Int. J. Electrochem. Sci., 17 **(2022)** Article Number: 22059, doi: 10.20964/2022.05.14

International Journal of
ELECTROCHEMICAL
SCIENCE
www.electrochemsci.org

A Novel Approach for the Open-circuit Voltage Estimation of Lithium-ion Batteries by epsilon SVR

Bin Xiao[1], *Houmin Wu*[1], *Xiubin Zhang*[1], *Rongjie Wu*[2], *Yonggui Liu*[3,*]

[1] School of Information Engineering, Guangzhou Vocational College of Technology &Business, Guangzhou 511442, China
[2] Paitan Middle School of Zengcheng District, Guangzhou 511385, China
[3] College of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China
[*]E-mail: auygliu@scut.edu.cn

Received: 3 January 2022 / *Accepted:* 26 February 2022 / *Published:* 5 April 2022

Open-circuit voltage (OCV) estimation is an essential challenge for battery management systems (BMSs). Accurate OCV estimation can benefit the state estimation of lithium-ion batteries (LiBs), including state of charge (SoC) and state of health (SoH) estimations. When obtaining a SoC-OCV curve, a battery test method based on low current discharge cannot overcome the effect of battery polarization, and the discharge time usually takes several hours. Hence, an ε-support vector regression (ε-SVR) model for battery OCV estimation was proposed in this research. In accordance with the voltage relaxation behaviour of LiBs, the sample data were collected by hybrid pulse power current (HPPC) experiments on LiBs under different ageing degrees. The features were selected from the data samples using grey association analysis (GRA), and the hyperparameters of the ε-SVR model were obtained by K-folding cross-validation (CV). To validate this approach, the proposed model was trained and tested over the dataset acquired from the LiBs with varying degrees of ageing. Based on the experimental results, the model only needs some short-term battery characteristic data to achieve high-precision OCV estimation.

Keywords: Lithium-ion batteries; Open circuit voltage; Support vector regression; Cross-validation; hybrid pulse power current

1. INTRODUCTION

With their excellent performance, LiBs are widely applicate in various industrial applications such as EVs, portable devices, and smart grids [1]. The OCV is associated with the natural properties of an LiB, which depend on the Gibbs energy from the battery's electrochemical reactions. Precise estimation of the battery OCV is necessary for the BMS to monitor battery status, which includes SoC estimation [2], SoH estimation [3], and state of power (SoP) estimation [4]. When estimating the battery SoC, the SoC

initialization and recalibration of estimated results always depend on the Coulomb counting-based estimation together with an OCV-based method. The voltage curve of the SoC-OCV can be adopted to acquire the battery ageing degree for battery SoH estimation through differential voltage analysis [5, 6]. In addition, the correlation between the battery SoC-OCV is also an essential characteristic parameter to obtain the peak current for battery SoP estimation [7]. Hence, OCV is related to many aspects of battery technology and plays a key role.

1.1. Review of OCV Estimation Approaches

Direct measurement is one of the standard methods for battery OCV estimation. The polarization voltage of a battery usually decreases slowly during an experimental test and even disappears after enough relaxation time. As a result, the battery OCV can be measured directly from the cell terminal voltage. The method for measuring a battery OCV with a low rate current uses a tiny minimal constant current to test the [8] discharge of the battery. The battery SoC changes continuously during the test, and then the battery SoC-OCV curve can be obtained. However, although the current is tiny and minimal, the polarization effect of the battery still exists due to the prolonged continuous current excitation. Significantly, in the battery SoC end range, the accuracy of the OCV is greatly affected [9]. During an incremental OCV characteristic test, the battery is discharged incrementally, and then a timed relaxation process is carried out [10-11]. According to previous conclusions in the literature [8], the reliability of the SoC-OCV correlation obtained by the incremental OCV characterization test is higher than that obtained by the low-rate current OCV characterization test. The main reason is that in the incremental OCV characterization test, the polarization effect of the battery is almost eliminated after a long relaxation process. Obviously, due to the long relaxation process time of the battery, it takes a while to use the incremental OCV characterization test method to obtain the SoC-OCV curve in the whole range.

For more accurate estimation, the model-based method provides a new method for battery OCV estimation. Generally, this method uses the voltage relaxation model to fit the cell terminal voltage after the interruption current and can use the voltage relaxation model to predict the OCV of the cell. Wladislaw Waag et al. proposed a new method [12] estimating the electromotive force (EMF) by using the OCV relaxation process within only a few minutes after the current interruption. After online fitting of the OCV relaxation model and measured OCV relaxation curve, the method is verified by LiPB SoC and SoH estimation. Lei Pei et al. developed a new model based on an improved linear resistor–capacitor (RC) model and proposed a fast battery OCV prediction method [13]. The final battery static OCV was predicted within a few minutes by a linear regression technique. Kun Qian et al. used a second-order equivalent circuit model to describe the voltage relaxation process [14] and obtained, in Ohms, characteristic parameters of electrochemical and concentration depolarization. These parameters are alive to electrochemical states and can provide the basis for battery SoH estimation. Although the method can quickly obtain the OCV, model parameter recognition is needed, and the expansibility is not good.

Based on modern control theory, various observers or filters, such as the extended Kalman filter (EKF), H-infinity filters and adaptive observers, can be used to implement battery OCV estimation. To solve the problem of straightness in the middle of the OCV-SoC curve, Chang Zhang et al. proposed two

model-based real-time SoC estimation methods [15]. A weighted recursive least square (WRLS) method was proposed to achieve real-time battery OCV estimation, and another approach used the EKF to achieve state estimation. To solve the effect of truncation error and further improve the accuracy of battery SoC estimation, Pan Haihong and others addressed the state prediction problem according to the grey prediction model [16], presenting a new OCV model based on cubic Hermite interpolation. In reference [17], the authors used H-infinity filters to extract OCV-SoC relationships from existing current–voltage measurements and validated the method under constant current and dynamic conditions. According to the Lyapunov stability criterion, Yi-Hsien Chiang et al. developed an adaptive control algorithm to achieve accurate estimation of battery OCV and battery internal resistance [18]. Moreover, the proposed method was not limited by the input signal and enhanced the applicability in an onboard power supply system. By obtaining a battery online OCV, these methods avoid the effect of voltage polarization caused by a long relaxation process. However, the accuracy of the obtained OCV depends on the model adopted. Because of the limited computing power and storage of onboard microcontrollers, the models have difficulty ensuring the accuracy of OCV estimation. In addition, the design of an observer or filter and its parameter tuning are tedious and time-consuming processes.

