,2,3,4)$ represents the signal variance; l_m represents the length of the scale parameter; and α is the proportional mixing parameter.

To make better predictions than a single learner, ensemble learning can mix many learning approaches[43]. One of them, stacking is an excellent ensemble learning technique. By enhancing the capability of the base learner and employing a meta-learner to integrate the outcomes predicted by the base learner, stacking aims to improve generalization[44,45].

The two-layer model in this research is constructed using the stacking approach. Its structure is shown in Fig.2. The base learner and mata learner are the first layer and second layer in the model, respectively. The first layer employs the combination of the Matern32, Matern52, rational quadratic (RQ) kernel, and Gaussian kernel function (SE) in the GPR model. The ultimate prediction result of the entire model is output when SVR receives the output of the base learner as input. The specific details and steps of GPR and SVR model can be seen in Refs.[31,37,39,42].

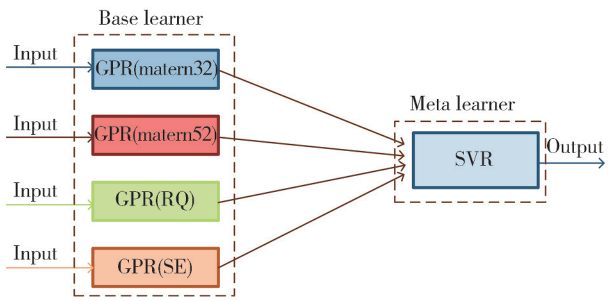


Fig. 2 EGPR model. Base learner is composed of GPR with different kernel functions and meta learner is an SVR model

2.2 Gannet optimization algorithm(GOA)

The optimal hyperparameters of GPR kernel function are typically determined by the gradient descent approach, but this method often converges to local optimum. To address this issue, various meta-heuristic algorithms are commonly employed, such as particle swarm optimization (PSO), sparrow search algorithm (SSA), hunger games search (HGS) [46], dragonfly algorithm (DA) [47], etc. In this work, a novel meta-heuristic optimization technique named gannet optimization algorithm (GOA) is used to optimize the hyperparameters of GPR, which models the predatory behavior of gannets through two phases, exploration and exploitation[48]. There are two different sorts of dives in the exploration phase: U-shaped (Eq. (5)) and V-shaped (Eq. (6)), namely

$$a=2t_1\cos(2\pi r_1), \tag{5}$$

$$b=2t_1V(2\pi r_2), \tag{6}$$

where $V(x)=\begin{cases}-\dfrac{1}{\pi}x+1, & x\in(0,\pi),\\ \dfrac{1}{\pi}x-1, & x\in(\pi,2\pi),\end{cases}$

$t_1=1-\dfrac{t}{t_{\max}}$, t and $t_{\max}$ represent the current iteration count and the maximum iteration count respectively, and r_1, r_2 are random numbers within $[0,1]$. The location of the gannet is updated to

$$MX_i(t+1)=\begin{cases}X_i(t)+u_1+a(2r_3-1)[X_i(t)-X_r(t)], & q\geqslant 0.5,\\ X_i(t)+v_1+b(2r_4-1)[X_i(t)-X_m(t)], & q<0.5,\end{cases} \tag{7}$$

where r_3 and r_4 are both random numbers within $[0,1]$; u_1 and v_1 are random numbers generated from intervals $[-a, a]$ and $[-b,b]$; $X_i(t)$ and $X_r(t)$ represent the ith individual and a randomly selected individual from the current population respectively, while $X_m(t)$ denotes the average position of individuals within the same population.

The gannet needs to perform two activities mentioned above, as well as two more, to mature. To avoid the gannet, the fish in the water are typically accompanied by a swift rotating motion. To catch the fish that are frantically attempting to escape, the gannet also uses a lot of energy. The capture capacity is defined by

$$C=\frac{1}{Rt_2}, \tag{8}$$

where $t_2=1+\dfrac{t}{t_{\max}}$, $R=\dfrac{M\times vel^2}{0.2+1.8r_5}$, r_5 is a random number within (0, 1), M=2.5 kg and vel=1.5 m/s represent the weight of the gannet and its speed in the water, respectively.

As the gannet accumulates sufficient energy, it will initiate fish-catching maneuvers. However, over time, the gannet's energy depletes, hindering its ability to successfully complete the capture. The formula for updating its position is

$$X_i(t+1)=\begin{cases} t_1 C\left|X_i(t)-X_{\mathrm{Best}}(t)\right|\left[X_i(t)-X_{\mathrm{Best}}(t)\right]+X_i(t), & C \geqslant c, \\ X_{\mathrm{Best}}(t)-\left[X_i(t)-X_{\mathrm{Best}}(t)\right] Levy(dim) t_1, & C < c, \end{cases} \tag{9}$$

where $c=0.2$ is a constant value that was established after conducting several experiments, $X_{\mathrm{Best}}(t)$ represents the best-performing individual in the current population, and $Levy(\cdot)$ is the $Levy$ flight function determined by

$$Levy(dim)=0.01 \times \frac{\mu\sigma}{|v|^{\frac{1}{\beta}}}, \tag{10}$$

where μ, σ and v are random values between 0 and 1.

2.3 Global forecasting process

The EGPR model is trained and constructed with GOA. In order to evaluate the prediction accuracy, three evaluation indexes are introduced for quantitative analysis as

$$MAE=\frac{1}{N}\sum_{i=1}^{N}\left|y_i-y_i^{*}\right|, \tag{11}$$

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-y_i^{*})^2}, \tag{12}$$

$$R^2=1-\frac{\sum_{i=1}^{N}(y_i-y_i^{*})^2}{\sum_{i=1}^{N}(y_i-\bar{y}_i)^2}, \tag{13}$$

where y_i, $\bar{y}_i$, and y_i^{*} represent the true value, mean of true value and predicted value of the ith cycle, respectively, and N is the number of samples. The range of mean absolute error (MAE) and root mean squared error (RMSE) is $(0,+\infty)$, the closer it is to 0, the smaller the error result of the model; and the range of coefficient of determination (R^2) is $(0,1)$, the closer it is to 1, the better the fitting effect of the model.

