CityU Scholars

Quality Classification of Lithium Battery in Microgrid Networks Based on Smooth Localized Complex Exponential Model

Huang, Zhelin; Yang, Fangfang

Document Version:
Final Published version, also known as Publisher's PDF, Publisher's Final version or Version of Record

Publication record in CityU Scholars:
Go to record

WILEY | Hindawi

Research Article

Quality Classification of Lithium Battery in Microgrid Networks Based on Smooth Localized Complex Exponential Model

Zhelin Huang [iD][1] **and Fangfang Yang** [iD][2]

[1]*Department of Statistics, College of Economics, Shenzhen University, Shenzhen, China*
[2]*School of Data Science, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China*

Correspondence should be addressed to Fangfang Yang; fangfyang2-c@cityu.edu.hk

Accurate prediction of battery quality using early-cycle data is critical for battery, especially lithium battery in microgrid networks. To effectively predict the lifetime of lithium-ion batteries, a time series classification method is proposed that classifies batteries into high-lifetime and low-lifetime groups using features extracted from early-cycle charge-discharge data. The proposed method is based on a smooth localized complex exponential model that can extract battery features from time-frequency maps and self-adaptively select the time-frequency resolution to maximize the discrepancy of data from the two groups. A smooth localized complex exponential periodogram is then calculated to obtain the time-frequency decomposition of the whole time series data for further classification. The experimental results show that, by using battery features extracted from the first 128 charge-discharge processes, the proposed method can accurately classify batteries into high-lifetime and low-lifetime groups, with classification accuracy and specificity as high as 95.12% and 92.5%, respectively.

1. Introduction

Energy storage technology is regarded as the last kilometer of new energy development. Energy storage lithium battery in microgrid is a small power generation and distribution system composed of distributed generation, energy storage device, energy conversion device, load, and monitoring and protection device. In order to ensure the safety of power consumption load within the power supply scope of microgrid, the iron lithium battery energy storage system as an important backup of microgrid is essential. At present, the application process of power grid system mainly includes acid battery, flow battery, and iron lithium battery. Lithium-ion batteries have become the most promising solutions for applications in microgrid networks due to their high energy density and high power density.

To ensure safety, battery quality must be assessed and guaranteed within a limited time during production. Assessment of battery quality, which usually uses cycle life as an indicator, requires observation of the actual failure of sample batteries during the evaluation period, which is time-consuming. It is thus important to develop effective methods that can identify the quality of sample batteries and better monitor their health.

Several studies have examined unit-quality monitoring. For examples, Ribeiro [1] and Omitaomu et al. [2] applied SVM for machine fault detection and battery quality control, respectively. Park et al. [3] proposed a dual-feature functional SVM for fault detection of lithium-ion batteries. Peng et al. [4] designed example selection criteria and an active learning algorithm and selected a representative time series instance for battery event detection. However, traditional machine learning methods generally use the capacity degradation curve to evaluate battery quality, which is limited. The limitation of existing methods is as below. First, KNN, SVM, and other machine learning methods are often used in discrete sample classification, but less used in time series classification. Second, the classification results obtained by machine learning method are difficult to be explained by the physical state of batteries. Third, the capacity degradation

curve cannot reflect all the characteristics of the battery. It is abundantly clear that other features, such as voltage and temperature, are highly sensitive to the health of the cells. More specifically, the degree of fluctuation of those extracted features increases with cell aging. When battery cells are fresh, the degree of variation is small; however, this variation increases as battery age increases. Also, low-lifetime batteries tend to show large fluctuations and large variances at an early stage [5, 6]. Therefore, in this work, we explore the use of more features for classifying battery quality and propose a new classification method to explore the change of the degree of variation.

In this paper, a set of battery features extracted from temperature, voltage, and current information are adopted as degradation indicators. These indicators are obtained by applying signal processing methods, such as differential voltage analysis (DV) and incremental capacity analysis (IC), which are thought to be useful for evaluating battery health. For instance, Zhou et al. [7] proposed an average voltage attenuation to estimate remaining useful battery life. Weng et al. [8] used the peak position in the IC curve as a conditional indicator to track the state of health (SOH) of lithium-ion batteries in electric vehicles. Wang et al. [9] adopted two inflection points from the location interval of the DV curve as the SOH indicator. These extracted features are time series data from a series of discharging cycles. The time series data is time-correlated or time-dependent, and their vibration frequency and variance usually differ at various time periods, which make the time series classification problem challenging.