In the literature, there are other types of battery OCV estimation methods. Fangdan Zheng and others studied the relationship between battery OCV-SoC, which depends on temperature, and its effect on SoC estimation results [19]. Alexander Farmann et al. studied the various factors affecting battery OCV estimation in detail [20]. It turns out that ambient temperature and the degree of battery ageing can seriously affect the relaxation behaviour of a battery. There is a linear relationship between the cell relaxation time and ambient temperature. To actively reduce the polarization voltage, Jufeng Yang et al. proposed an improved OCV test method [21]. Based on third-order ECM, two sets of current pulses are used to accelerate the convergence of the battery terminal voltage. Compared with the incremental OCV testing method, the test time is effectively shortened.

The direct measurement method is suitable for engineering testing [8-11]. Because of the existence of the battery relaxation process, it takes a great deal of time to measure. The method, based on OCV modelling, can estimate the battery OCV quickly [12-14], but it does not have good expansibility. A method based on the observer or filter can estimate OCV online [15-18]. However, the complex design and parameter adjustments present challenges to engineers. For battery testing through current excitation, the battery terminal voltage obtained involves not only OCV but also overvoltage and is related to a series of polarization effects. Due to the effect of battery polarization, it is difficult to achieve accurate direct OCV measurements. Moreover, battery relaxation may even last for several hours after the current is interrupted. As a result, the end voltage of a cell can be approximately equal to the OCV [22]. After an equal length of battery relaxation, the polarization effect almost disappears. It is difficult to achieve high-precision and high-efficiency OCV measurements in the complete battery SoC range. Furthermore, with the rapid development of EVs, the problem of recycling and ladder utilization of retired power batteries needs to be solved.

To conquer these challenges, an approach implemented by ε-SVR is proposed for LiB OCV estimation. The main contributions of the OCV estimation of this research are as follows.

A new method for LiB OCV estimation is proposed. The ε-SVR model just needs some short-term sample data and is able to work efficiently for accurate measurement, and it does not rely on complex mathematical calculations, the battery ECM, or inconvenient parameter tuning.

Based on the relaxation process of the LiB, the samples in the battery degradation process are measured by a set of current pulses that take only a few minutes. This non-destructive and non-invasive method for short-term feature acquisition is simple to realize in engineering applications.

2. MODELLING AND METHODOLOGY

2.1. Battery OCV and Voltage Relaxation Behaviour

Generally, battery OCVs can be classified into three categories: battery static balance voltage (OCV), battery dynamic equilibrium voltage (DEV), and battery quasi-equilibrium voltage (QEV). The static balance voltage of the battery reflects the inherent balance between the chemical potential and the potential between the positive and negative electrodes in the battery [23]. The dynamic equilibrium voltage of the battery, which is added to the phase formation process of the counter-electrode particles and the phase exchange process between different particles, is a relatively stable part of the dynamic operating voltage of the battery [24]. The quasi-balance voltage of the battery is an approximation of the static equilibrium voltage of the battery.

Generally, the distribution of ions and electrons in an electrode is unbalanced during an experimental test of battery charging and discharging. This is named battery polarization [25]. Because of the phenomenon of battery polarization, the electrode potential deviates from the equilibrium potential. As a result, there are server main manifestations of the battery [26]: the first is ohmic polarization, which leads to a decrease in battery voltage due to the resistance of the electrode, electrolyte and diaphragm. The second is the electrochemical polarization formed by the charge transfer of the battery, which is related to the lithium ions in the battery. The third is the concentration polarization, which is determined by the lithium ions in the solid phase diffusion of the battery. When the battery test process stops, the battery voltage still does not reach a stable state during the relaxation time, as shown in Fig. 1.a, which can be understood as the depolarization process of the battery.

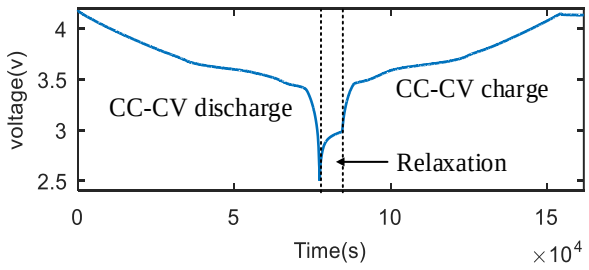


Figure 1.a. Constant-current and constant-voltage (CC-CV) discharge/charge test of lithium-ion battery

Fig. 1 (b) shows a typical voltage relaxation curve of an LiB, in which the battery voltage is discharged to 3.3 at a rate of 0.1 C according to the polarization phenomenon of the battery. Then, the discharge current is terminated, and the battery voltage enters the relaxation process. The battery voltage rises sharply from the cut-off voltage V0 to V1 at the last moment of battery discharge, which is caused by the rapid response of ohmic depolarization and the redistribution of electrons. The process from V1 to V2 is the activation and concentration depolarization process of the cell. Due to the kinematic differences in the diffusion of lithium ions at the electrode–electrolyte interface and the active materials, the voltage relaxation process will be affected by the depolarization process of the battery. Because the voltage relaxation phenomena include direct information about lithium-ion and electron transport, they can also be applied to battery SoH estimation.

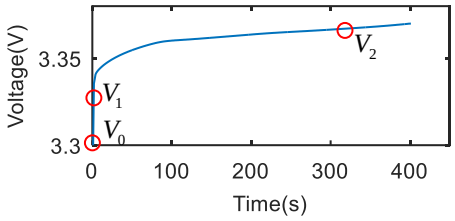


Figure 1.b. A typical voltage relaxation curve

2.2. Problem Formulation of OCV

As normal, the dynamic behaviour of the battery is able to be described by an equivalent circuit model (ECM) realized by several RC combinations, as shown in Fig. 2, which is a second-order RC cell ECM model.

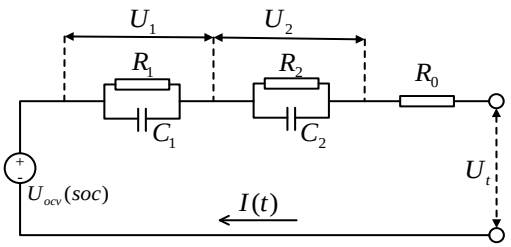


Figure 2. The second-order battery ECM

The battery ECM model consists of three parts: the ohmic resistance (R_0), U_{ocv}, and the polarization part, which includes two RC-equivalent circuits. The parameters U_0, U_1, and U_2 are the ohmic overpotential, diffusion overpotential, and charge transfer overpotential, respectively. Thus, in the process of battery relaxation, the battery voltage can be defined as a relaxation voltage:

$$U_t = U_{ocv} + U_1 + U_2 + U_0 \quad (1)$$

The ohmic overpotential U_0 release rate is extremely fast, and it can be considered that the ohmic overpotential disappears after the external battery circuit is disconnected. The release speed of U_1 and U_2 is relatively slow, and it will gradually decrease as the OCV time increases. Thus, Eq. (1) can be rewritten as follows:

$$U_{rlx,k} = \begin{cases} U_{ocv} + U_{1,k} + U_{2,k}, & 0 < t_k \le T_{ct} \\ U_{ocv} + U_{1,k}, & T_{ct} < t_k \le T_d \\ U_{ocv}, & T_d < t_k \end{cases} \quad (2)$$

where t_k, T_d, and T_{ct} are the battery open-circuit, diffusion polarization, and transfer polarization times, respectively.