As mentioned above, HI extraction is a crucial phase that has an impact on SOH prediction performance as a result of the complex degradation process that goes along with battery packaging[22, 41]. Therefore, it is essential to extract the features most correlated with capacity and use them as inputs for the EGPR model.

In summary, SOH prediction process of LIBs based on EGPR was carried out as shown in Fig. 3 with following steps:

1) From the discharge voltage and temperature curves of the LIB dataset, six HIs were isolated and then filtered using the Pearson correlation coefficient. The selected HIs were used as the EGPR input after normalization.

2) The meta-level regression model was trained using the entire training set, with the result of the base learner serving as its input. The GPR kernel function in base learner was optimized via GOA to avoid entering a local optimum.

3) The SOH of LIBs was predicted with EGPR and the accuracy of prediction results was evaluated.

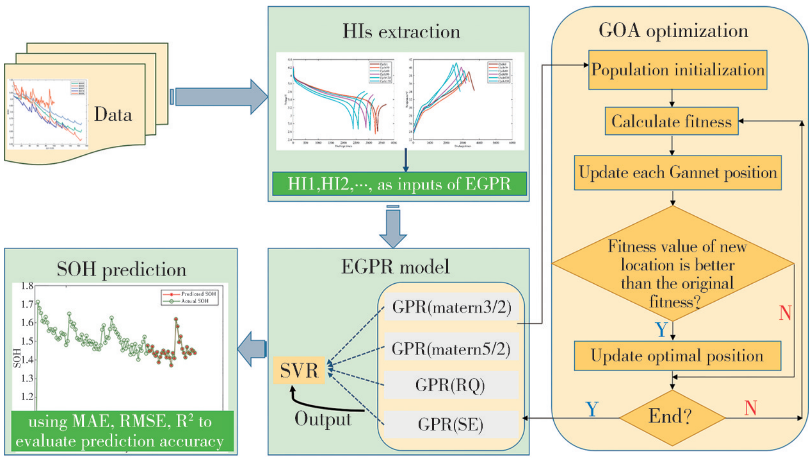
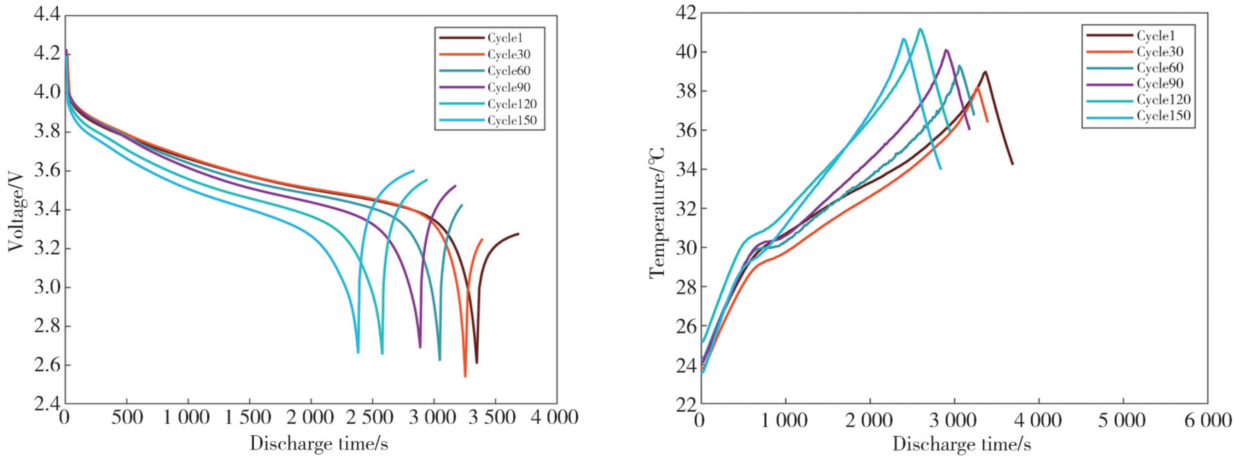


Fig. 3 Framework for SOH prediction of LIBs

3 Results and discussion

3.1 Feature extraction and correlation analysis

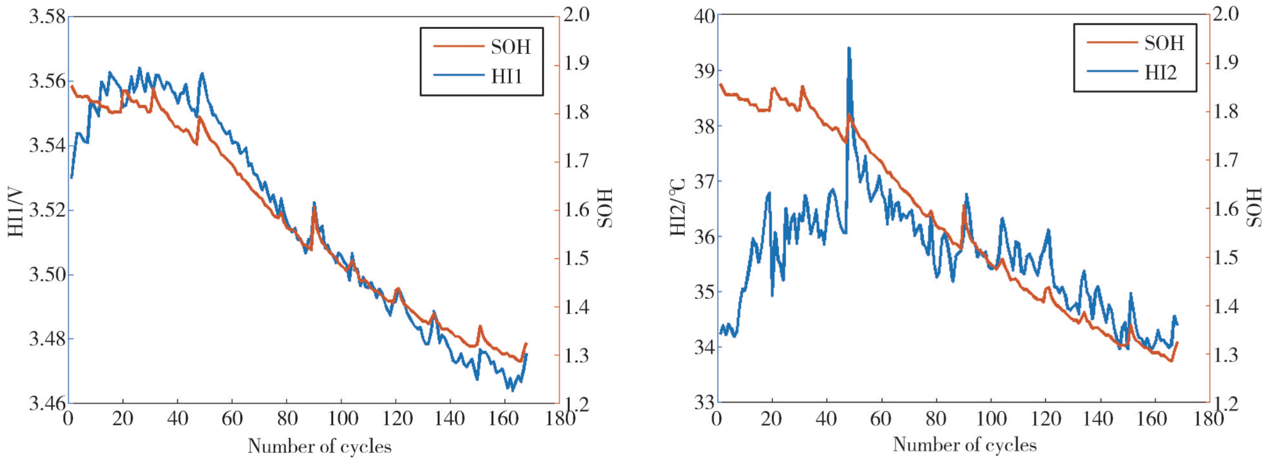
In this study, a battery's SOH is predicted utilizing direct external variables that are simple to get from the BMS and appropriate for dynamic operation as indirect HIs. As seen in Fig.4, which depicts the variation of the depleted battery voltage as well as the temperature under various cycles, the B0005 battery is utilized as an example to extract HIs. The graph shows that, as the number of cycles rises, the discharge voltage trough shifts to the left, indicating a decline in discharge capacity. To extract features from the curves of current, temperature, and time, the peak of the discharge temperature shifts to the left as the number of cycles rises, suggesting that the voltage and temperature change at different phases.