Time domain approaches and frequency domain approaches have been proposed for time series classification. The abovementioned features can be transformed in a time domain and a frequency domain, respectively. Fu et al. [10] combined traditional dynamic time wrapping with uniform scaling. However, when calculating the distance between two time series, the method assigns the same weight to each pair of observations, which may cause errors. The disadvantage of considering only the time domain information is that such information ignores the phase difference between the reference value and the test value, which may cause misclassification when shape similarity is compared. Various methods extract information from the frequency domain. Lines et al. [11] proposed a shapelet transform and constructed a decision tree to carry out the classification. Pulli [12] investigated seismic and explosion discrimination problems using the ratio of spectra. Frequency domain information can only reflect the characteristics if such characteristics remain constant throughout the time period. However, if such characteristic changes in different time periods of the whole time series, then frequency domain information cannot detect these changes.

These works have certain limitations because battery systems usually display nonstationary and nonlinear properties, which means that the degree of fluctuation of the characteristics of the battery will change at different time periods. It is abundantly clear that some features, such as temperature, voltage, and capacity, are highly sensitive to the health of the cell. Their variance is visible between cells.

In this paper, we proposed a smoothed localized complex exponential (SLEX) model, which is based on the SLEX transform for adaptive signal decomposition. The reasons why we use the SLEX transform are as follows. First, the SLEX transform corresponds to a time-frequency decomposition, which enables us to calculate the SLEX periodogram for a specific time-frequency region; thus, it can well capture the differences of the variance in different time period and as a basis for classification. Second, the SLEX transform has been successful in various data analyses [13]. However, in battery life analysis, it remains unexplored. Third, in the proposed method, we use a best tree method (see Section 2.2) to self-adaptively select the best SLEX-basis functions so that our method can maximize the discrepancy between two different groups of the time series, leading to higher prediction accuracies.

The rest of this paper is organized as follows. Some technical details of the SLEX model and the classification method based on the SLEX model are introduced in Section 2. The data and preprocessing methods are introduced in Section 3. The experimental results are reported and discussed in Section 4, and our conclusions are drawn in Section 5.

2. Battery Classification Based on SLEX Transformation

2.1. The SLEX Transform. In this section, we briefly introduce the SLEX transform and some properties of this model.

The SLEX library is a set of bases. Each base has orthogonal vectors with time support which are obtained by dividing the time series of length T in a binary manner (Figure 1). An SLEX basis vector $\psi_{S,\omega}(t)$ that has support on the discrete time block $S = [a_0 + 1, \ldots, a_1]$ and oscillates at frequency ω has the form

$$\psi_{S,\omega}(t) = \Psi_{S,+}(t)\exp\left(i2\pi\omega\frac{t}{|S|}\right) + \Psi_{S,-}(t)\exp\left(-i2\pi\omega\frac{t}{|S|}\right), \tag{1}$$

where a_0 and a_1 are start and end points of the time block S. The windows $\Psi_{S,+}(t)$ and $\Psi_{S,-}(t)$ are two particular smooth windows, with the following forms:

$$\Psi_{S,+}(t) = r^2\left(\frac{t - a_0}{\epsilon}\right)r^2\left(\frac{a_1 - t}{\epsilon}\right),$$

$$\Psi_{S,-}(t) = r^2\left(\frac{t - a_0}{\epsilon}\right)r^2\left(\frac{a_1 - t}{\epsilon}\right), \tag{2}$$

where $r(\cdot)$ is the sine rising cut-off function:

$$r(u) = \sin\left(\frac{\pi}{4}(1 + u)\right). \tag{3}$$

The SLEX library is a collection of bases, each having orthogonal vectors with time support that is obtained by segmenting the time series of length T, in a dyadic manner. The library is constructed by first specifying the finest resolution level J (smallest time block has length $T/2^J$). At resolution level j (where $j = 0, \ldots, J$), the time series is

S(0, 0)			
S(1, 0)		S(1, 1)	
S(2, 0)	S(2, 1)	S(2, 2)	S(2, 3)

FIGURE 1: The SLEX library.

divided into 2^j blocks. We denote the block b on level j to be $S(j, b)$ and $M_j = T/2^j$. The SLEX vectors on block $S(j, b)$ are allowed to oscillate at different frequencies $\omega_k = k/M_j$, where $k = -(M_j/2) + 1, \ldots, (M_j/2)$.