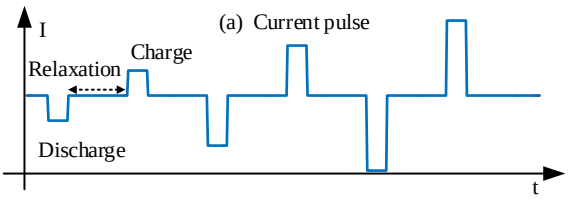
2.3. Feature Construction

During the charging and discharging process of the battery test, the distribution of ions and electrons in the electrode is unbalanced, resulting in battery polarization. According to the polarization phenomena, a typical voltage relaxation curve for an LiB is easy to obtain in battery test processes. Thus, the battery voltage response under a mixed-pulse current test, containing discharging and charging, is used for feature construction. A test schematic is shown in Fig. 3, which includes the current pulse in Fig.3 (a), the voltage response in Fig. 3 (b), and voltage relaxation curve in Fig. 3 (c).

The mixed-pulse current is charged at a current rate of 1 C and discharged at a current ratio of 1 C, and then charged and discharged at 2 C and 3 C, respectively. In Fig. 3 (a), the duration of the battery current pulse is 10 s, the pulse relaxation time is 30 s, and the amplitudes are 1C, 2C, and 3C. Furthermore, a relaxation time occurs after each pulse discharge and charge. Under the current pulse test, the battery voltage response can be acquired. Then, the voltage relaxation curve, as shown in Fig. 3 (c), is able to acquire from the discharge and charge response voltages. Finally, the mean, maximum, summation, and time of the gradient from the voltage relaxation curve reduced to zero are analysed as features. Through short-term current testing, these characteristics can be conveniently extract in practical engineering applications.

(1) When discharging at 1C, F11–F14 represent the mean, maximum, summation, and the time of the gradient from the voltage relaxation curve reduced to zero; when charging at 1C, F15–F18 represent the mean, maximum, summation, and the time of the gradient from the voltage relaxation curve reduced to zero.

(2) In the same way, F21–F28 and F31–F38 features can be obtained. According to this feature construction method, a total of 24 sample features are obtained.



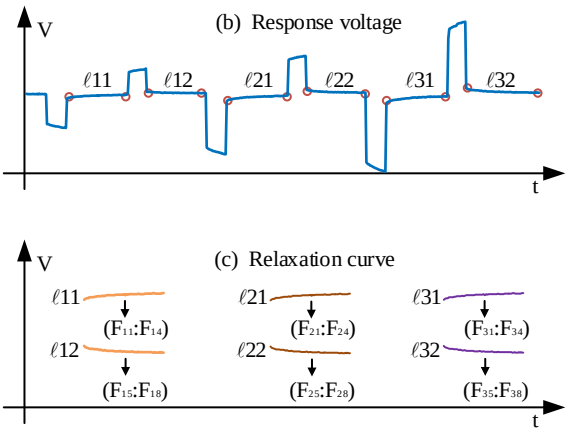


Figure 3. The schematic of the current pulse test

2.4. Feature Extraction

Sample feature selection is a vital part of modelling based on machine learning (ML). However, with the increasing number of data features, the comprehensive performance of the model may not be improved accordingly. Moreover, the final performance of the model rely mostly on the way the sample features are constructed and the number of features. Thus, to research the relation between the sample features and OCV, the GRA method [27] is applied for data sample feature selection. Based on grey system theory (GST), the relational grade is measured on the basis of the similarity between each factor. Hence, the GRA method implements a quantitative measure of system evolution and provides an analytical method for dynamic process analysis.

(1) A sequence for reference $Y=\{y(k)\,|\,k=1,2,...,n\}$, where $y(k)=OCV(k)$. The comparative sequence is $X_i=\{x_i(k)\}$, here $X_i=F_i$;

(2) Data normalization process;

(3) Compute the relational coefficients:

$$\xi_i(k)=\frac{\min\limits_i\max\limits_k|y(k)-x_i(k)|+\rho\max\limits_i\max\limits_k|y(k)-x_i(k)|}{|y(k)-x_i(k)|+\rho\max\limits_i\max\limits_k|y(k)-x_i(k)|} \tag{3}$$

where ρ is the identification coefficient, and $\rho\in(0,1)$

(4) Compute the relational grade r:

$$r_i=mean(\sum_{i=1}^{n}\xi_i(k)) \tag{4}$$

The correlation between the sample features and the battery OCV can be obtained under the above calculation process of the GRA method, as shown in Table 1. The closer the correlation grade is to 1, the greater the correlation is. The results in the table show that the correlation grade of the selected sample features is close to 1. Thus, there is a high correlation grade between the data features chosen and the battery OCV, which shows that it provides a useful sample feature for constructing the model to realize the battery OCV estimation. There may be redundant items in the selected data sample features, but the features calculated by the GRA method will not affect the final battery OCV estimation.

Table 1. The GRA between features and OCV

Fij	1	2	3	4	5	6	7	8
F1	0.723	0.722	0.723	0.727	0.753	0.744	0.753	0.705
F2	0.72	0.738	0.72	0.73	0.766	0.719	0.766	0.707
F3	0.719	0.727	0.719	0.726	0.768	0.734	0.768	0.724

2.5. The Battery OCV Estimation Model

To establish the model, the battery OCV is taken as the state variable in the estimation model, and the mean, maximum, summation, and time of the gradient from the voltage relaxation curve reduced to zero are used as the input variables. Then, the core problem of battery OCV modelling is to establish the relationship between these variables and battery OCV estimation, which can be expressed as follows:

$$OCV(n) = f(F_{1i}, F_{2i}, F_{3i}) + w, i = 1, 2, ..., 8 \tag{5}$$

The function $f(F_{1i}, F_{2i}, F_{3i})$ is adopted to describe the battery OCV estimation model, and w represents the model error. The SVR model is very good at solving problems related to small sample size and nonlinear and high-dimensional pattern recognition [28]; it has been widely used in practical applications, so the SVR algorithm is employed to approximate the complex electrochemical reaction process. A well-trained SVM regression model can be used to predict physical parameters. The following is a given training dataset.

$$\begin{cases} T = \{(x_1, y_1), \cdots, (x_2, y_2), \cdots (x_n, y_n)\} \\ x_i \in X = R^m, y_i \in Y = R, (i = 1, 2, \cdots, n) \end{cases} \tag{6}$$

With the inner product of the kernel function, an SVR can quickly achieve nonlinear transformation from the space of inputs to a higher dimensional feature space. Thus, the nonlinear regression problem can be achieved through this method to attain high-dimensional feature space mapping and then be converted into a linear regression problem, which can be expressed as follows.