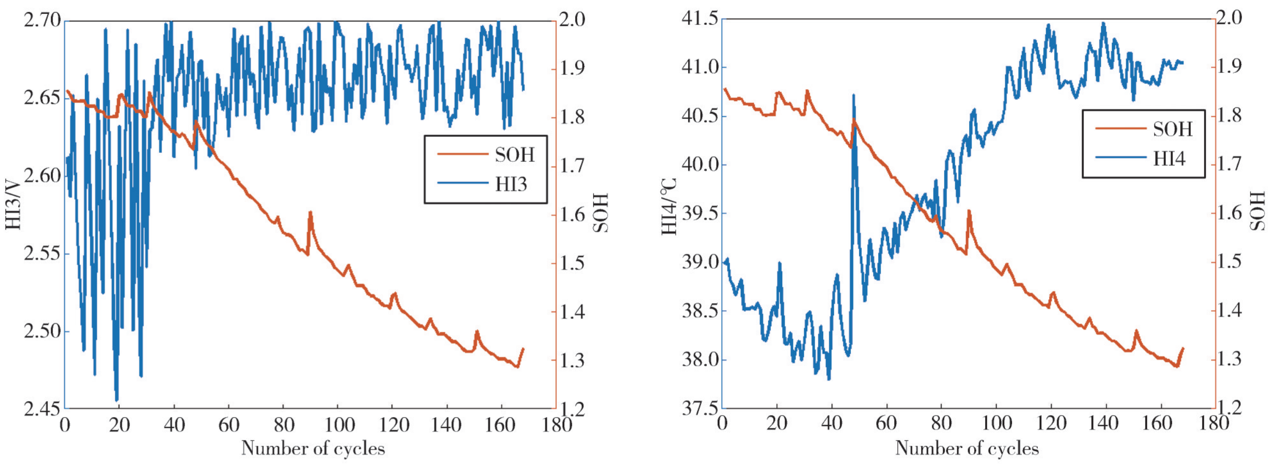
(a) Variation of voltage (b) Variation of temperature

Fig. 4 Variation of battery in different cycles

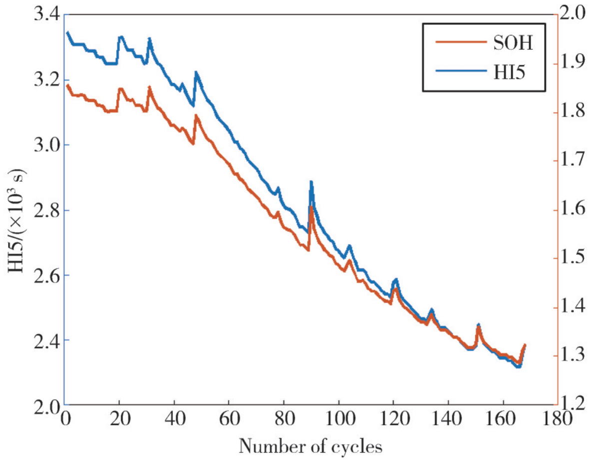
Fig. 5 shows t he His with the number of cycles for B0005. The average discharge voltage is given by HI1; the final discharge temperature is given by HI2; the lowest voltage is given by HI3; the greatest discharge temperature is given by HI4; the time taken to reach the lowest discharge voltage is given by HI5; and HI6 provides the amount of time needed to reach the highest discharge temperature.



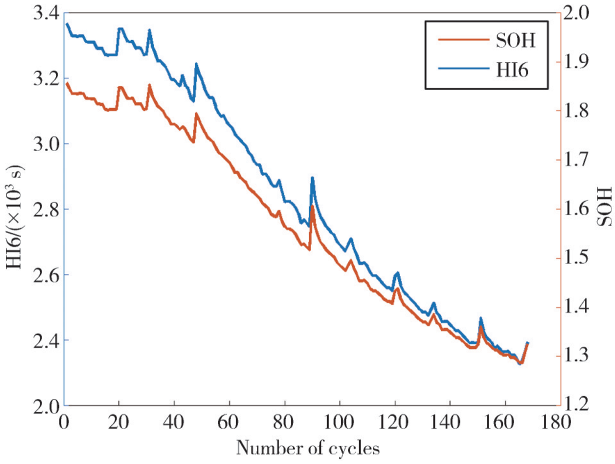
(a) Average discharge voltage (b) Final discharge temperature



(c) Lowest voltage (d) Greatest discharge temperature



(e) Time to reach the lowest discharge temperature



(f) Time to reach the highest discharge temperature

Fig. 5 HIs with the number of cycles for B0005

As shown in Table 2. The absolute values of the Pearson correlation coefficients for HI1, HI4, HI5, and HI6 are above 0.65, indicating a very good correlation with capacity. As a result, these indicators were chosen as the HIs for the SOH prediction in this work and renamed as HI1, HI2, HI3, and HI4.

