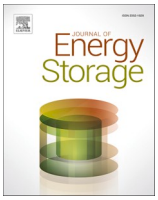
Journal of Energy Storage 62 (2023) 106790

Contents lists available at ScienceDirect

Journal of Energy Storage

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



Research Papers

Early prediction of lithium-ion battery cycle life based on voltage-capacity discharge curves

Wei Xiong [a], Gang Xu [a,*], Yumei Li [a], Feng Zhang [b], Peng Ye [b], Ben Li [c]

[a] School of Mechanical Engineering, Hubei Engineering University, Xiaogan 432000, China
[b] School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070, China
[c] School of Mechanical and Power Engineering, East China University of Science and Technology, Shanghai 200237, China

ARTICLE INFO

Keywords:
Early prediction
Cycle life
Voltage-capacity discharge curves
HIs
WLS-SVM

ABSTRACT

Accurately predicting the lifetime of lithium-ion batteries is critical for accelerating technological advancements and applications. Nevertheless, the complex aging mechanisms and dynamic operating conditions of lithium-ion batteries have remained major challenges. This paper proposes a method for early predicting lithium-ion batteries cycle life based on weighted least squares support vector machine (WLS-SVM) with health indicators (HIs) as input. The HIs are extracted from lithium-ion batteries voltage-capacity discharge curves, since these curves are easy to measure and strongly correlate to battery cycle life. Taking into account the nonlinearity of batteries cycle life, a support vector machine (SVM) that is capable of strong generalization is used to predict cycle life. As a solution to the problem of misleading results from outlier data in SVM, error square term combined with weight functions is used in this study to improve robustness and prediction accuracy. The datasets of 41 cells are used to verify the proposed early prediction method. The results show that the root mean square error (RMSE) and mean absolute error (MAE) of cycle life prediction results by the proposed method are much lower, and the cycle life early prediction errors of the test cells are all <9 %, which notes that the proposed method is accurate and valid for cycle life early prediction.

1. Introduction

Lithium-ion batteries have been widely employed as an energy storage device due to their high specific energy density, low and falling costs, long life, and lack of memory effect [1,2]. Unfortunately, like with many chemical, physical, and electrical systems, lengthy battery lifespan results in delayed feedback of performance, which cannot reflect the degradation of lithium-ion batteries in time [3]. This has posed obstacles to improving the development and use of lithium-ion batteries. In the light of this, accurate lifetime prediction based on early-cycle data is critical, and would offer up new avenues for battery manufacturing, usage, and improvement. Manufacturers, for example, can shorten the cell development cycle, quickly evaluate novel manufacturing techniques, and sort/grade new cells based on their anticipated lifetime. Moreover, end-users may also find out how long their batteries will last [4]. Since lithium-ion batteries are popular, it is increasingly important to predict their lifetimes early. However, it is challenging because of their nonlinear degradation with cycling and wide-ranging variability, as well as the need to meet operating conditions.

Many prior publications have attempted to early predict the lithium-ion battery cycle life. Summarizing these studies, it is not difficult to find that methods for early prediction of lithium-ion battery's cycle life can be categorized into two main types: model-based method and data-driven method [5]. Model-based methods rely on models that describe the internal chemical reactions or degradation mechanisms of the battery, predicting the cycle life by taking into account changes in some specific mechanisms. Among these models, semi-empirical models have been used previously. As early as 2001, Bloom et al. and Broussely et al. simulated power and capacity loss using semi-empirical models [6,7], and identified the model parameters in a specific way to prepare for predicting battery cycle life. In order to predict remaining useful life (RUL), Lian et al. proposed a bat-based particle filter algorithm (PF) applying a semi-empirical model [8], which can achieve acceptable results under certain conditions. To improve prediction accuracy, researchers Xue et al. proposed an integrated algorithm which combines adaptive unscented kalman filter (AUKF) and genetic algorithm optimized support vector regression (GA-SVR) [9]. By using the state space model with double exponential, the RUL was predicted. The results show

* Corresponding author.
E-mail address: xugang524@163.com (G. Xu).

https://doi.org/10.1016/j.est.2023.106790
Received 18 September 2022; Received in revised form 5 January 2023; Accepted 27 January 2023
Available online 13 February 2023

that the proposed method is more accurate in terms of battery remaining lifetime prediction, especially early on [10]. Despite the fact that these mechanism-specific models have achieved some predictive success, it remains difficult to develop models for full cells to predict cycle life and solve the model parameters for early cycle life prediction with only limited data, since there are a variety of degradation modes in a cell, which are coupled to mechanical and thermal heterogeneities. As these challenges cannot be overcome, there are currently few reports on model-based methods for predicting battery cycle life.

Data-driven methods, unlike model-based ones, do not take into account the internal reactions of the battery or analyze the reasons for failure, but instead simply find and learn the degradation law through past experience [11]. Therefore, with technological advancements, data-driven methods are widely employed in the prediction of cycle life, such as neural networks (NNs), Gaussian process regression (GPR), and support vector machines (SVMs), etc. [12]. It has become increasingly common for data-driven techniques to be used in predicting battery lifespan based on data collected both in laboratories and online. With the help of machine-learning tools, Severson et al. calculated and classified cell cycle life using voltage-capacity discharge curves from early cycles. For the first 100 cycles, the best models achieve a 9.1 % test error [3]. Nevertheless, the selected health indicators (HIs) of this prediction method must calculate the capacity change value under the corresponding voltage, which is a very tedious process. In order to simplify the prediction process, Ren et al. developed an integrated deep learning approach for RUL prediction of lithium-ion battery by integrating autoencoder with deep neural network (DNN) [13]. To reach more accurate results, Braco et al. analyzed 58 HIs acquired from incremental capacity analysis, partial charging, constant current and constant voltage stages, and internal resistance [14]. With a multi-dimensional feature extraction method, the proposed approach produced more accurate cycle life predictions than others. Zhang et al. employed the long short-term memory (LSTM) recurrent neural network (RNN) to learn the long-term dependencies among the degraded capacities of lithium-ion batteries [15]. The developed method is able to predict the battery's RUL independent of offline training data. To improve RUL prediction, a prognostic framework shared by multiple batteries is proposed by Li et al. [16]. Compared with other data-driven methods, the experiments carried on NASA dataset demonstrate the proposed method hits lower average root mean square. However, these method is mainly designed for offline prediction [17–19]. At present, many practical applications require not only that the prediction method can be used offline, but also that online prediction can be easily realized. For online prediction, Lin et al., using battery energy storage system monitoring data obtained from the battery management system (BMS), presented a data-driven technique that is capable of estimating the RUL of the battery with a high degree of accuracy and reliability [20]. However, this method requires the collection of a large amount of historical battery charge-discharge cycle data. In order to achieve accurate cycle life prediction in the case of limited early cycle data of the battery, this method still needs to be improved. Tong et al. developed an adaptive dropout long short-term memory (ADLSTM)-based prediction algorithm that combines the features of Monte Carlo simulation with deep learning. By analyzing only 25 % of the degradation data, ADLSTM-MC can achieve accurate prediction [21]. However, for LFP/graphite cells with long cycle life, 25 % of the degradation data is already a lot, and it is not entirely an early prediction. A true early prediction should be based on relatively early cycle data, and the number of cycles should be much $<$25 % of the total cycle life. By using early cycle data, the Long-Short-Term Memory (LSTM) model for early prediction combines Broad Learning System (BLS) and (LSTM) is proposed by Zhang et al. [22]. Yet, this method only uses early cycle life data, whose nonlinear characteristics are quite different from the cycle life data at later stages of the battery. In order to achieve accurate prediction of cycle life in the entire battery life cycle, it is not enough for machine learning methods to learn only early cycle life data. Therefore, this method still has room for improvement in prediction accuracy. In order to improve the accuracy of cycle life prediction, the changes of certain features during battery aging and the correlation between the changes of these features and cycle life have received increasing attention. Based on 19 features extracted from the dQ/dV and dV/dQ curves, Afshari et al. presented a practical method for accurately predicting batteries' RUL at an early stage [4]. Although the extraction of 19 features can improve the prediction accuracy to some extent, it increases the difficulty of execution HIs and reduces the prediction success rate. This method has certain problems, but it also provides a good direction for battery cycle life prediction. Adopting suitable HIs and data-driven method are important to overcome the challenges posed by the nonlinearity of cycle life and to achieve accurate cycle life prediction.