$$f(x) = w * \varphi(x) + b \tag{7}$$

In the above formula, $f(x)$ is the approximation for OCV estimation, and x, w, b and $\varphi(x)$ are the input data, weight, offset item and feature space, respectively. SVR must obtain a function that can maximize the deviation of $f(x)$ with a dataset, where all the deviations are less than a given parameter ε. Therefore, the problem can be converted to the following formula:

$$\begin{cases} \min R(w, C, \varepsilon) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} L_\varepsilon(y_i - f(x_i)) \\ L_\varepsilon(y - f(x)) = \begin{cases} 0 & if\ |y - f(x)| \le \varepsilon \\ |y - f(x)| - \varepsilon & otherwise \end{cases} \end{cases} \tag{8}$$

L_ε represents the ε-insensitive loss function. When the deviation is less than ε, no penalty will be imposed. When the deviation is greater than ε, the corresponding penalty will be imposed.

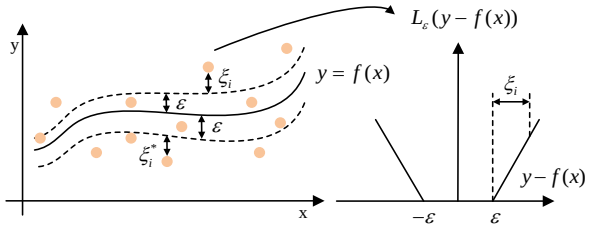


Figure 4. The schematic diagram of ε-SVR

The parameter C is a penalty factor. Its role is to weigh the flatness of the model and empirical risk. To facilitate the calculation of the ε insensitive loss function, ξ_i and ξ_i^* slack variables are needed, as shown in Fig. 4. Then, the solving problem for model parameters can be transformed into the following convex optimization problem:

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$$
$$s.t.\begin{cases} y_i - w\cdot\varphi(x) - b \le \varepsilon + \xi_i \\ w\cdot\varphi(x) + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{9}$$

According to Karush Kuhn Tucker's conditions, the Lagrange function is introduced. The following dual constrained function can be obtained:

$$\min \frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)[\varphi(x_i)\cdot\varphi(x_j)]$$
$$+\varepsilon\sum_{i=1}^{N}(\alpha_i^* + \alpha_i) - \sum_{i=1}^{N} y_i(\alpha_i^* - \alpha_i)$$
$$s.t.\begin{cases} \sum_{i=1}^{N}(\alpha_i^* - \alpha_i) = 0 \\ \alpha_i, \alpha_i^* \in [0, C], i = 1, 2, ...n \end{cases} \tag{10}$$

When the Lagrange multipliers are obtained, the above model can be transformed into the following expression:

$$f(x) = \sum_{i=1}^{N}(\alpha_i^* - \alpha_i)K(x_i, x) + b \tag{11}$$

Here, $K(x_i, x)$ is the kernel function; as calculated by Eq. (12), introducing a kernel function can avoid the complex calculation of the $\varphi(x)$ transformation.

$$K(x, x_i) = \Phi(x)^T\Phi(x_i) \tag{12}$$

The kernel functions of SVR are linear, polynomial, radial basis and sigmoid kernel functions. In this paper, we chose the most commonly used kernel function, the radial kernel function (RBF).

$$K(x, x_i) = \exp(-\gamma\cdot\|x_k - x_i\|^2) \tag{13}$$

Parameter *gama* is the variance of the kernel function. Using the above analysis, the battery OCV estimation model can be expressed as Eq. (14).

$$OCV_{n,estimation} = \sum_{i=1}^{N}(\alpha_i^* - \alpha_i)\exp(-\gamma\cdot\|x_n - x_i\|^2) + b \tag{14}$$

The OCV estimation model based on ε-SVR is established by Eq. (14), and its structure diagram is shown below.

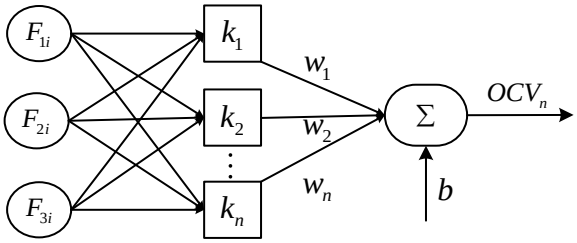


Figure 5. ε-SVR model for battery OCV estimation

2.6. Hyperparametric Optimization for the SVR Model

For better estimation accuracy of the battery OCV, optimal kernel parameters and penalty factors ought to be selected before SVR model training and testing. Therefore, parameters C, gamma and ε are selected by K-fold CV [29]. Depending on the parameters, in model training, the K datasets are chosen as the validation set, and then the remaining K-1 datasets form the training set. Therefore, the trained model can obtain K possible estimators and K corresponding validation errors, and the average value of the K validation error is the final validation error of the model. By combining grid search (GS) methods, the last error of the training model is minimized to select the best model hyperparameters. Five CV processes are shown in Fig. 6. The yellow part of the data represents the validation set of the model, and the data in the purple section represent the training set of the model.

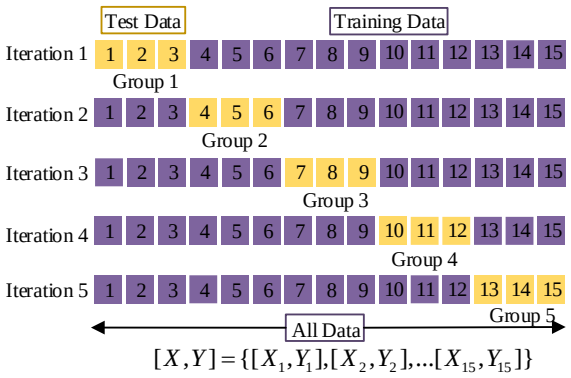


Figure 6. A schematic of K-fold CV

3. RESULTS AND DISCUSSION

This section mainly introduces the experimental analysis results, including data collection and description, the process of model training and verification, and experimental results.

3.1. Data Acquisition

The platform for battery testing is shown in Fig. 7 and is employed to perform task configuration and battery testing. The test bench is composed of a lithium-ion battery, an electronic load test system and a computer for user testing. The battery is connected to the electronic load system and the DC power test system, and the test data are acquired by the computer to carry through the battery test process. In the experiment, 18650 lithium-ion batteries were tested, and their cycle life was between 1000 and 1500 cycles. Other battery parameters are listed in Table 2.