Table 2 Pearson correlation coefficients of different HIs

HI No.	Pearson correlation coefficient			
	B0005	B0006	B0007	B0018
HI1	0.982 357	0.965 189	0.961 071	0.985 595
HI2	0.586 301	0.934 800	−0.021 81	0.496 173
HI3	−0.487 070	−0.489 720	−0.314 74	0.117 606
HI4	−0.935 270	−0.850 420	−0.749 46	−0.695 210
HI5	0.999 947	0.999 915	0.999 725	0.999 774
HI6	0.999 813	0.999 871	0.999 201	0.999 415

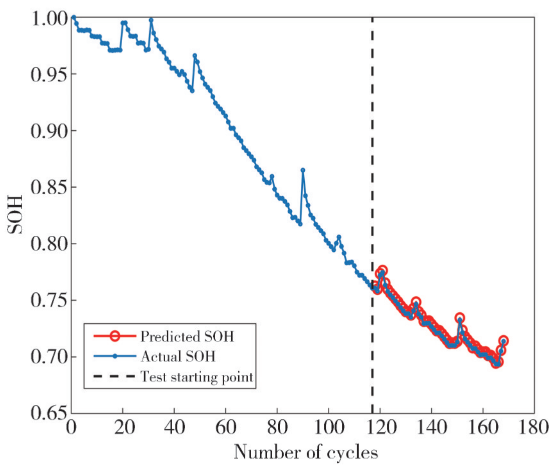
3.2 SOH predicted by EGPR

The EGPR model's parameters are searched using GOA, with the parameter yielding the lowest MSE selected as the ideal model parameter. For the batteries B0005, B0006, B0007, and B0018, the HIs that were obtained in Section 3.1 were normalized. In this study, the first 70% of the sample set constitutes the training set, while the remaining 30% is used as the testing set. Fig. 6 presents the predicted results, showing that the estimated SOH closely matches the actual values for these four batteries.

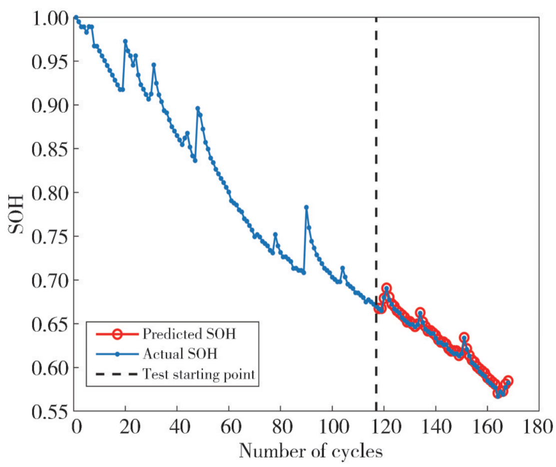
The findings demonstrated a high level of accuracy and minimal error of the EGPR model. The majority of prediction errors fall within the range of −0.01 to 0.01, with the maximum error rate being less than 1%. Table 3 clearly illustrates the model's predictive accuracy, as evidenced by MAE of less than 0.16%, RMSE of less than 0.20%, and R^2 value ranging from 0.988 9 to 0.999 5 across the four batteries.

Table 3 Errors of SOH prediction by EGPR

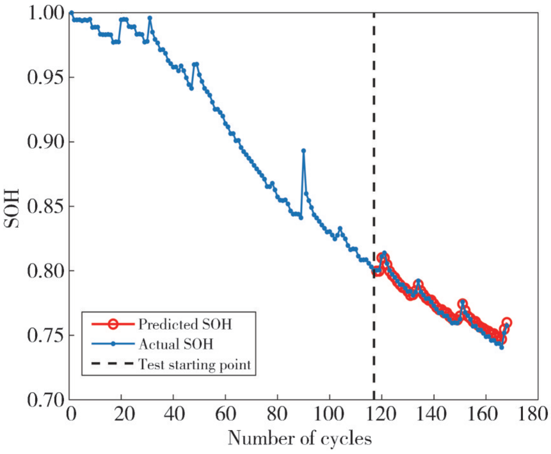
Battery No.	MAE	RMSE	R^2
B0005	0.042 8	0.048 5	0.999 5
B0006	0.159 5	0.192 5	0.996 5
B0007	0.085 7	0.106 4	0.996 9
B0018	0.131 8	0.158 7	0.988 9