To achieve better predictions, this paper proposes a method for predicting lithium-ion batteries cycle life based on early voltage-capacity discharge curves and weighted least squares support vector machines (WLS-SVMs). The proposed method extracts HIs from voltage-capacity discharge curves; the selected HIs has a high correlation with cycle life and is easy to extract. By incorporating all training data as support vectors, the proposed method can fully characterize the degradation of batteries and utilize weight functions to weight error variables, resulting in high accuracy in early predicting the battery's cycle life. The primarily contributes to this work in the following areas.

1) First, this work extracts the area change in voltage-capacity discharge curves between different cycles as HI. The proposed HI acquisition method is simple and efficient compared to others, and it has a strong correlation with battery cycle life. This ensures the accuracy and feasibility of early prediction.
2) Second, this work uses an WLS-SVM based method to solve the problem of lithium-ion batteries cycle life prediction with limited training data. The proposed method can fully characterize the degradation of batteries and utilize weight functions to weight error variables, achieving precise early prediction.

This paper is organized as follows: Section 2 involves battery cycle life tests and the extraction of the HIs for this study. Section 3 discusses the WLS-SVM algorithm. In Section 4, we report the results of the cycle life prediction and analyze them. The conclusions of this study are summarized in Section 5.

2. Battery cycle life test and HIs extraction

2.1. Dataset of lithium-ion batteries cycle life test

Detailed information about the dataset employed in this paper is given in [3]. In order to obtain the correlation between battery cycle life and certain characteristics, and to provide training and validation data for machine learning methods, this work selects the first batch of experimental battery data from the paper of Severson et al., a total of 41 cells. In this batch of batteries, the data of 30 cells are selected to train the prediction model of the machine learning method. The remaining 11 battery data are used to confirm the prediction method proposed in this paper. These 41 cells that provided data in this paper have reached the end of their life (EOL) at the end of the experiment. The EOL of a cell is determined by the number of cycles it has taken to reach 80 % of its normal capacity. In order to generate the data, 1.1 Ah nominal capacity LFP/graphite cells (A123 Systems) were cycled under variable fast-charging conditions and discharging conditions (4C to 2.0 V, 1C equals 1.1 Amps) under temperature-controlled conditions (30 °C). Data generation is conducted at charging rates ranging from 3C to 8C in order to investigate the characteristics of the cells under a variety of charging conditions. The operating parameters of the batteries are shown in Table 1. As a result, the dataset captures a variety of cycle lives since charging conditions are deliberately varied over a wide range of time. Among these 41 cells, the shortest cycle life is 534 cycles, the longest is

Table 1
Operating parameters of the batteries.

Battery	Observed cycle life	Charging policy	Charging cut-off voltage	Discharging policy	End-off voltage	Category
#1, #2	534, 559	5.4C(80 %)-5.4C	3.6 V	4C	2 V	#1 training data,◆#2 Test data
#3	616	8C(35 %)-3.6C	3.6 V	4C	2 V	Training data
#4, #5	617,625	7C(40 %)-3.6C	3.6 V	4C	2 V	Training data
#6	636	4.8C(80 %)-4.8C	3.6 V	4C	2 V	◆Test data
#7	651	8C(25 %)-3.6C	3.6 V	4C	2 V	◆Test data
#8	691	5.4C(70 %)-3C	3.6 V	4C	2 V	Training data
#9	703	7C(30 %)-3.6C	3.6 V	4C	2 V	◆Test data
#10	704	7C(40 %)-3C	3.6 V	4C	2 V	◆Test data
#11	709	6C(50 %)-3.6C	3.6 V	4C	2 V	Training data
#12	731	6C(60 %)-3C	3.6 V	4C	2 V	Training data
#13	742	7C(30 %)-3.6C	3.6 V	4C	2 V	Training data
#14	757	6C(60 %)-3C	3.6 V	4C	2 V	Training data
#15	788	5.4C(50 %)-3C	3.6 V	4C	2 V	Training data
#16	788	5.4C(70 %)-3C	3.6 V	4C	2 V	Training data
#17	842	6C(40 %)-3.6C	3.6 V	4C	2 V	Training data
#18	854	6C(40 %)-3C	3.6 V	4C	2 V	Training data
#19	857	5.4C(60 %)-3.6C	3.6 V	4C	2 V	◆Test data
#20	860	6C(50 %)-3C	3.6 V	4C	2 V	Training data
#21	862	5.4C(60 %)-3.6C	3.6 V	4C	2 V	Training data
#22	870	4.8C(80 %)-4.8C	3.6 V	4C	2 V	Training data
#23	870	6C(40 %)-3.6C	3.6 V	4C	2 V	Training data
#24	876	6C(50 %)-3.6C	3.6 V	4C	2 V	Training data
#25	879	5.4C(40 %)-3.6C	3.6 V	4C	2 V	◆Test data
#26	880	5.4C(60 %)-3C	3.6 V	4C	2 V	◆Test data
#27	891	6C(30 %)-3.6C	3.6 V	4C	2 V	Training data
#28	897	5.4C(50 %)-3.6C	3.6 V	4C	2 V	Training data
#29	906	5.4C(50 %)-3C	3.6 V	4C	2 V	Training data
#30	917	6C(50 %)-3C	3.6 V	4C	2 V	Training data
#31	966	8C(15 %)-3.6C	3.6 V	4C	2 V	◆Test data
#32	1014	6C(30 %)-3.6C	3.6 V	4C	2 V	Training data
#33	1017	6C(40 %)-3C	3.6 V	4C	2 V	Training data
#34	1051	8C(15 %)-3.6C	3.6 V	4C	2 V	◆Test data
#35	1054	5.4C(40 %)-3.6C	3.6 V	4C	2 V	Training data
#36	1074	4.4C(80 %)-4.4C	3.6 V	4C	2 V	Training data
#37, #38, #39	1177, 1179, 1190	3.6C(80 %)-3.6C	3.6 V	4C	2 V	Training data
#40, #41	1226,1227	4C(80 %)-4C	3.6 V	4C	2 V	◆#40 Test data,#41 training data