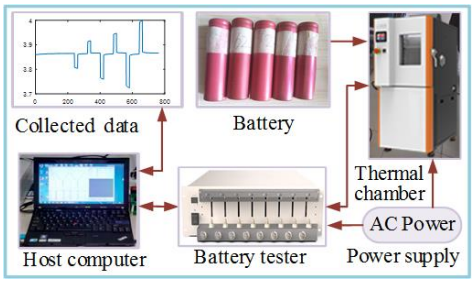


Figure 7. The experimental setup of the test platform

The CC-CV charge/discharge method is used for the battery ageing test. The charging and discharging currents are 5C (A), and the battery can be quickly aged with a large multiplier for the charging and discharging tests, while a specific ageing result can be obtained after multiple cycles.

Table 2. Cell parameters

Brand name	Battery weight	Nominal capacity	Nominal voltage	Charge/discharge cut-off voltage
Sanyo	45 g	2.4 Ah	3.7 V	4.2 V/3.0 V

After 100, 350, 500, 750, and 900 cycle battery ageing tests, battery sample data are collected by the mixed-pulse current test. The batteries are discharged at a current rate of 1 C, charged at a current rate of 1 C and then discharged and charged at 2 C and 3 C, respectively. The duration of pulse charging and discharging is 10 s. Each pulse has a relaxation time of 30 s after charging and discharging. Moreover, the depth of discharge (DoD) is 5%, 10%, 15%, 20%, and 25%. A total of 12 batteries were tested in the experiment, and the sampling time was set to 1 s. According to the CC-CV test method, the battery is quickly aged, and the aged battery is charged and discharged according to 5 different discharge depths. A sum of 60 data samples are gathered for model training, and 3 test sets (20 samples) are randomly selected for model verification.

3.2. Model Training

The experiment for OCV estimation model was conducted in Windows 8.0, with the MATLAB 2017 [30] and LibSVM Toolbox [31] platforms. The toolbox is implemented in the C/MATLAB development language and can be applied to ε-SVR simulation. The sequence sample data in which the datasets are sent to the estimation model is modified based on random functions.

To prevent the various dimensions of the original data from influencing the ε-SVR model training, data normalization have to be performed. Data normalization is implemented to clarify this effect and increase the convergence rate in the process of model training. The approach of min-max normalization is used to scale the data between 0 and 1 by the following equation:

$$x' = (x - \min_x) / (\max_x - \min_x) \tag{15}$$

where x is the raw data value, x' is the scaled data value, and max_x and min_x are the maximum and minimum data values, respectively.

Adaptability evaluation plays a significant role. Generally, the model of ε-SVR is selected as the evolutionary standard, and the CV mean square error (MSE) is used to evaluate the ε-SVR performance. Specifically, the CV MSE is defined as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(OCV_i - OCV_i^*)^2 \tag{16}$$

where OCV_i is the measured value, OCV_i^* is the estimated value, and n is the number of samples.

To select a suitable kernel function, the ε-SVR model is configured with a linear kernel (LK), polynomial function kernel (PNK), and radial basis function (RBF) kernel. A 5-fold CV is performed to optimize the model parameters. Table 3 provides the CV results, showing the best ε-SVR parameters and CV MSE errors of the models. The model configured with the LK produces a large CV error, which shows that this model has poor comprehensive performance on the validation set. The CV error of the model with the PNK is 5.35%, which is a little larger than that of the model with the LK. This shows that the ε-SVR model with LK or PNK cannot achieve the desired result. However, because of the introduction of the Gaussian function (GF), the structure of the RBF model is more sophisticated than that of the other models. Therefore, the CV error of the model configured with RBF is only 0.49%. The model with RBF has the best generalization performance. Then, the RBF kernel may be the best choice for the ε-SVR model.

Table 3. The CV results of the ε-SVR model training

Kernel function	Model parameters	MSE errors
LK	null	5.22%
PNK	c=3.5, gama=0.3, d=3	5.35%
RBF	c=3.5, gama=0.3, e=0.005	0.49%

3.3. Performance Metrics

For a comprehensive evaluation of ε-SVR performance of the estimation model, five performance indicators are introduced to evaluate model performance, including determination of coefficients (R^2), root mean square error (RMSE), average absolute error (MAE), average absolute percentage error (MAPE), and OCV estimation errors (Error). R^2 denotes the degree of interpretation of the input variable to the output variable. When the value is 1, the fitting effect of the model is the best. RMSE represents the sample standard deviation of the difference between the OCV estimated value and the OCV actual value. MAE, MAPE, and the error range error used to determine the OCV estimated results. The closer the values of RMSE, MAE, MAPE, and Error are to 0, the better the performance of the model. Each performance metric is defined as follows.

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(OCV_k - OCV_k^*)^2}{\sum_{k=1}^{n}(OCV_k - M_{ocv})^2} \tag{17}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(OCV_k - OCV_k^*)^2} \tag{18}$$

$$MAE = \frac{1}{n}\sum_{k=1}^{n}|OCV_k - OCV_k^*| \tag{19}$$

$$MAPE = \frac{1}{n}\sum_{k=1}^{n}\left|\frac{OCV_k - OCV_k^*}{OCV_k}\right| \tag{20}$$

$$Error = \frac{OCV_k - OCV_k^*}{OCV_k} \tag{21}$$

where OCV_k denotes the actual value, OCV_k^* denotes the estimated value, M_{ocv} is the mean OCV, and n is the number of test samples.

3.4. Performance Validation of the ε-SVR Model

To verify the performance of the ε-SVR model, the model configured with the RBF kernel is trained and tested by the training dataset and the test dataset. The test results are shown below. The test set used in the test process is different from the training set. The measured value is considered the actual battery OCV.