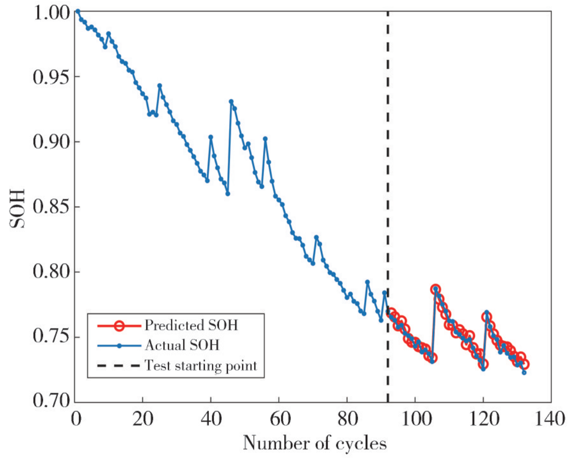
(a) B0005



(b) B0006



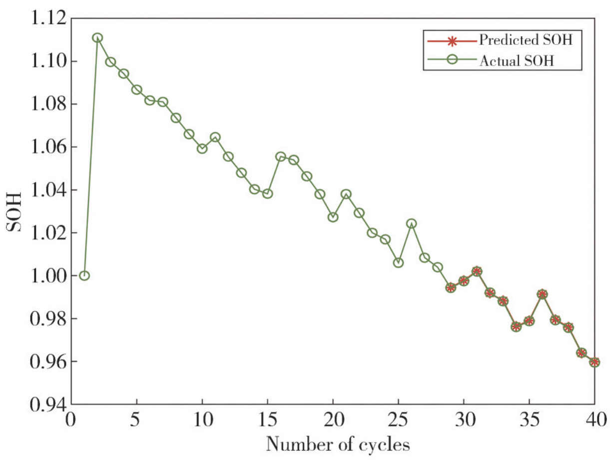
(c) B0007



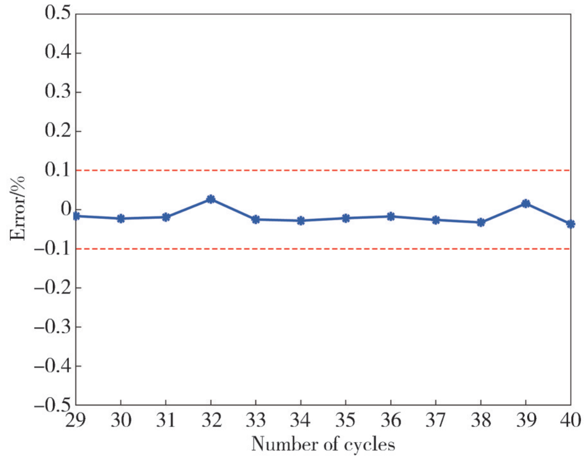
(d) B0018

Fig. 6 SOH prediction results by EGPR for battery

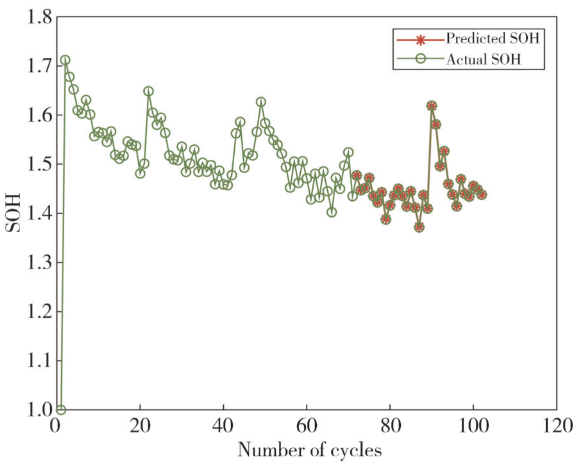
Furthermore, the EGPR model was also tested on two additional datasets, B0032 and B0056, which operates at various temperatures from NASA lithium battery datasets, to evaluate the robustness and generalization ability of the constructed model. The prediction results and relative errors are presented in Fig.7.



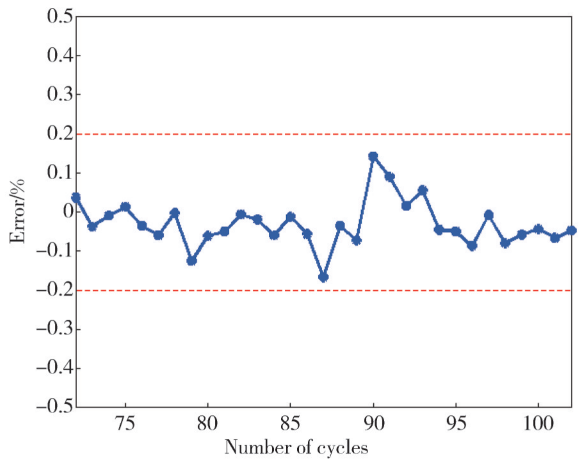
(a) SOH of B0032



(b) Prediction error of B0032



(c) SOH of B0056



(d) Prediction error of B0056

Fig. 7 Prediction results by EGPR for B0032 and B0056

It was observed that the original curve and predicted curve nearly overlap, indicating a high level of prediction accuracy. The model accurately captured the overall degradation trend within a minimal error of less than 0.2%. These findings further demonstrate the excellent robustness, generalization ability and prediction reliability of the proposed model when handling diverse datasets.

3.3 Model comparison

To verify the effectiveness of GOA in EGPR parameter optimization, taking B0005 as an example, GOA was compared with the whale optimization algorithm (WOA), genetic algorithm (GA), and gray wolf optimization algorithm (GWO). The iteration curves are shown in Fig. 8. It can be seen that, compared with WOA, GA, and GWO, GOA has a faster convergence rate and lower error.

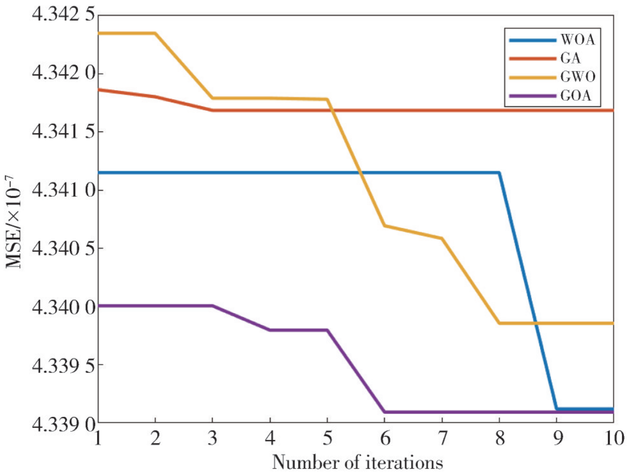


Fig. 8 Comparison of various optimization algorithms for hyperparameters of EGPR

To further validate the model performance, multiple sets of comparison experiments were designed. Firstly, the EGPR model was compared with the Gaussian process regression (GPR) with SE kernel, RQ kernel, Matern32 and Matern52 kernels under the same experimental environment and samples, using the battery B0018 as an example. The experimental results are shown in Fig. 9, where the MSE and RMSE of EGPR are smaller than GPR with a single kernel function, and R^2 is closest to 1, which indicates that EGPR has better learning ability and higher prediction accuracy than individual GPR.

Secondly, the SOH prediction accuracy of EGPR model proposed in this work was compared with other previously developed machine learning regression models such as linear regression, SVR, and Ridge regression under identical experimental settings. Taking battery B0005 as an example, the comparison between the prediction results of four different algorithms and the real value are shown in Fig. 10. It can be seen that, the EGPR algorithm offers more accurate predictions than those produced by other algorithms.