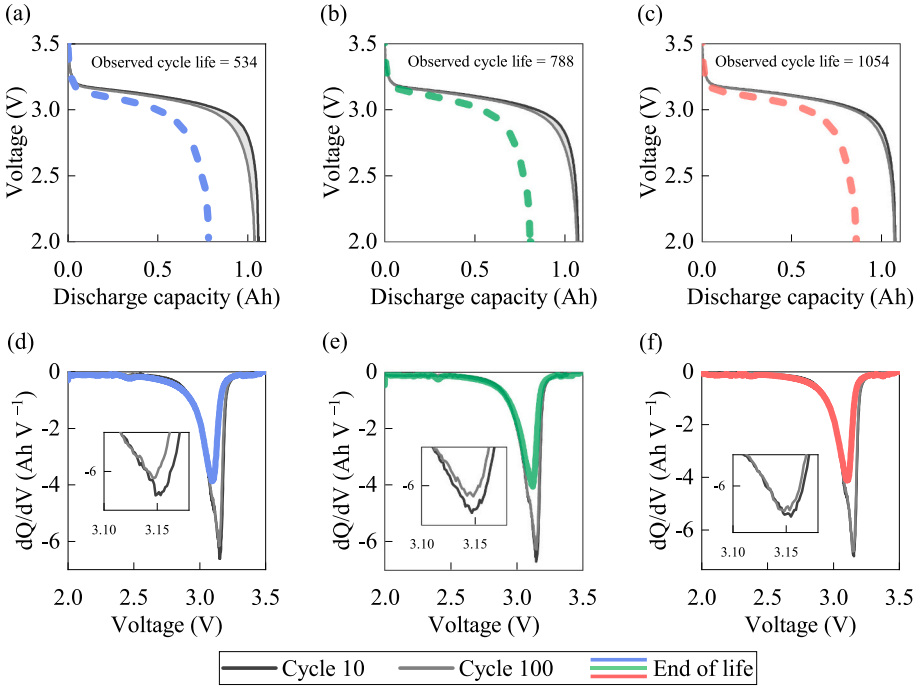


Fig. 1. Voltage–capacity discharge curves and dQ/dV curves. (a) Voltage–capacity discharge curves of cell with a cycle life of 534. (b) Voltage–capacity discharge curves of cell with a cycle life of 788. (c) Voltage–capacity discharge curves of cell with a cycle life of 1054. (d) dQ/dV curves of cell with a cycle life of 534. (e) dQ/dV curves of cell with a cycle life of 788. (f) dQ/dV curves of cell with a cycle life of 1054.

1227 cycles. The cycle life of the remaining cells is randomly distributed between the two.

2.2. Extraction of HIs

In the early cycles, capacity degradation is negligible; however, the voltage-capacity discharge curve changes. It has been demonstrated by Dubarry et al. [23] that degradation modes in LFP/graphite cells result in shifts in dQ/dV and dV/dQ derivatives for diagnostic cycles at C/20 as a result of dQ/dV and dV/dQ degradation modes. According to their findings, one degradation mode (loss of active material of the delithiated negative electrode) results in a shift in discharge voltage without a change in capacity. As a result, losing lithium ions from a delithiated negative electrode does not affect the overall capacity of the electrode, but changes the potentials [23,24].

To capture the relationship between potential and cycle life in early cycling, several features are measured via the discharge voltage curve and differential capacity curve (dQ/dV) as shown in Fig. 1. Fig. 1(a) and (d) show the discharge data for a battery with a cycle life of 534 cycles. The discharge data include the voltage-capacity curves and dQ/dV curves of the 10th cycle, the 100th cycle and the 534th cycle, respectively. And Fig. 1(b) and (e) show the discharge data for a battery with a cycle life of 788 cycles, including data from the 10th cycle, the 100th cycle and the 788th cycle, respectively. Fig. 1(c) and (f) show the discharge data for a battery with a cycle life of 1054 cycles, including data from the 10th cycle, the 100th cycle and the 1054th cycle, respectively. These three batteries were randomly selected. According to the data presented in Fig. 1(a), (b), and (c), the changes in battery voltage-capacity curves appear to reflect battery cycle life. Specifically, we analyze how V(Q) changes from cycle to cycle as a function of voltage. In Fig. 1(a), (b), and (c), from the 10th cycle to the 100th cycle the voltage–capacity discharge curves shift to the left. Upon examining the voltage-capacity discharge curves of the 100th and 10th cycles, it appears that the larger the shift between the two curves, the shorter the battery's cycle life. In other words, based on Fig. 1(a), (b), and (c), it appears that the broader the area change between the 10th and 100th cycle of the voltage-capacity discharge curve, the shorter the battery's cycle life. There might be a correlation between this shifting and battery cycle life in early cycles.

In Fig. 1(d), (e), and (f), there is also a shift in the dQ/dV curves from the 10th cycle to the 100th cycle, and the magnitude of the shift varies with battery cycle life. But these mappings do not seem to be as clear as those between voltage–capacity discharge curves and cycle life. In addition, the dQ/dV curves need to be measured at a low rate of current. And it needs to be smoothed before it can be used, which makes it difficult to achieve efficient application in cycle life prediction. As a result, the dQ/dV curves are not suitable for the extraction of HI.

In light of this, we decided to look for HI from the voltage–capacity discharge curves. According to the intuitive changes in Fig. 1(a), (b) and (c), the area change in voltage-capacity discharge curves between cycles is evaluated. For ease of writing, the area change in voltage-capacity discharge curves between cycles 10 and 100 denote ΔA_{10-100} (Ah·V) = $A_{voltage\_cycle10}$ (Ah·V) − $A_{voltage\_cycle100}$ (Ah·V), where the subscripts denote the cycle number. As can be seen in Fig. 1(a), (b), and (c), as the battery cycle life changes, ΔA_{10-100} also changes accordingly, with the longer the battery cycle life, the smaller the ΔA_{10-100}. To discuss this correlation in detail, as shown in Fig. 2, the cycle life and ΔA_{10-100} of the three cells (the same batteries in Fig. 1) are displayed on the same coordinate system. The cycle life of the three cells is similar to ΔA_{10-100}, and the trend of the cycle life curve seems to correlate well with the ΔA_{10-100} curve. The results of analysis for the three cells show that the correlation between cycle life and the area change of voltage-capacity discharge curves is high. Here are just three randomly selected batteries and cycle intervals for reference. Further analysis of the correlation of the HI, using all the samples and cycle intervals, will be needed.

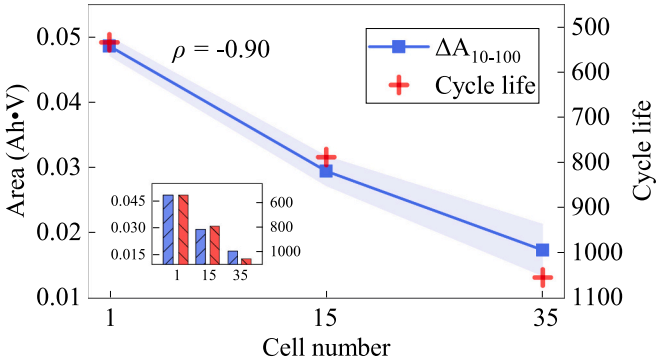


Fig. 2. Cycle life curve and ΔA_{10-100}.