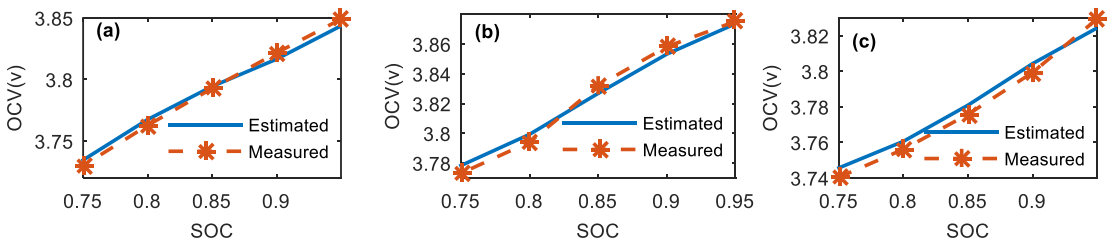


Figure 8. The test results of OCV estimation when SoH equals 0.45

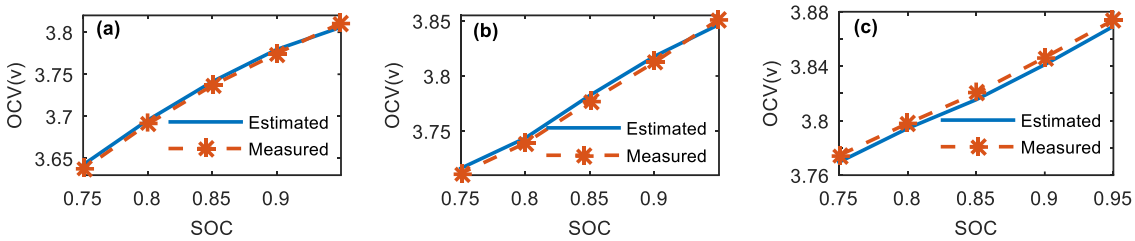


Figure 9. The test results of OCV estimation when SoH equals 0.75

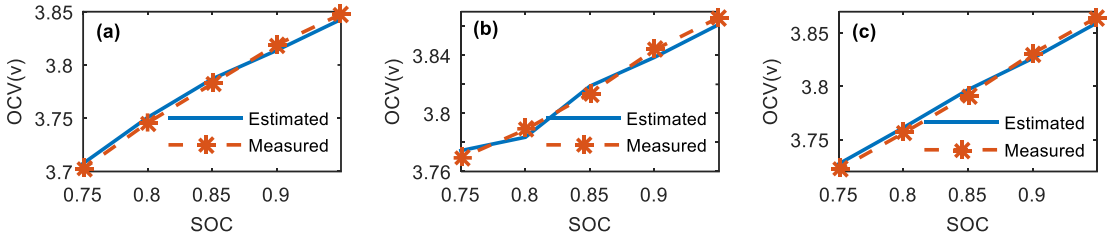


Figure 10. The test results of OCV estimation when SoH equals 0.9

Figs. 8–10 show the OCV estimation results of the trained model tested on the three test sets in the case of multiple SoH values. When SoH is equal to 0.45, the test results of randomly selecting test data three times are shown in Fig. 8 (a, b, c). According to Fig. 8, the OCV increases monotonically with battery SoC. When the SoH is equal to 0.75, the estimation results are well fitted, as shown in Fig. 9, and the OCV and SoC curves show some nonlinear characteristics. When the SoH is equal to 0.9, the estimation results are also well fitted. Through the above analysis, the ε-SVR model has achieved satisfactory OCV estimation.

The statistics of the experimental test results are shown in Table 4. The ε-SVR model obtained the worst estimation results on the No. 3 test set, and the obtained RMSE, MAE, and R^2 values were 0.5%, 0.5%, and 98.94%, respectively. However, the ε-SVR model obtained better estimation results on the No. 2 test set, and its RMSE, MAE, and R^2 values were 0.49%, 0.49%, and 99.39%, respectively. These results exceed the estimated results obtained by the ε-SVR model on the No. 3 test set. Moreover, the ε-SVR model obtained the best estimation results on the No. 1 test set, and the corresponding RMSE, MAE and R^2 values were 0.47%, 0.46% and 98.76%, respectively.

Table 4. Statistical results of OCV estimation

Kernel	Test case	RMSE	MAE	R^2 (%)	Error (%)
Radial	No.1	0.47	0.46	98.76	[-0.14 0.14]
Basis	No.2	0.49	0.49	99.39	[-0.14 0.15]
Function	No.3	0.5	0.5	98.94	[-0.14 0.14]

As shown in Table 4, the ε-SVR model obtained the best test results on the first test set, and the corresponding performance metrics reached satisfactory results. In contrast, the model obtained the worst estimation results on test set No. 3. Although the estimation results of the model on the three test sets are different, the overall performance is still excellent; in particular, the error range of battery OCV estimation is very small, which is always between -0.14% and 0.15%. Therefore, the above test results fully show that the ε-SVR model configured with RBF kernel can overcome the nonlinear relationship between feature variables and battery OCV and realize the accurate estimation of battery OCV.

3.5. The Influence of the Kernel on the ε-SVR Model

To verify the rationality of the model's kernel function selection, two other ε-SVR models with LK and PNK functions were trained and tested. The test results are shown in Table 5.

Table 5. The statistical results of OCV estimation

Kernel	Test case	RMSE	MAE	R^2 (%)	Error (%)
Linear kernel	No.1	4.86	3.88	31.19	[-2.35 1.28]
	No.2	5.5	4.62	22.87	[-2.43 2.74]
	No.3	5.29	4.47	16.6	[-2.57 2.05]
Polynomial kernel	No.1	4.95	3.94	36.42	[-2.43 0.94]
	No.2	5.65	4.71	18.86	[-2.55 2.76]
	No.3	5.46	4.68	24.3	[-2.43 1.72]

As shown in the above table, the ε-SVR model equipped with LK has the worst OCV estimation result on test set No. 3, with RMSE, MAE, and R2 values of 5.29%, 4.47%, and 16.6%, respectively. The ε-SVR model equipped with PNK has the best OCV estimation result on the first test set. Its RMSE, MAE and R2 values are 4.95%, 3.94% and 36.42%, respectively. In particular, the range of its estimation error is reduced to between -2.43% and 0.94%. From the overall test results, compared with the estimation result of the ε-SVR model equipped with LK, the estimation result of the ε-SVR model equipped with PNK has no obvious advantage. However, compared with the estimation result of the model configured with RBF, as shown in Table 4, the OCV estimation result of the model with RBF has obvious advantages.

From the above statistical results, it is clear that these two models cannot obtain ideal OCV estimation results. In contrast, the model using RBF achieves an accurate estimation of OCV. Because the RBF introduces the Gaussian kernel function, it can effectively overcome the nonlinear characteristics of the OCV. The ε-SVR model with RBF has achieved satisfactory estimation results, so it is the best choice for OCV estimation.

3.6. The Influence of Sample Features on the ε-SVR Model

To illustrate the rationality of sample feature selection, one sample feature is randomly reduced for each training sample, and then six ε-SVR models configured with the RBF kernel are trained by this training sample. The results of the experimental tests are shown in Table 6. When F19 is deleted, the test result of the model on T3 is the worst, the RMSE and MAE values reach 0.52%, the R2 value is also less than 99%, and the range of estimation error reaches [-0.14 0.15]. When F23 is deleted, the test result of the model on T1, T2, and T3 is the best, and the range of RMSE and MAE are between [0.45 0.51], and the value of R^2 is also satisfactory. In particular, range of estimation error is reduced to between -0.13 and 0.14. In Table 6, the comprehensive range of R^2 is between [98.67 99.39], and the range of error estimation error is between [-0.14 0.15].