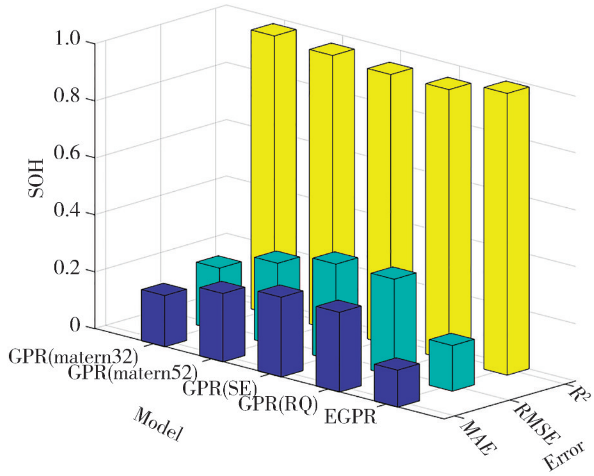


Fig. 9 Comparison of EGPR and GPR with different kernel functions for SOH prediction of battery B0018, where MAE and RMSE are percentage values

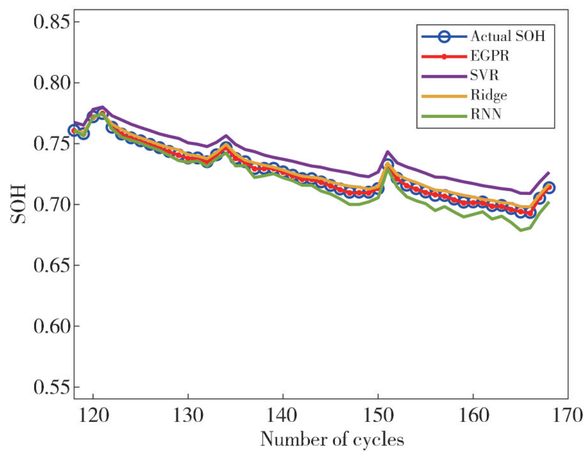


Fig. 10 Comparison of EGPR and other machine learning algorithms for SOH prediction of battery B0005

Table 4 compares the errors produced by the single regression models of Ridge regression, SVR, linear regression, and literature-based machine learning. According to Ref. [49], the charging and discharging curves were used to extract 15 health features, which were then filtered using Pearson correlation coefficients and downscaled using NCA to create the DEGWO-LSTM model. In Ref. [50], the link between HIs and SOH was assessed after using the voltage discharge time and temperature variation throughout the discharge process as HIs. Based on this, a particle swarm optimization-based multicore correlation vector machine (MKRVM) was built to increase the precision of SOH prediction and parameter selection. Zhang et al. [51] employed artificial neural networks (ANN) to predict SOH of LIB by extracting features from incremental capacity curves. From Table 4, MSE and RMSE values of EGPR model are both lower than those of other models, indicating that the proposed HIs and EGPR model can predict the SOH more precisely.

Table 4 Accuracy comparison between EGPR and other ML algorithms for SOH prediction of B0005

Model	MAE	RMSE	R^2
SVR	0.44	0.51	0.885
Ridge	0.21	0.24	0.974 3
RNN	0.62	0.96	0.596 4
DEGWO-LSTM[49]	1.028	1.048	0.941 4
PSO-MKRVM[50]	—	0.48	—
ANN[51]	2.79	4.01	—
EGPR	**0.13**	**0.16**	**0.988 9**

Finally, several ensemble approaches for SOH estimation have been proposed in previous research[52-55]. Table 5 compares and constrasts the model structure, feature sources, and SOH estimation accuracy of those models with the EGPR model proposed in this work for NASA battery datasets. As shown in Table 5, the proposed method offers several advantages over the existing methods. First of all, it exhibits a stronger resistance to noise, which is crucial for accurately predicting the SOH of LIBs. Secondly, the proposed method leverages the extraction of HIs from voltage and temperature data. These HIs capture the complete battery degradation process, providing a comprehensive assessment of the battery's health status. This leads to a more accurate estimation of SOH. Thirdly, the suggested model performs well in terms of SOH estimation accuracy. The model has a low RMSE, which demonstrates that the approach is highly capable of estimating the SOH of LIBs with great accuracy. In summary, the proposed method is effective for dynamic SOH prediction of LIBs due to its strong robustness, thorough HI extraction, and precise SOH estimation.

Table 5 Comparison between EGPR and other ensemble methods for SOH prediction on NASA dataset

Model	Base learner	Meta learner	Feature source	RMSE
Ref.[52]	Extreme learning machine (ELM)	A reliable decision-making rule	Short-term CC voltage curves	0.785
Ref.[53]	SVR	LSTM	Entire CC-CV current and voltage curves	0.38
Ref.[54]	SVR	Linear regression	Short-term CC voltage curves	1.21
Ref.[55]	SVR, GPR, BFNN, RF	Ridge regression	Voltage curves DT and IC curves	0.41
This work	**GPR**	**SVR**	**Entire CC-CV voltage and temperature curves**	**0.13**

4 Conclusions

This study proposed an EGPR-based battery health prediction model to address the challenge of accurately forecasting battery capacity reduction. The SOH of LIBs was precisely assessed, and the experimental batteries' aging parameters were accurately extrapolated.

1) Six HIs were extracted from the discharge voltage and temperature curves of lithium battery dataset. These HIs were then filtered using the Pearson correlation coefficient, and the selected HIs were normalized and used as inputs for the EGPR model.

2) GOA was utilized to optimize the parameters for the kernel function parameter problem in the EGPR, and various optimization strategies were contrasted. The results demonstrate that GOA achieves faster and more accurate convergence.

3) With 70% of the data, EGPR was trained and compared to various ML models. The experimental results indicate that the proposed approach offers excellent precision and strong generalization.

Acknowledgement

This work was supported by Fundamental Research Program of Shanxi Province (No. 202203021211088); and Shanxi Provincial Natural Science Foundation (No.202204021301049).

Declaration of conflicting interests

The authors have no conflict of interests related to this publication.

References

[1] LIN M Q, WU D G, MENG J H, et al. A multi-feature-based multi-model fusion method for state of health estimation of lithium-ion batteries. Journal of Power Sources, 2022, 518: 230774.

[2] TIAN H X, QIN P L, LI K, et al. A review of the state of health for lithium-ion batteries: Research status and suggestions. Journal of Cleaner Production, 2020, 261: 120813.