2.3. Correlation analysis

To assess the HI in this paper, Spearman's correlation coefficient is applied. Spearman's correlation coefficient is widely used to evaluate the correlation between variables [25]. It is computed using the original sorting order of the data to solve, ascending or descending [26]. Therefore, when outliers are present in the data, the results are not affected too much [27]. Data expressions do not have to conform to normal distributions, continuous variables, or linear relationships [28]. In general, strong correlation is defined as Spearman's coefficient >0.6 when the absolute values of the two variables exceed this threshold [29,30]. The Spearman's correlation coefficient is

$$d_i = x_i' - y_i' \tag{1}$$

$$r_{xy} = 1 - \frac{6 \cdot \sum_{i=1}^{w} d_i^2}{w \cdot (w^2 - 1)} \tag{2}$$

where x_i' is the sorted position of HIs, y_i' is the battery capacity, y_i' is the sorted position of the battery capacity, w is the sample size, d_i is the difference of the sorted position, r_{xy} is the value of Spearman's correlation coefficient. Based on Eqs. (1) and (2), Spearman's correlation coefficient can be calculated between HIs and battery capacity.

According to the result in Section 2.2, to find out the strongest correlation HI, the correlation between the area change and cycle life is examined in two steps based on Spearman's correlation coefficient in this section. In the first step, the correlation between ΔA_{10-100} and cycle life of all the samples is discussed. As shown in Fig. 3, ΔA_{10-100} of 41 cells are extracted to analyze the correlation between cycle life and ΔA_{10-100}. The cells are sorted by their cycle life, with short cycle life first and long cycle life last. With the increase of battery cycle life, ΔA_{10-100} gradually decreases, showing a negative correlation, and the correlation coefficient is −0.83. The correlation analyses reveal that the area change in voltage-capacity discharge curves between cycle 10 and cycle 100 is highly correlated with cycle life for 41 cells. However, this result only

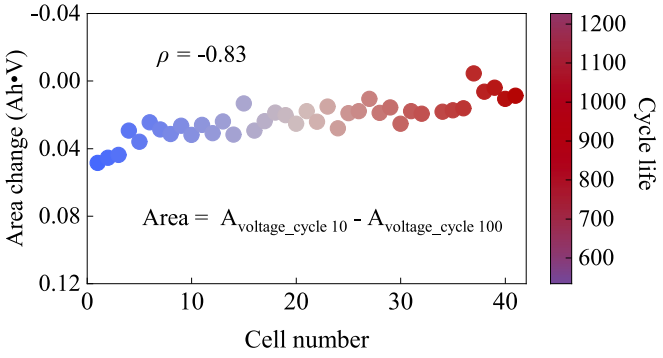


Fig. 3. Cycle life and area change.

shows that the ΔA_{10-100} is highly correlated with cycle life, it does not mean that HIs with higher correlations do not exist between other cycle intervals.

In the second step, the correlation of HI extracted from other cycle intervals is discussed. The area change in voltage-capacity discharge curves is between different numbers of cycles. A change of the cycle interval may result in a change in the area. Therefore, to determine the ΔA (area change in voltage-capacity discharge curves) with the highest correlation to cycle life, it is necessary to take into account the correlation based on different cycle intervals. As shown in Fig. 4, changes in the cycles interval during ΔA extraction lead to changes in the correlation coefficient. In order to more clearly indicate the generalization ability of the selected HI, the sample size is taken as a parameter. In Fig. 4, the sample sizes range from 10 samples to 41 samples. The cycle interval is between the 10th cycle and the n th cycle, and the value of n ranges from 11 to 260. Cycle numbers larger than 260 are not considered, as they do not belong to the early working data of the battery. According to Fig. 4, when the sample size is small, the correlation coefficient changes greatly. And the correlation coefficient is unstable regardless of the change in the cycle interval. When the sample size is >35, starting from cycle interval 100, the wider the cycle interval, the higher the correlation between ΔA and cycle life. The correlation coefficient becomes more stable.

The results of correlation coefficient analysis show that for beyond 35 samples, the larger the cycle interval, the higher the correlation, and the more stable the correlation. When the cycle interval is 100, the correlation coefficient can reach above −0.8, and when the cycle interval is increased to 200, the correlation coefficient can reach above −0.9. According to Spearman's correlation rules, a correlation coefficient above 0.6 indicates a strongly correlated pair of variables. A Spearman's correlation coefficient of above 0.8 indicates a high degree of correlation between the two variables [31]. Therefore, ΔA_{10-100} has a high correlation with cycle life, and ΔA_{10-100} is extracted from the early 100th cycle data, which meets the requirements of early prediction for eigenvectors. ΔA_{10-200} is more correlated with cycle life. Although ΔA_{10-200} is extracted from the early 200th cycle data, which is not regarded as early cycle data for batteries with short cycle life, the 200th cycle data for batteries with longer cycle life can still be counted as early cycle data. As a result, this paper will discuss using ΔA_{10-100} and ΔA_{10-200} as HI to predict battery cycle life, respectively. The predicted results of ΔA_{10-100} and ΔA_{10-200} as HI will be analyzed in Section 4, which is critical for practical applications.

3. WLS-SVM algorithm

In this section, we discuss the research done concerning a method for predicting the cycle life of lithium-ion batteries using WLS-SVMs. The support vector machines (SVM) has high generalization capability, and after training, it learns the relationship between the data. It can predict the nonlinear attenuation of battery capacity over a long period of time

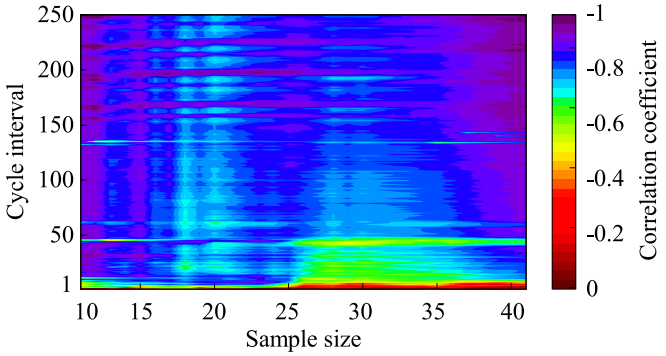


Fig. 4. Correlation coefficient under different conditions.

during its life cycle. However, standard SVMs can provide accurate estimation results if they are used under certain conditions in order to achieve the intended result. Despite this, the SVM lacks robustness due to the possibility of being misled by abnormal training points. In addition, an SVM only utilizes a limited number of data points for training purposes. In the case of abnormal training points, the prediction error using the SVM method will be significantly increased. In this paper, an WLS-SVM based method is proposed in order to address this issue. The purpose of this study is to determine a trustworthy early prediction of cycle life using a limited set of training data which extracts early cycle data from batteries. The proposed method utilizes early cycle data to comprehensively characterize lithium-ion battery degradation and weights the error variables, allowing for cycle life early prediction with a high level of accuracy.