Table 6. The statistical results of battery OCV estimation

Minus feature	Test case	RMSE	MAE	R^2	Error
	No.1	0.49	0.48	98.67	[-0.13 0.14]
F1	No.2	0.5	0.5	99.36	[-0.14 0.14]
	No.3	0.51	0.51	98.9	[-0.14 0.15]
	No.1	0.49	0.48	98.67	[-0.13 0.14]
F3	No.2	0.5	0.5	99.36	[-0.14 0.14]
	No.3	0.51	0.51	98.9	[-0.14 0.15]
	No.1	0.48	0.46	98.73	[-0.14 0.15]
F8	No.2	0.5	0.5	99.35	[-0.14 0.14]
	No.3	0.51	0.51	98.9	[-0.14 0.14]
	No.1	0.46	0.42	98.84	[-0.14 0.14]
F10	No.2	0.5	0.5	99.36	[-0.14 0.15]
	No.3	0.51	0.51	98.91	[-0.14 0.15]
	No.1	0.48	0.47	98.72	[-0.14 0.14]
F19	No.2	0.49	0.49	99.38	[-0.14 0.14]
	No.3	0.52	0.52	98.87	[-0.14 0.15]
	No.1	0.47	0.45	98.77	[-0.13 0.14]
F23	No.2	0.49	0.49	99.39	[-0.14 0.13]
	No.3	0.51	0.51	98.92	[-0.14 0.14]

Compared with the results in Table 4, the value of R^2 and error change little. This shows that deleting any of these features has little effect on R^2 and error in the model estimation results. Moreover, compared with the results in Table 4, if a feature variable is deleted in the training dataset, the impacts on R^2 and error in the model estimation results are minimal, and the impacts on RMSE and MAE are slightly larger. This is precisely because the correlation coefficient between each feature and OCV is between [0.7 0.77], as shown in Table 1. Moreover, among the test results, the result of test 1 is the best, and the result of test 3 is the worst. Therefore, no matter which features are deleted from the data sample, the accuracy

of the OCV estimation result of the ε-SVR model will be reduced accordingly. Therefore, as shown in Table 1 above, the selected features are essential for the ε-SVR model to achieve battery OCV estimation.

3.7. The Influence of Hyperparametric on the ε-SVR Model

The ε-SVR regression is very easy to use with only three parameters to tune: the cost parameter C, the gamma in the kernel function, and the epsilon in the loss function. The parameter C is the penalty coefficient, indicating the tolerance for estimation errors. A larger C indicates that the error cannot be tolerated and is easy to overfit. The smaller the value of C is, the easier the model is to under fit. If the parameter C is too large or too small, the model generalization ability will be poor. As shown in Fig. 11, when C is greater than 3.5, both RMSE and MAE tend to be minimum, while the value of R^2 reaches the maximum value of 100%.

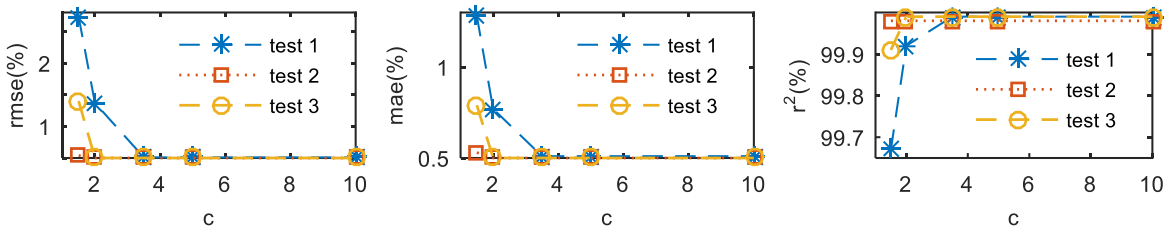


Figure 11. The relationship between parameter C and performance metrics

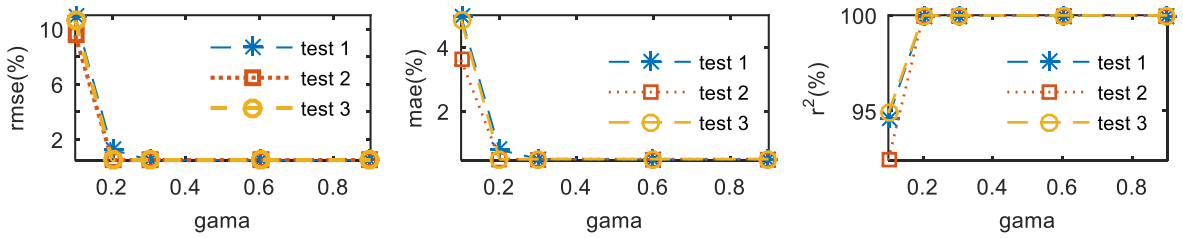


Figure 12. The relationship between parameter gama and performance metrics

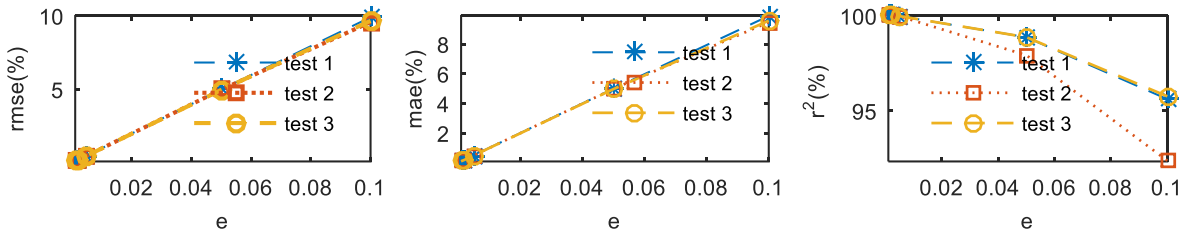


Figure 13. The relationship between parameter ε and performance metrics

The parameter gamma is a parameter in the RBF kernel. Based on this parameter, the data distribution can be implicitly determined after the data are mapped to the new feature space. Therefore, the larger the value of the parameter gamma is, the smaller the support vector of the ε-SVR model, and vice

versa. As shown in Fig. 12, when gamma is greater than 0.2, both RMSE and MAE tend to 1%, while the value of R^2 reaches 100%. Epsilon is a sensitive parameter in the loss function; when the deviation is less than epsilon, no penalty will be imposed. When the deviation is greater than epsilon, the corresponding penalty will be imposed. As shown in Fig. 13, epsilon and RMSE show a linear correlation and monotonically increase, and the same is true for epsilon and MAE. There is a piecewise linear relationship between ε and R^2. When ε is less than 0.03, the value of R^2 decreases with increasing ε, and when ε is greater than 0.03, the value of R^2 decreases rapidly. After epsilon is less than 0.01, the value of R^2 is close to the optimal value. To summarize, by analysing the impact of the three parameters on the performance metrics RMSE, MAE, and R^2, the parameter of the ε-SVR model can be easily determined.