The SLEX transform consists of the set of coefficients that corresponds to all SLEX vectors defined in the library. The SLEX coefficients on block $S = S(j, b)$ are defined to be

$$\hat{\theta}_{S,k} = \frac{1}{\sqrt{M_j}} \sum_t X(t) \overline{\psi_{S,\omega_k}(t)}, \tag{4}$$

where M_j is the number of points on block S_j and the frequency is $\omega_k = k/M_j$ and $k = -(M_j/2) + 1, \ldots, (M_j/2)$.

The SLEX periodogram is defined as

$$\hat{\alpha}_{S,k} = \left| \hat{\theta}_{S,k} \right|^2. \tag{5}$$

For more details about the SLEX model, see [14].

In the following section, we first introduce a best-base selection method that can self-adaptively select the best SLEX base. This method will then be applied to battery data classification.

2.2. Best Base Selection.

Because the SLEX transform corresponds to a multilevel decomposition, we denote $S(j, b)$ as the block index. These blocks are of dyadic lengths (Figure 1). For example, if we rescale the entire time series into $(0, 1)$, the blocks of dyadic length will consist of $(0, 1/2)$, $(1/2, 1)$, $(0, 1/4)$, $(1/4, 1/2)$, $(1/2, 3/4)$, $(3/4, 1)$, and so on. The division of $\cup_{S(j,b)}$ determines the time domain resolution and the frequency resolution. For example, if the length of a time series is 1024 and we have a depth of three decompositions, the corresponding division of the time axis is $(0, 1/4)$, $(1/4, 1/2)$, and $(1/2, 1)$ $(\cup_{S(j,b)} = S(2, 0), S(2, 1), S(1, 1))$ and the frequency resolutions are 256, 256, and 512, respectively.

We show how to choose the best base $\cup_{S(j,b)}$ that best reflects the time-frequency decomposition characteristics of the process. Suppose we have two time series, x^1 and x^2.

For a given decomposition depth J, $2^{2^{J-1}}$ different SLEX bases exist. The selection algorithm outlined here is based on the idea of optimal pruning trees, which was first introduced by [15] in the signal-processing literature. This algorithm can automatically find the best base among the possible bases.

(1) Set the maximum depth J to which the tree is grown, which depends on the length T of the time series.

(2) For $j = 0, \ldots, J$, divide the time series into 2^j blocks: block (j, b), $b = 0, \ldots, 2^j - 1$, block (j, b) corresponding to node (j, b) of the tree.

(3) For $j = 0, \ldots, J$ and $b = 0, \ldots, (2^j - 1)$, compute $\hat{\alpha}_{1,j,b,k}$ and $\hat{\alpha}_{2,j,b,k}$ of the spectrum in block (j, b) for two time series x^1 and x^2, respectively.

(4) Compute $R_{j,b} = \text{disparity}(\alpha_{1,j,b,k}, \alpha_{2,j,b,k})$, where

$$\text{disparity}(\alpha_{1,j,b,k}, \alpha_{2,j,b,k}) = \sum_{k=1}^{M_j/2} \frac{\alpha_{1,j,b}}{\|x^1\|^2} \log \frac{\left(\alpha_{1,j,b}/\|x^1\|^2\right)}{\left(\alpha_{2,j,b}/\|x^2\|^2\right)}. \tag{6}$$

(5) For $j = J$ to $j = 0$, $b = 0$ to $2^j - 1$. If $j = J$, label block (j, b) as terminal. If $j < J$ and if $R(j, b) \geq R(j + 1, 2b) + R(j + 1, 2b + 1)$, then label the block (j, b) as terminal; otherwise, leave block (j, b) unlabeled and set $R(j, b) = R(j + 1, 2b) + R(j + 1, 2b + 1)$.

Final segmentation = set of highest labeled blocks = {block (j, b): block (j, b) is labeled and its ancestors are unlabeled}.