3.1. Principle of SVM algorithm

As mentioned earlier, SVMs were originally developed as a solution to classification problems that needed to be solved. The biggest feature of the SVM is that according to the structural risk minimization criterion, the optimal classification hyperplane is constructed with a maximum classification interval [32]. This improves the generalization ability of the learning machine. When dealing with classification problems, SVMs generate decision surfaces for regions based on samples within these regions, allowing the determination of the category of unknown samples within the region [33,34]. By continuously improving SVMs, they can be utilized to deal with regression problems as well. SVM regression is a supervised learning algorithm used to predict discrete values. The regression and classification principles of SVMs are the same, which are based on finding the best-fit line. During the SVM regression process, the line of best fit is the hyperplane with the most points. A number of practical applications have shown that SVMs perform well for regression analysis, particularly for approximating functions that have high dimensions [35,36].

In order to solve problems involving linear regression, SVMs are used to address convex quadratic programming problems aiming to obtain the most optimal objective function $(x) = w^T x + b$:

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

$$\text{s.t.}\begin{cases} ((w \bullet x_i) + b) - y_i \le \varepsilon + \xi_i, i = 1, \cdots, n \\ y_i - ((w \bullet x_i) + b) \le \varepsilon + \xi_i^*, i = 1, \cdots, n \\ \xi_i \ge 0, \xi_i^* \ge 0, i = 1, \cdots, n \end{cases} \tag{3}$$

where n is the size of the training data set. ξ_i and ξ_i^* describe slack variables.

For simplicity, the primal function, kernel trick, and corresponding constraints may be incorporated into Eq. (3). By using the Lagrange multiplier and Karush-Kuhn-Tucker condition [37], the dual problem can be derived from Eq. (3).

$$\min_{a_i^{(*)}} \frac{1}{2}\sum_{i,j=1}^{n}(a_i^* - a_i)\left(a_j^* - a_j\right)(x_i \bullet x_j) + \varepsilon\sum_{i=1}^{n}(a_i^* + a_i) - \sum_{i=1}^{n} y_i(a_i^* - a_i)$$

$$s.t.\begin{cases} \sum_{i=1}^{n}(a_i^* - a_i) = 0 \\ 0 \le a_i^{(*)} \le C\ i = 1, \cdots, n \end{cases} \tag{4}$$

where a_i and a_i^* are Lagrange multipliers, and only the points that meet the condition $(a_i - a_i^*) \neq 0$ are support vectors. The objective function

can be formulated as:

$$f(x) = \sum_{i=1}^{n} (a_i^* - a_i)(x_i^* \bullet x) + b \tag{5}$$

In order to overcome the problem of nonlinearity and to reduce the level of complexity in the computation, the kernel function is utilized in Eq. (5). The newly defined objective function can be stated as follows:

$$f(x) = \sum_{i=1}^{n} (a_i^* - a_i) k(x_i, x) + b \tag{6}$$

where $K(x_i, x)$ is the kernel function. Since it is capable of handling nonlinear problems well [38], the radial basis kernel is widely used [34]. It can be stated as follows:

$$K(x_i, x) = exp(-\gamma \|x_i - x\|^2) \tag{7}$$

where γ is the radius of the radial basis kernel function. γ is an important factor in determining prediction accuracy.

For this work, grid search is utilized to find the parameter, and cross-validation is used to measure grid search on the training data [39]. Specifically, the minimal square error (MSE) is applied to the cross-validation calculation for evaluating the choice of γ. Whenever the MSE reaches a minimum, it is considered a suitable γ.

$$MSE = \frac{1}{l} \sum_{i=1}^{l} (f(x_i) - y_i)^2 \tag{8}$$

where $f(x_i)$ is the estimated cycle life, y_i is the actual value.

It can be seen from Eq. (4) that only training samples that fulfill the requirement are used as support vectors. It is obvious that not all training data satisfy the stated condition, which is the reason why the SVM is sparse. It is generally considered an advantage that the sparsity of a SVM makes it simpler to carry out the calculations. In spite of this, due to SVM's sparsity, the objective function may not accurately characterize the training data if anomalous training points exist [40,41]. The abnormal training points result in an untimely shift of the decision hyperplane to the abnormal data points, This makes the objective function unable to accurately describe the characteristics of the training data set, and ultimately leads to a decrease in the robustness and accuracy of the prediction results of the support vector machine.

3.2. SVM with weight function to enhance the robustness

To improve the performance of the SVM, the WLS-SVM-based method is developed in this paper. Since the error square term avoids solving the quadratic programming problem, the complexity of the SVM is reduced. The improvement measures start from the loss function associated with machine learning, use the two-norm in the objective function of its optimization problem, and use equality constraints as a replacement for constraints on inequality in the standard SVM. The solution of the optimization problem becomes the solution of a series of linear equations derived from the Kuhn-Tucker condition, thereby reducing the complexity of the calculation process. By weighting the error variables, the accuracy and robustness of the SVM is improved. After the error variable is weighted by the weighting function, the contribution of abnormal training data to the SVM decision hyperplane is limited. This results in the misleading of the prediction result being greatly reduced, thereby improving the robustness and accuracy of the prediction result [1]. As a result, the optimization effort can be redirected to solving Eq. (9).

$$J(w, e_i) = \frac{1}{2} \|w^2\| + \frac{1}{2} C \sum_{i=1}^{N} v_i e_i^2 \tag{9}$$

The equation is constrained by the following:

$$y_i = W^T \varphi(x_i) + b + e_i \tag{10}$$

where v_i is the weight vector. The formula for obtaining v_i is as follows:

$$v_i = \begin{cases} 1 & |e_i/\widehat{S}| \le m_1 \\ \dfrac{m_2 - |e_i/\widehat{S}|}{m_2 - m_1} & m_1 < |e_i/\widehat{S}| \le m_2 \\ 10^{-4} & otherwise \end{cases} \tag{11}$$

where $\widehat{S} = 1.48 MAD(e_i)$ is the standard deviation. In order to minimize Eq. (9), using the Lagrange multiplier a_i, the Lagrange function can be derived as follows:

$$\mathscr{L}(w, a_i, b, e_i) = J(w, e_i) + \sum_{i=1}^{N} a_i \left[y_i - w^T \varphi(x_i) - b - e_i \right] \tag{12}$$

The Karush-Kuhn-Tucker conditions can be summarized [42] as follows:

$$\begin{cases} \dfrac{\partial L}{\partial \omega} = 0 \Rightarrow w = \sum_{i=1}^{N} a_i \varphi(x_i) \\ \dfrac{\partial L}{\partial a_i} = 0 \Rightarrow w^T \varphi(x_i) + b + e_i \\ \dfrac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N} a_i = 0 \\ \dfrac{\partial L}{\partial e_i} = 0 \Rightarrow a_i = \gamma v_i e_i \end{cases} \tag{13}$$

Taking the variables e_i and w out of the equation, Eq. (14) can be derived as follows:

$$\begin{bmatrix} 0 & (l_n)^T \\ l_n & \Omega + I/\gamma v \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{14}$$

where $a = (a_1, a_2, \ldots a_n)$, $\Omega_{ij} = \varphi(x_i)^T \varphi(x_j)$, $y = (y_1, y_2, \ldots y_n)$, $v = diag\ (v_1, v_2 \ldots v_n)$. The WLS-SVM can be obtained as follows:

$$\hat{y}(x) = \sum_{i=1}^{N} a_i K(x_i, x) + \mathrm{b} \tag{15}$$

where $K(x_i, x)$ is the kernel function specified by Eq. (7).