[3] SHEN S Q, LIU B C, ZHANG K, et al. Toward fast and accurate SOH prediction for lithium-ion batteries. IEEE Transactions on Energy Conversion, 2021, 36(3): 2036-2046.

[4] LIN C P, CABRERA J, YANG F F, et al. Battery state of health modeling and remaining useful life prediction through time series model. Applied Energy, 2020, 275: 115338.

[5] CHEN D, MENG J H, HUANG H Y, et al. An empirical-data hybrid driven approach for remaining useful life prediction of lithium-ion batteries considering capacity diving. Energy, 2022, 245: 123222.

[6] GE M F, LIU Y B, JIANG X X, et al. A review on state of health estimations and remaining useful life prognostics of lithium-ion batteries. Measurement, 2021, 174: 109057.

[7] ZHANG M, YANG D F, DU J X, et al. A review of SOH prediction of Li-ion batteries based on data-driven algorithms. Energies, 2023, 16(7): 3167.

[8] NG K S, MOO C S, CHEN Y P, et al. Enhanced coulomb counting method for estimating state-of-charge and state of health of lithium-ion batteries. Applied Energy, 2009, 86(9): 1506-1511.

[9] KIEL M, SAUER D U, TURPIN P, et al. Validation of single frequency z measurement for standby battery state of health determination//2008 IEEE 30th International Telecommunications Energy Conference, September 14-18, 2008, San Diego, CA, USA. New York: IEEE, 2008: 1-7.

[10] XIONG R, TIAN J P, MU H, et al. A systematic model-based degradation behavior recognition and health monitoring method for lithium-ion batteries. Applied Energy, 2017, 207: 372-383.

[11] ZENATI A, DESPREZ P, RAZIK H. Estimation of the SOC and the SOH of Li-ion batteries by combining impedance measurements with the fuzzy logic inference//36th Annual Conference on IEEE Industrial Electronics Society, November 7-10, 2010, Glendale, AZ, USA. New York: IEEE, 2010: 1773-1778.

[12] LOVE C T, VIRJI M B V, ROCHELEAU R E, et al. State-of-health monitoring of 18650 4S packs with a single-point impedance diagnostic. Journal of Power Sources, 2014, 266: 512-519.

[13] BRESSEL M, HILAIRET M, HISSEL D, et al. Extended Kalman filter for prognostic of proton exchange membrane fuel cell. Applied Energy, 2016, 164: 220-227.

[14] DONG G Z, CHEN Z H, WEI J W, et al. Battery health prognosis using Brownian motion modeling and particle filtering. IEEE Transactions on Industrial Electronics, 2018, 65(11): 8646-8655.

[15] YAN W Z, ZHANG B, ZHAO G Q, et al. A battery management system with a lebesgue sampling-based extended Kalman filter. IEEE Transactions on Industrial Electronics, 2019, 66(4): 3227-3236.

[16] QIU X H, WU W X, Wang S F. Remaining useful life prediction of lithium-ion battery based on improved cuckoo search particle filter and a novel state of charge estimation method. Journal of Power Sources, 2020, 450: 227700.

[17] HE J, WEI Z, BIAN X, et al. State-of-health estimation of lithium-ion batteries using incremental capacity analysis based on voltage-capacity model. IEEE Transactions on Transportation Electrification, 2020, 6(2): 417-426.

[18] RUAN H K, HE H W, WEI Z B, et al. State of health estimation of lithium-ion battery based on constant-voltage charging reconstruction. IEEE Journal of Emerging and Selected Topics in Power Electronics, 2023, 11(4): 4393-4402.

[19] BIAN X, WEI Z, LI W, et al. State-of health estimation of lithium-ion batteries by fusing an open circuit voltage model and incremental capacity analysis. IEEE Transactions on Power Electronics, 2022, 37(2): 2226-2236.

[20] SUN H L, YANG D F, DU J X, et al. Prediction of Li-ion battery state of health based on data-driven algorithm. Energy Reports, 2022, 8: 442-449.

[21] XIA F, WANG K G, CHEN J J. State of health and remaining useful life prediction of lithium-ion batteries based on a disturbance-free incremental capacity and differential voltage analysis method. Journal of Energy Storage, 2023, 64: 107161.

[22] TIAN Y K, WEN J, YANG Y R, et al. State-of-health prediction of lithium-ion batteries based on CNN-BiLSTM-AM. Batteries, 2022, 8(10): 155.

[23] FU P Y, CHU L, LI J H, et al. State of health prediction of lithium-ion battery based on deep dilated convolution. Sensors, 2022, 22(23): 9435.

[24] SHEN S, SADOUGHI M, CHEN X Y, et al. A deep learning method for online capacity estimation of lithium-ion batteries. The Journal of Energy Storage, 2019, 25: 100817.

[25] HU C, JAIN G, SCHMIDT C, et al. Online estimation of lithium-ion battery capacity using sparse Bayesian learning. Journal of Power Sources, 2015, 289: 105-113.

[26] GAO D, HUANG M. Prediction of remaining useful life of lithium-ion battery based on multi-kernel support vector machine with particle swarm optimization. Journal of Power Electronics, 2017, 17(5): 1288-1297.

[27] LI Y, ZOU C F, BERECIBAR M, et al. Random forest regression for online capacity estimation of lithium-ion batteries. Applied Energy, 2018, 232: 197-210.

[28] LI X Y, YUAN C, LI X, et al. State of health estimation for li-ion battery using incremental capacity analysis and gaussian process regression. Energy, 2020, 190: 116467.

[29] WU Y T, XUE Q, SHEN J W, et al. State of health estimation for lithium-ion batteries based on healthy features and long short-term memory. IEEE Access, 2020, 8: 28533-28547.

[30] XIA Z Y, ABU QAHOUQ J A. Lithium-ion battery ageing behavior pattern characterization and state-of-health estimation using data-driven method. IEEE Access, 2021, 9: 98287-98304.