2.3. Classification Rules.

Given two groups of battery data with high lifetime and low lifetime, respectively, each group has $m = 1, \ldots, N_{\text{low}}$ and $n = 1, \ldots, N_{\text{high}}$ samples of time series with known group labels (low lifetime or high lifetime), respectively. Suppose we have a new sample with unknown group label, and this sample has a periodogram $\hat{\alpha}_{\text{unknown},i,k}$. We compute the SLEX periodograms from the blocks in $\bigcup S(j, b) = \cup_i S_i$ selected in the last section ($\cup_i S_i$ is a particular dyadic segmentation of the time series) and derive our criteria based on a time-frequency domain approach using the log-likelihood ratio. We introduce the following method to classify the battery with unknown group labels: for each selected time block S_i and frequency k, we denote $\alpha_{\text{low},m,i,k}$ as periodogram of the mth sample in the low quality group and $\alpha_{\text{high},n,i,k}$ as periodogram of the nth sample in the high quality group; then,

$$\alpha_{\text{low},i,k} = \frac{1}{N_{\text{low}}} \sum_{m=1}^{N_{\text{low}}} \alpha_{\text{low},m,i,k}, \tag{7}$$

$$\alpha_{\text{high},i,k} = \frac{1}{N_{\text{high}}} \sum_{n=1}^{N_{\text{high}}} \alpha_{\text{high},n,i,k}. \tag{8}$$

Denote $\alpha = \{\alpha_{i,k}\}$. Let $p_1(\alpha)$ and $p_2(\alpha)$ be the density under processes π_1 and π_2, respectively, and let $\alpha_{\text{low}} = \alpha_1$ and $\alpha_{\text{high}} = \alpha_2$ be the periodogram under these two processes. The log-likelihoods under these two densities are then, respectively,

$$l(\alpha_{\text{low}} \mid \hat{\alpha}_{\text{unknown}}) = -\sum_{\cup S_i} \sum_{k=0}^{M_i/2} \left[\log \alpha_{\text{low},i,k} + \frac{\hat{\alpha}_{\text{unknown},i,k}}{\alpha_{\text{low},i,k}} \right],$$

$$l(\alpha_{\text{high}} \mid \hat{\alpha}_{\text{unknown}}) = -\sum_{\cup S_i} \sum_{k=0}^{M_i/2} \left[\log \alpha_{\text{high},i,k} + \frac{\hat{\alpha}_{\text{unknown},i,k}}{\alpha_{\text{high},i,k}} \right]. \tag{9}$$

The classification criterion is based on the likelihood ratio, that is, we classify the time series x with unknown type into π_1 if $l(\alpha_{\text{low}}|\widehat{\alpha}_{\text{unknown}}) > l(\alpha_{\text{high}}|\widehat{\alpha}_{\text{unknown}})$; otherwise, it is classified into π_2. Let the classification statistic be

$$D_T\left(\alpha_{\text{low}}, \alpha_{\text{high}}, \widehat{\alpha}_{\text{unknown}}\right) = l\left(\alpha_{\text{low}} \mid \widehat{\alpha}_{\text{unknown}}\right) - l\left(\alpha_{\text{high}} \mid \widehat{\alpha}_{\text{unknown}}\right)$$
$$= \sum_{\cup_{S_i}} \sum_k \left[\log \frac{\alpha_{\text{high},i,k}}{\alpha_{\text{low},i,k}} + \left(\widehat{\alpha}_{\text{unknown},i,k}\right)\left(\frac{1}{\alpha_{\text{high},i,k}} - \frac{1}{\alpha_{\text{low},i,k}}\right) \right]. \tag{10}$$

A flowchart of the proposed classification method to determine the quality of batteries is given in Figure 2.

2.4. Battery Classification Based on SLEX Transformation.
We propose that the SLEX model is used to solve the battery classification problem for the following reasons. First, the SLEX model handles nonstationary time series, where the capacity degradation curves and other features such as temperature and voltage curves are nonstationary and nonlinear. Therefore, the model is suitable for processing battery data. Second, the classification method based on the SLEX model is consistent, which means that the more data used for the classification test, the greater the classification accuracy so that the accuracy of classification is guaranteed theoretically [13]. Third, the capacity variance is minimal, when battery cells are fresh, but it increases with increasing battery age [5]. It is abundantly clear that other features, such as voltage and temperature, are highly sensitive to the health of the cells. More specifically, the degree of fluctuation of those extracted features increases with cell aging. Therefore, low-quality batteries tend to show large fluctuations and large variances at an early stage. The SLEX transform corresponds to a time-frequency decomposition, so it can well capture these differences and carry out the classification.