In general, WLS-SVM uses a quadratic loss function, resulting in all the training points becoming support vectors. The weight function can be implemented to weight error variables in the training data. These improvements have greatly improved the prediction robustness and accuracy of the WLS-SVM [42]. Thus, the WLS-SVM can be utilized for the efficient construction of robust models for cycle life early prediction under a variety of working conditions. The procedure for cycle life early prediction based on the proposed method is shown in Fig. 5. The input for the prediction model is ΔA, namely, the area change in voltage-capacity discharge curves between different cycles. The output of the prediction model is the cycle life of prediction. The complete prediction process is:

(1) Randomly classify the training dataset (75 %) and test dataset (25 %) from the battery cycle life test database.
(2) Select the kernel function Eq. (7).
(3) Construct WLS-SVM by using the weight function and the error square term.
(4) Train the WLS-SVM model using the training set.
(5) Verify the WLS-SVM model using the test set.
(6) Evaluate the predicted results of battery cycle life.

Due to applying the two-norm in the objective function of the optimization problem, and employing equality constraints as a replacement

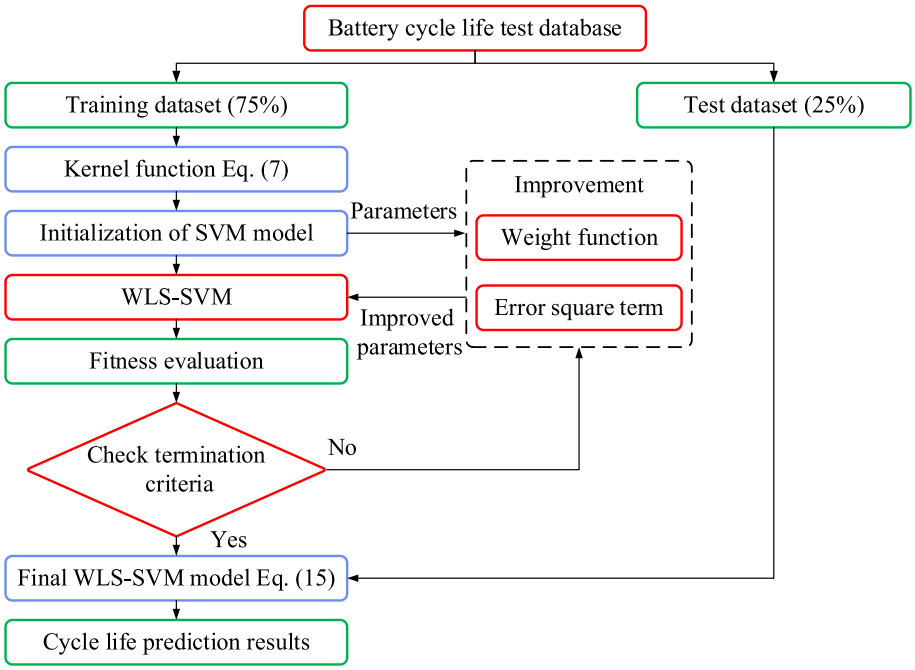


Fig. 5. Early cycle life prediction flowchart based on the WLS-SVM.

for inequality constraints, the optimization problem of the WLS-SVM becomes the solution of a series of linear equations, thereby reducing the complexity of the calculation process, and speeding up the prediction model training time. Once the prediction model has been trained, the process of using the model to predict battery cycle life is efficient. This process can be completed in a short period of time. In the following section, we will analyze the results and evaluate the accuracy of cycle life early prediction.

4. Simulation results and analysis

As a means of verification, the proposed method as well as four representative methods for predicting battery cycle life have been tested. Backpropagation neural network (BPNN) [43,44], GPR [45,46], and SVM compared in this paper are widely adapted to prediction applications. In addition, the linear regression (LR) method is also used to compare with the proposed method [14]. The prediction results of the four methods are compared with the proposed method to demonstrate the advantages of the proposed method. In cycle life early prediction, the inputs for the proposed method are ΔA_{10-100} and ΔA_{10-200}, and the compared methods (BPNN, GPR, and SVM) are ΔA_{10-100}. The output of these data-driven methods is the predicted cycle life. Applying Eq. (16) to Eq. (18), cycle life early prediction results are evaluated based on mean absolute error (MAE) and root mean square error (RMSE), where smaller values indicate better performance.

$$\text{error} = \frac{Cycle\ Life_{k,real} - Cycle\ Life_{k,predict}}{Cycle\ Life_{k,real}} \tag{16}$$

$$\text{MAE} = \frac{1}{N}\sum_{k=1}^{N}|error| \tag{17}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(error)^2} \tag{18}$$

For the purpose of predicting cycle life, the data of 30 cells are used to train the prediction model. The data from the remaining 11 cells is used to evaluate the performance of the prediction method. The results

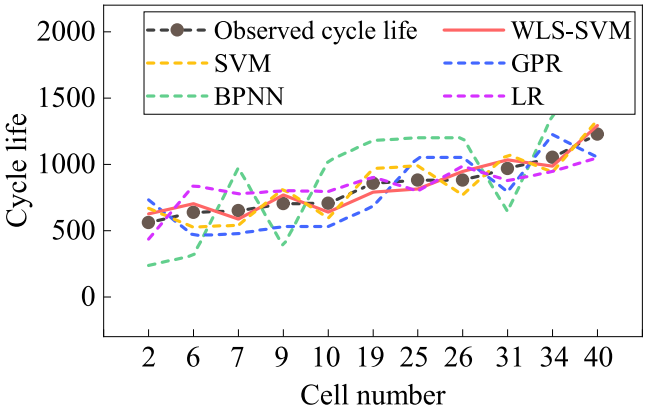


Fig. 6. Cycle life prediction curves using different methods.

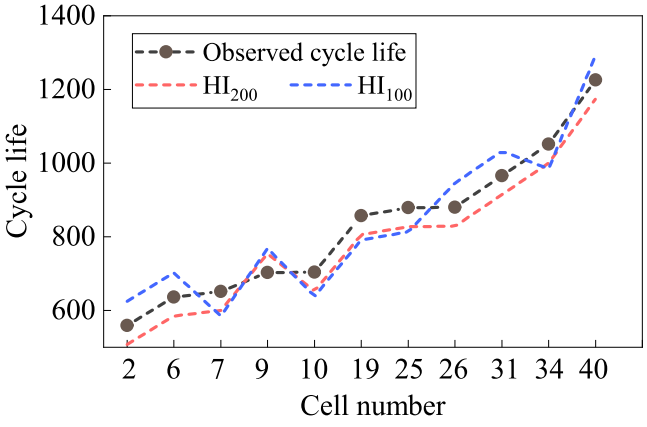


Fig. 7. Cycle life prediction curves using different HIs.

Table 2
The error of cycle life prediction.