3.8 Performance comparison

In the literature [13], based on the improved model of the second-order resistance and capacitance, the author proposes a fast OCV prediction method. Using the linear regression technique, the static OVC estimation can be completed within 20 minutes, and the error is less than 1mV. However, the second-order resistance-capacitance model does not consider the environmental temperature factor, and this factor has a great influence on the battery charging and discharging process and the use efficiency. Therefore, the OCV estimation method established using this method is severely affected by temperature. In [5], the author proposed a novel OCV-SOC model for lithium batteries based on fractional calculus. Compared with ordinary models, fractional-order modelling methods can effectively improve the accuracy of OCV estimation. However, the fractional-order modelling process is complex, and parameter identification is also difficult. Therefore, using this kind of method to establish battery OCV estimation, although it can improve the accuracy of the model, it will also increase some challenges. In [17], the author designed an H-infinity filter that can extract the battery OCV-SoC relation table in seconds. Experimental results show that the OCV estimated using this method can achieve accurate battery SoC estimation with a maximum error of 1%. Considering the effect of modelling error and parameter uncertainty on OCV estimation, an H-infinity filter is used to identify the ensemble parameter set. The observer parameter tuning process is cumbersome and the parameter adjustment process is difficult. The dependency on the model is high and it is difficult to apply to different objects. This paper proposes an SVM-based OCV estimation model that uses only voltage response curves as sample features extracted from HPPC tests. Not only can high-accuracy OCV estimates be obtained, but also battery modelling and complex parameter tuning procedures can be avoided.

4. CONCLUSIONS

In this article, we focus on model design and feature construction based on machine learning technology to establish a battery OCV estimation method. On the one hand, the ε-SVR model is good at solving classification and regression problems and has excellent performance in processing high-dimensional, nonlinear and small samples datasets. Therefore, we choose this model to set up a battery OCV estimation model. On the other hand, the battery voltage response under the mixed-pulse current test

is employed for feature construction. This non-destructive and non-invasive short-term feature acquisition method is convenience to implement in engineering applications.

Through this study, the conclusions are as follows. First, the best battery OCV estimation results can be obtained by the ε-SVR model with the RBF kernel. Therefore, considering the generalization performance of the model, the RBF kernel is the best choice for battery OCV estimation by the ε-SVR model. Second, based on the voltage relaxation behaviour of lithium-ion batteries, only 4 minutes are demanded to acquire the voltage response in the mixed-pulse current test for sample feature construction. This method used for short-term feature acquisition is fit well for rapid on-site measurements in engineering applications. Third, a new OCV estimation method for a LiB is proposed. The ε-SVR model only requires some short-term sample data, which can effectively be apply to realize rapid measurements without relying on complex mathematical calculations or cumbersome parameter adjustments. Finally, through an analysis and discussion of the experimental results, it is shown that the method has high battery OCV estimation accuracy, and the construction method of sample features is convenient, which shows that the technique has superior performance.

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.

ACKNOWLEDGMENTS

This work is supported by the Department of Education of Guangdong Province(2019GKTSCX084), the National Natural Science Foundation of China (61573153, 62006052), and the Natural Science Foundation of Guangdong Province (2021A1515011520).

References

1. Z. B. Ghassan, D. L. Rodolfo, C. Monica, G. Pasaoglu, *Renew Sust Energ Rev*, 89(2018) 292-308.
2. S. Zhang, X. Guo, X. Zhang, *J. Energy Storage*, 32(2020) 101980.
3. J. Kim, L. Krüger, J. Kowal, *J. Energy Storage*, 32(2020) 101841.
4. M. J. Esfandyari, Y. D. Hairi, M. R. Esfahanian, V. Masih-Tehrani, M. Nehzati, H. Shekoofa, *J. Energy Storage*, 24(2019) 100758, .
5. Q. Zhang, N. Cui, Y. Li, B. Duan, C. Zhang, *J. Energy Storage*, 27(2020) 100945.
6. Z. Zhou, Y. Cui, X. Kong, J. Li, Y. Zheng, *J. Energy Storage*, 32(2020) 101830.
7. A. Farmann, D. U. Sauer, *Appl Energ*, 225(2018) 1102-1122.
8. C. Campestrini, S. Kosch, A. Jossen, *J. Energy Storage*, 12(2017) 149-156.
9. Newman, Srinivasan. *Electrochem Solid-State Lett*, 9(2006) A110-4.
10. A. Barai, W. D. Widanage, J. Marco, A. McGordon, P. Jennings, *J. Power Sources,* 295(2015) 99-107.
11. S. Arora, W. Shen, A. Kapoor, *J. Power Sources* ,350(2017) 117-26.
12. W. Wladislaw, S. D. Uwe, *Appl Energ,* 111(2013) 416-427.
13. L. Pei, T. S. Wang, R. G. Lu, C. B. Zhu, *J. Power Sources* ,253(2014) 412-418.
14. K. Qian, B. H. Huang, A. H. Ran, Y. B. He, B. H. Li, F. Y. Kang, *Electrochimica Acta* ,303(2019) 183-191.
15. C. Zhang, K. Li, L. Pei, C. Zhu, *J. Power Sources,* 283(2015) 24-36.

16. H. H. Pan, Z. Q. Lu, W. L. Lin, J. Z. Li, L. Chen, *Energy,* 138(2017) 764-775.
17. R. Xiong, Q. Q. Yu, L. Y. Wang, *Appl Energ* ,207(2017) 346-353.
18. C. Yi-Hsien, W. Y. Sean, J. C. Ke, *J. Power Sources,* 196(2011) 3921-3932.
19. F. D. Zheng, Y. J. Xing, J. C. Jiang, B. X. Sun, J. H. Kim, M. Pecht, *Appl Energ* ,183(2016) 513-525.
20. A. Farmann, D. U. Sauer, *J. Power Sources,* 347(2017) 1-13.
21. J. F. Yang, W. X. Huang, B. Xia, C. Mi, *Appl Energ* ,237(2019) 682-694.
22. B. Pan, D. Dong, J. G. Wang, J. B. Nie, S. Y. Liu, Y. H. Cao, Y. Z. Jiang, Electrochimica Acta, 362(2020) 137101.
23. J. B. Goodenough, Y. Kim, *J. Power Sources* ,196(2011) 6688-6694.
24. W. Dreyer, J. Jamnik, C. Guhlke, R. Huth, J. Moskon, M. Gaberscek, *Nature Materials,* 9(2010) 448-453.
25. J. C. Jiang, Q. J. Liu, C. P. Zhang, W. Zhang, *IEEE T Ind Electron,* 61(2014) 6844-6851.
26. P. D. Weidman, D. Ahn, R. Raj, *J Power Sources*, 249(2014) 219-230.
27. N. Tosun, *Int J Adv Manuf Tech*, 28(2006) 450-455.
28. C. Cortes and V.N. Vapnik, *Mach Learn*, 20(1995) 273-297.
29. R. Kohavi, *International Joint Conference on Artificial Intelligence (IJCAI)*, 1995, Montreal
30. Matlab 2015.online: https://www.mathworks.com/
31. libsvm. [Online] https://www.csie.ntu.edu.tw/~cjlin/libsvm/