3. Li-Ion Battery Data and Data Preprocessing

3.1. Data.
The capacity of a rechargeable battery is usually defined as the available power in ampere-hours (Ah). It is a measure of the battery's basic performance. Cycle life refers to the number of times a charge and discharge cycle is completed before a battery's nominal capacity falls below a predetermined fraction of its initial capacity. Although it is desirable to maintain the initial capacity as much as possible during battery use, the capacity may be reduced by repeated cycles. Figure 3 shows the remaining capacities of 123 samples of commercial graphite cells as the charge and discharge cycle progressed. The data came from [16], who randomly selected the battery cells and cycled them in a temperature-controlled chamber (30°C) under various fast-charging but identical discharging conditions (4°C to 2.0 V). Internal resistance, cell temperature, current, and voltage were continuously measured during the cycle. In the cycle life test, each sample was checked for a specific threshold over a specified number of cycles. The threshold level and number of cycles are usually predetermined based on industry standards (e.g., the threshold value is usually

defined as 80% of its initial capacity over 500 cycles; if the capacity is greater than 80% of its initial capacity over 500 cycles, then we define the battery as a high-lifetime battery; otherwise, it is defined as a low-lifetime battery). According to the requirements of the battery, the battery is then assigned to the high-lifetime or low-lifetime group (Figure 3).

3.2. Feature Extraction.
In this work, the following features, with reference to [17, 18], were extracted for battery quality classification:

(1) $F1$: capacity

(2) $F2$: $(dQ/dV)_{\text{max}}$ which records the y-axis of maximum values on the IC curve

(3) $F3$: $(dQ/dV)_{\text{max}}(V)$ which records the x-axis of maximum values on the IC curve

(4) $F4$: $\text{mean}(\Delta Q)$ which records the mean of difference of Q in adjacent cycles

(5) $F5$: $\text{var}(\Delta Q)$ which records the variance of difference of Q in adjacent cycles

(6) $F6$: IR is internal resistance

(7) $F7$: T_{max} is maximum temperature of the cell in each cycle

(8) $F8$: $\text{min}(\Delta Q)$ which records the minimum value of the difference of Q in adjacent cycles

(9) $F9$: T_{min} is minimum temperature of the cell in each cycle

$F2$ and $F3$ were extracted from the incremental capacity (IC) curve (peak and voltage shift, respectively). The IC describes the relationship between a capacity change and a voltage change $(\Delta Q/\Delta V)$ during a discharge process (Figures 4(e) and 4(f)). An IC curve is obtained by charging or discharging a battery under a very low current (e.g., 1/30°C) to ensure that the battery operates in a "near equilibrium" condition [19]. Although the battery has a large charge and discharge current when used in a vehicle, as shown in [20, 21], the peak on the IC curve can still be recognized, which reveals important features about battery health for normal charging and discharging data. A study of the relationship between coulombic efficiency and capacity degradation of commercial lithium-ion batteries is shown in [22]. By observing the gradual development of the peak of the IC curve throughout the life cycle, we can understand the battery's aging mechanism.
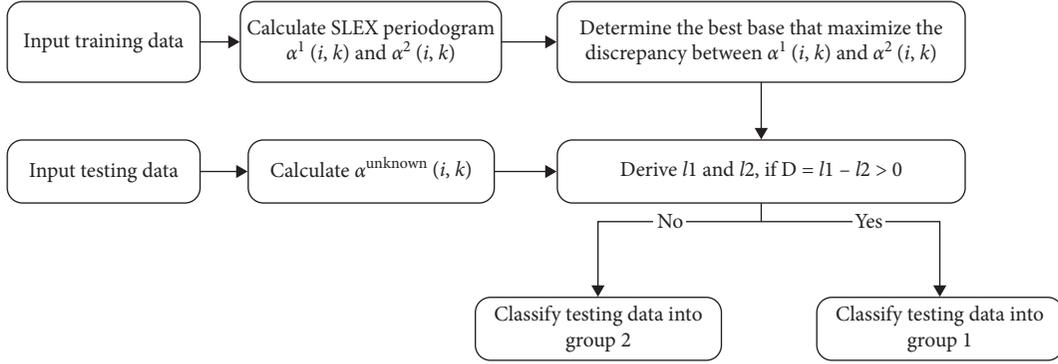
FIGURE 2: Flowchart of classification process, where training data with low lifetime are in group 1 and training data with high lifetime are in group 2.
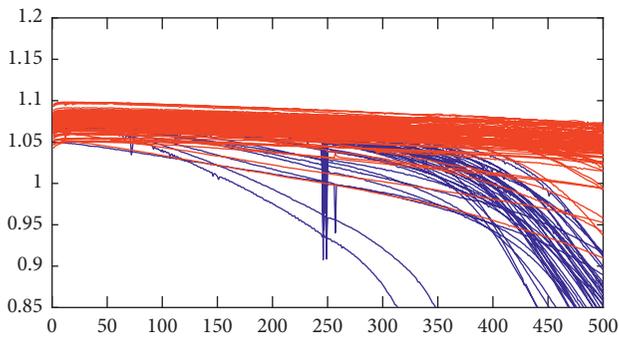