Method	WLS-SVM		SVM	LR	GPR	BPNN
	HI_{100}	HI_{200}	HI_{100}	HI_{100}	HI_{100}	HI_{100}
MAE (%)	8.60	8.42	13.98	15.53	22.00	40.80
RMSE (%)	8.83	8.69	14.33	16.12	22.54	41.80

of cycle life prediction are shown in Figs. 6, 7, and Table 2, respectively. In Figs. 6 and 7, the predicted curves of cycle life obtained from the proposed method correspond more closely to the actual values. Applying ΔA_{10-100}, the RMSE of WLS-SVM is 8.83 %, which is significantly lower than the RMSE of BPNN's 41.80 %, the RMSE of GPR's 22.54 %, the RMSE of SVM's 14.33 %, and the RMSE of LR's 16.12 %. The MAE of WLS-SVM based method is also smaller than other methods. The RMSE and MAE of the compared BPNN and GPR methods are larger, which is because the compared machine learning methods rely on a large amount of data to obtain high-precision predictions. But the proposed method, WLS-SVM, can achieve a better machine learning effect and high-precision prediction with a small number of samples. Furthermore, the proposed method weights the error variables, which can effectively suppress the negative impact of abnormal data on the decision hyperplane. As a result, the RMSE and MAE of the WLS-SVM are smaller than the SVM. The cycle life of lithium-ion batteries is strongly nonlinear, so the prediction effect of the LR method has certain limitations. The results show that applying ΔA_{10-100} the proposed method has higher prediction accuracy than the compared methods. Applying two different HIs, the MAE of WLS-SVM is 8.60 % and 8.42 % respectively. Using ΔA_{10-200}, the cycle life prediction is more accurate, the RMSE is 0.14 % lower than that of ΔA_{10-100}. By utilizing ΔA_{10-200}, the cycle life prediction effect is approximately the same as that of ΔA_{10-100}. In this case, it would be more appropriate to use ΔA_{10-100}, as it is able to significantly shorten the test time. Using ΔA_{10-100}, the early prediction errors of the proposed method for the cycle life of the test cells are all <9 %. The results show that the early prediction method proposed in this article for lithium-ion battery cycle life is feasible and effective.

5. Conclusions

In this work, a novel method has been proposed for accurate early prediction of lithium-ion battery cycle life. The novelties arise from using the area change in voltage-capacity discharge curves during early cycles as the HI and applying the WLS-SVM-based method. First of all, since the degradation mode (the loss of active material from the delithiated negative electrode) of lithium-ion batteries results in a change in discharge voltage without a change in capacity in the early cycles, the area change in voltage-capacity discharge curves is extracted as HI, which is efficient to obtain. Then applying Spearman's correlation coefficient, the correlation between HIs and battery cycle life of all the samples is examined. The results show that HIs, such as ΔA_{10-100}, and ΔA_{10-200}, extracted from the early cycle data have a high correlation with battery cycle life. Moreover, utilizing weight functions to weight error variables and transforming the quadratic programming problem into solving linear equations to reduce computational complexity, the WLS-SVM based method is employed to train the prediction model and predict battery cycle life. The proposed WLS-SVM based method can solve cycle life early prediction with limited training data and fully characterize the degradation of batteries. Finally, by using lithium-ion battery data from early cycles under various aging conditions, the proposed method is verified. The results of the proposed method are evaluated in comparison with different machine learning methods, such as BPNN, GPR, and SVM. The results show that applying ΔA_{10-100} the RMSE of WLS-SVM is 8.83 %, which is significantly lower than the RMSE of BPNN's 41.80 %, the RMSE of GPR's 22.54 %, the RMSE of SVM's 14.33 %, and the RMSE of LR's 16.12 %. The early prediction errors of the proposed method for cycle life are all <9 %. The proposed method in this paper is feasible and effective, which can be applied to lithium-ion battery production, use, and optimization in the future.

CRediT authorship contribution statement

Wei Xiong: Conceptualization, Methodology, Software. **Gang Xu:** Methodology. **Yumei Li:** Data curation, Writing – original draft. **Feng Zhang:** Visualization, Investigation. **Peng Ye:** Software, Validation. **Ben Li:** Writing – review & editing.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by the Natural Science Project of Xiaogan (No. XGKJ2022010092).

References

[1] W. Xiong, Y.M. Mo, C. Yan, Online state-of-health estimation for second-use lithium-ion batteries based on weighted least squares support vector machine, IEEE Access 9 (2021) 1870–1881.
[2] W. Xiong, Y.M. Mo, C. Yan, Lithium-ion battery parameters and state of charge joint estimation using bias compensation least squares and the alternate algorithm, Math. Probl. Eng. 2020 (2020).
[3] K.A. Severson, P.M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, et al., Data-driven prediction of battery cycle life before capacity degradation, Nat. Energy 4 (2019) 383–391.
[4] S.S. Afshari, S. Cui, X. Xu, X. Liang, Remaining useful life early prediction of batteries based on the differential voltage and differential capacity curves, IEEE Trans. Instrum. Meas. 71 (2022) 1–9.
[5] D.W. Pan, H.F. Li, S.J. Wang, Transfer learning-based hybrid remaining useful life prediction for lithium-ion batteries under different stresses, IEEE Trans. Instrum. Meas. (2022) 71.
[6] I. Bloom, B.W. Cole, J.J. Sohn, S.A. Jones, E.G. Polzin, V.S. Battaglia, et al., An accelerated calendar and cycle life study of Li-ion cells, J. Power Sources 101 (2001) 238–247.
[7] M. Broussely, S. Herreyre, P. Biensan, P. Kasztejna, K. Nechev, R.J. Staniewicz, Aging mechanism in Li ion cells and calendar life predictions, J. Power Sources 97–8 (2001) 13–21.
[8] Y.C. Lian, J.V. Wang, X.L. Deng, J.Q. Kang, G.R. Zhu, K. Xiang, Remaining useful life prediction of lithium batteries using semi-empirical model and bat-based particle filter, IEEE Int Symp Circ (2020) 1–5.
[9] Z. Xue, Y. Zhang, C. Cheng, G. Ma, Remaining useful life prediction of lithium-ion batteries with adaptive unscented Kalman filter and optimized support vector regression, Neurocomputing 376 (2020) 95–102.
[10] L.S. Yan, J. Peng, D.Z. Gao, Y. Wu, Y.J. Liu, H. Li, et al., A hybrid method with cascaded structure for early-stage remaining useful life prediction of lithium-ion battery, Energy 243 (2022).
[11] Y.C. Song, D.T. Liu, C. Yang, Y. Peng, Data-driven hybrid remaining useful life estimation approach for spacecraft lithium-ion battery, Microelectron. Reliab. 75 (2017) 142–153.
[12] J.T. Qu, F. Liu, Y.X. Ma, J.M. Fan, A neural-network-based method for RUL prediction and SOH monitoring of lithium-ion battery, IEEE Access 7 (2019) 87178–87191.
[13] L. Ren, L. Zhao, S. Hong, S. Zhao, H. Wang, L. Zhang, Remaining useful life prediction for lithium-ion battery: a deep learning approach, IEEE Access 6 (2018) 50587–50598.
[14] E. Braco, I. San Martin, P. Sanchis, A. Ursúa, D.-I. Stroe, Health indicator selection for state of health estimation of second-life lithium-ion batteries under extended ageing, J. Energy Storage (2022) 55.
[15] Y. Zhang, R. Xiong, H. He, M.G. Pecht, Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries, IEEE Trans. Veh. Technol. 67 (2018) 5695–5705.
[16] P. Li, Z. Zhang, Q. Xiong, B. Ding, J. Hou, D. Luo, et al., State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network, J. Power Sources 459 (2020).