[31] WU T Z, HUANG Y H, XU Y H, et al. SOH prediction for lithium-ion battery based on improved support vector regression. International Journal of Green Energy, 2023, 20(3): 227-236.

[32] CUI S M, JOE I. A dynamic spatial-temporal attention-based GRU model with healthy features for state-of-health estimation of lithium-ion batteries. IEEE Access, 2021, 9: 27374-27388.

[33] MAWONOU K S R, EDDAHECH A, DUMUR D, et al. State-of-health estimators coupled to a random forest approach for lithium-ion battery aging factor ranking. Journal of Power Sources, 2021, 484: 229154.

[34] HU X S, CHE Y H, LIN X K, et al. Battery health prediction using fusion-based feature selection and machine learning. IEEE Transactions on Transportation Electrification, 2021, 7(2): 382-398.

[35] XING J, ZHANG H L, ZHANG J P. Remaining useful life prediction of lithium batteries based on principal

component analysis and improved Gaussian process regression. International Journal of Electrochemical Science, 2023, 18(4): 100048.

[36] GONG D, GAO Y L, KOU Y L, et al. State of health estimation for lithium-ion battery based on energy features. Energy, 2022, 257: 124812.

[37] FENG H L, SHI G L. SOH and RUL prediction of li-ion batteries based on improved Gaussian process regression. Journal of Power Electronics, 2021, 21: 1845-1854.

[38] ZHENG X Y, DENG X G. State-of-health prediction for lithium-ion batteries with multiple gaussian process regression model. IEEE Access, 2020, 7: 150383-150394.

[39] LIU K L, HU X S, WEI Z B, et al. Modified gaussian process regression models for cyclic capacity prediction of lithium-ion batteries. IEEE Transactions on Transportation Electrification, 2020, 5(4): 1225-1236.

[40] GOEBEL K, SAHA B, SAXENA A, et al. Prognostics in battery health management. IEEE Instrumentation & Measurement Magazine, 2008, 11(4): 33-40.

[41] YU Z, LIU N, ZHANG Y, et al. Battery SOH prediction based on multi-dimensional health indicators. Batteries, 2023, 9(2): 80.

[42] RASMUSSEN C E, WILLIAMS C K I. Gaussian processes for machine learning. Cambridge, Mass.: MIT Press, 2006.

[43] LI G Z, LI B, LI C, et al. State-of-health rapid estimation for lithium-ion battery based on an interpretable stacking ensemble model with short-term voltage profiles. Energy, 2023, 263: 126064.

[44] WU H, LEVINSON D. The ensemble approach to forecasting: A review and synthesis. Transportation Research Part C: Emerging Technologies, 2021, 132: 103357.

[45] SHARMA R S, SINGH B, KAUR M. A novel approach of ensemble methods using the stacked generalization for high-dimensional datasets. IETE Journal of Research, 2022, 69(10): 6802-6817.

[46] JIN Z, LI X B, YU D W, et al. Lithium-ion battery state of health estimation using meta-heuristic optimization and Gaussian process regression. Journal of Energy Storage, 2023, 58: 106319.

[47] RICHARDSON R R, OSBORNE M A, HOWEY D A. Gaussian process regression for forecasting battery state of health. Journal of Power Sources, 2017, 357: 209–219.

[48] PAN J S, ZHANG L G, WANG R B, et al. Gannet optimization algorithm: A new metaheuristic algorithm for solving engineering optimization problems. Mathematics and Computers in Simulation, 2022, 202: 343-373.

[49] MA Y, SHAN C, GAO J W, 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.

[50] YANG Y R, WEN J, SHI Y H, et al. State of health prediction of lithium-ion batteries based on the discharge voltage and temperature. Electronics, 2021, 10(12): 1497.

[51] ZHANG S, ZHAI B, GUO X, et al. Synchronous estimation of the state of health and remaining useful lifetime for lithium-ion battery using the incremental capacity and artificial neural networks. Journal of Energy Storage, 2019, 26: 100951.

[52] GOU B, XU Y, FENG X. An ensemble learning-based data-driven method for online state-of-health estimation of lithium-ion batteries. IEEE Transactions on Transportation Electrification, 2021, 7(2): 422-436.

[53] LIU G F, ZHANG X W, LIU Z M. State of health estimation of power batteries based on multi-feature fusion models using stacking algorithm. Energy, 2022, 259: 124851.

[54] GUO Y, HUANG K, YU X, et al. State-of-health estimation for lithium-ion batteries based on historical dependency of charging data and ensemble SVR. Electrochimica Acta, 2022, 428: 140940.

[55] ZHANG B F, XU G, LIU J Y, et al. State-of-health estimation for lithium-ion batteries based on partial charging segment and stacking model fusion. Energy Science Engineering, 2023, 11(1): 383–397.

基于集成高斯过程回归的锂离子电池健康状态预测

惠周利，王瑞洁，冯娜娜，杨　明*

中北大学 数学学院，山西 太原 030051

摘　要：锂离子电池性能会随着时间推移而逐渐下降，实时预测电池的健康状态（State of health，SOH）至关重要。本文提出一种融合间接健康指标和集成高斯过程回归（Ensemble Gaussian process regression，EGPR）的锂离子电池SOH预测模型。首先，通过分析放电过程中电压和温度的变化提炼出能够反应电池退化过程的6个间接健康因子。其次，利用塘鹅优化算法（Gannet optimization algorithm，GOA）对EGPR模型中的参数进行优化并估计SOH。最后，该预测模型在多种实验场景下进行了测试和比较。针对NASA锂电池数据集的仿真实验表明，该方法具有较高的预测精度和泛化能力，均方根误差保持在0.20%以内，平均绝对误差低于0.16%。

关键词：锂离子电池；集成高斯过程回归；健康状态；健康因子；塘鹅优化算法

引用格式：HUI Zhouli, WANG Ruijie, FENG Nana, et al. State of health prediction for lithium-ion batteries based on ensemble Gaussian process regression. Journal of Measurement Science and Instrumentation, 2024, 15(3): 397-407.