[17] L. Ren, J. Dong, X. Wang, Z. Meng, L. Zhao, M.J. Deen, A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life, IEEE Trans. Ind. Inform. 17 (2021) 3478–3487.
[18] G. Ma, Y. Zhang, C. Cheng, B. Zhou, P. Hu, Y. Yuan, Remaining useful life prediction of lithium-ion batteries based on false nearest neighbors and a hybrid neural network, Appl. Energy 253 (2019).
[19] X. Li, L. Zhang, Z. Wang, P. Dong, Remaining useful life prediction for lithium-ion batteries based on a hybrid model combining the long short-term memory and elman neural networks, J. Energy Storage 21 (2019) 510–518.
[20] D. Lin, Y. Zhang, X.H. Zhao, Y.J. Tang, Z.R. Dai, Z.H. Li, et al., Early prediction of remaining useful life for grid-scale battery energy storage system, J. Energy Eng. (2021) 147.
[21] Z.M. Tong, J.Z. Miao, S.G. Tong, Y.Y. Lu, Early prediction of remaining useful life for lithium-ion batteries based on a hybrid machine learning method, J. Clean. Prod. (2021) 317.
[22] M. Zhang, L.F. Wu, Z. Peng, The early prediction of lithium-ion battery remaining useful life using a novel long short-term memory network, in: C Ind Elect Appl, 2021, pp. 1364–1371.
[23] M. Dubarry, C. Truchot, B.Y. Liaw, Synthesize battery degradation modes via a diagnostic and prognostic model, J. Power Sources 219 (2012) 204–216.
[24] D. Ansean, M. Dubarry, A. Devie, B.Y. Liaw, V.M. Garcia, J.C. Viera, et al., Fast charging technique for high power LiFePO4 batteries: a mechanistic analysis of aging, J. Power Sources 321 (2016) 201–209.
[25] M. Alipour, S.S. Tavallaey, A.M. Andersson, D. Brandell, Improved battery cycle life prediction using a hybrid data-driven model incorporating linear support vector regression and gaussian, ChemPhysChem 23 (2022).
[26] R.R. Ardeshiri, R. Razavi-Far, T. Li, X. Wang, C.B. Ma, M. Liu, Gated recurrent unit least-squares generative adversarial network for battery cycle life prediction, Measurement 196 (2022).
[27] D.Q. Chen, W.C. Hong, X.Z. Zhou, Transformer network for remaining useful life prediction of lithium-ion batteries, IEEE Access 10 (2022) 19621–19628.
[28] Y. Chen, W.X. Duan, Z.H. Ding, Y.L. Li, Battery life prediction based on a hybrid support vector regression model, Front. Energy Res. (2022) 10.
[29] J.W. Hu, B. Lin, M.F. Wang, J. Zhang, W.L. Zhang, Y. Lu, Health factor analysis and remaining useful life prediction for batteries based on a cross-cycle health factor clustering framework, J. Energy Storage (2022) 50.
[30] H. Li, L.J. Fu, Y. Zhang, Lithium battery remaining useful life prediction based on multi-kernel support vector regression with hybrid optimization algorithm, J. Electrochem. Energy (2022) 19.
[31] D.D. Ge, Z.D. Zhang, X.D. Kong, Z.P. Wan, Extreme learning machine using bat optimization algorithm for estimating state of health of lithium-ion batteries, Appl. Sci. (2022) 12. Basel.
[32] A. Ara, M. Maia, F. Louzada, S. Macedo, Regression random machines: an ensemble support vector regression modelwith free kernel choice br, Expert Syst. Appl. (2022) 202.
[33] P. Divya, B.A. Devi, Hybrid metaheuristic algorithm enhanced support vector machine for epileptic seizure detection, Biomed. Signal Process. (2022) 78.
[34] D. Faccini, F. Maggioni, F.A. Potra, Robust and distributionally robust optimization models for linear support vector machine, Comput. Oper. Res. 147 (2022).
[35] M.B. Haile, A.O. Salau, B. Enyew, A.J. Belay, Detection and classification of gastrointestinal disease using convolutional neural network and SVM, Cogent Eng. (2022) 9.
[36] D. Mou, Z.W. Wang, X.L. Tan, S.Y. Shi, A variational inequality approach with SVM optimization algorithm for identifying mineral lithology, J. Appl. Geophys. 204 (2022).
[37] M.S.P. Sharma, U. Meena, G.K. Sharma, Intelligent data analysis using optimized support vector machine based data mining approach for tourism industry, ACM Trans. Knowl. Discov. Data 16 (2022).
[38] Y. Simhamed, F. Ykhlef, A. Iratni, A new classification scheme based on extended Kalman filter and support vector machine, Electr. PowerSyst. Res. (2022) 210.
[39] J.F. Villa-Manriquez, R.Y. Sato-Berru, J. Castro-Ramos, J.L. Flores-Guerrero, Classification of trimethylamine-N-oxide, a cardiometabolic disease biomarker, by Raman spectroscopy and support vector machines, J. Phys. D. Appl. Phys. 55 (2022).
[40] C. Yang, S.K. Oh, B. Yang, W. Pedrycz, L. Wang, Hybrid fuzzy multiple SVM classifier through feature fusion based on convolution neural networks and its practical applications, Expert Syst. Appl. (2022) 202.
[41] H.J. Yu, X. Zhou, X.L. Zhang, M. Mooney, Enhancing earth pressure balance tunnel boring machine performance with support vector regression and particle swarm optimization, Autom. Constr. (2022) 142.
[42] L. Wang, J.H. Li, W.B. Zhang, Y. Li, Research on the gas emission quantity prediction model of improved artificial bee colony algorithm and weighted least squares support vector machine (IABC-WLSSVM), Appl. Bionics Biomech. 2022 (2022).
[43] Z. Tian, W.L. Gan, X.Z. Zou, Y. Zhang, W.Z. Gao, Performance prediction of a cryogenic organic Rankine cycle based on back propagation neural network optimized by genetic algorithm, Energy 254 (2022).
[44] D.H. Wu, H.M. Huang, S.J. Qiu, Y. Liu, Y.Z. Wu, Y. Ren, et al., Application of Bayesian regularization back propagation neural network in sensorless measurement of pump operational state, Energy Rep. 8 (2022) 3041–3050.
[45] O. Ali, M.K. Ishak, A.B. Ahmed, M.F.M. Salleh, C.A. Ooi, M.F.A.J. Khan, et al., On-line WSN SoC estimation using gaussian process regression: an adaptive machine learning approach, Alex. Eng. J. 61 (2022) 9831–9848.
[46] H. Chen, K. Ahmadi, Estimating pose-dependent FRF in machining robots using multibody dynamics and Gaussian process regression, Robot. Cim.Int. Manuf. (2022) 77.