FIGURE 3: Capacities in different cycles. x-axis: cycle number and y-axis: capacity. Red lines: batteries with high lifetime. Blue lines: batteries with low lifetime.

A loss of active material, loss of lithium inventory, and an increase of IR (Figures 4(g) and 4(h)) are three main processes that cause battery degradation [23]. These factors can be easily identified by an unbalanced drop in peak intensity (y-axis of the IC curve), a decrease in the ratio of peak intensities, and a shift in peak voltage position (x-axis of the IC curve), respectively [6].

We also propose characteristics from the field of lithium-ion batteries, such as cell temperature, initial discharge capacity, and charge time (Figures 4(a) and 4(b)). To capture the electrochemical evolution of a single cell during cycling, several features were calculated from the discharge voltage curve [24]. Specifically, we considered the cycle-to-cycle evolution of $Q(V)$ and the discharge voltage curve as a function of voltage for a given cycle (Figures 4(c) and 4(d)), respectively. Because the voltage range is the same for each cycle, we consider capacity as a function of voltage as the basis for the comparison period.

3.3. Training and Testing Data. There were a total of 83 batteries with high lifetimes and 40 batteries with low lifetimes. We randomly divided these batteries into five folds. The first four folds had 16 batteries with high lifetimes and 8 batteries with low lifetimes, whereas the fifth fold had 19 batteries with high lifetimes and 8 batteries with low lifetimes. Each time, we retained one fold of the time series

for classification and used the remaining time series for training a classifier. For each holdout time series, we first selected a basis; we then calculated the SLEX periodograms at the blocks in this basis; finally, we constructed the classifier and assigned the holdout time series. This procedure can be repeated for each fold of the holdout time series. For each feature we proposed, the classification process can be carried out by the following three steps:

1. Divide the data into roughly five equal parts.
(2) For each $i = 1, 2, 3, 4, 5$, train the classifier with other 4 parts and compute the number of TP, FP, FN, and TN (see Table 1) of the ith part.
(3) Sum the number of TP and TN in 5 parts. Calculate the accuracy, sensitivity, and specificity.

The proposed classification method based on the SLEX transform requires that the length of the input data to be a power of 2. Therefore, the experiments were conducted using battery data from the first 128 cycles and the first 256 cycles. The accuracy, sensitivity, and specificity were the three metrics used to evaluate the classification results. See Table 1, for more details.

1. Accuracy = (TP + TN)/(TP + TN + FP + FN)
(2) Sensitivity = TP/(TP + FN)
(3) Specificity = TN/(TN + FP)

"Accuracy" refers to the accuracy of the entire classification. "Sensitivity" refers to the proportion of samples that are actually positive and judged to be positive, whereas "specificity" refers to the proportion of samples that are actually negative and are classified to be negative. In this paper, we mainly focus on specificity and accuracy because these two indicators reflect the recognition accuracy of low-lifetime batteries and the overall recognition accuracy.

4. Experimental Results

We applied the proposed SLEX method to classify a battery time series as either high lifetime or low lifetime.

4.1. Univariate Case. Table 2 shows the classification result using data from the first 128 cycles. Using training data from
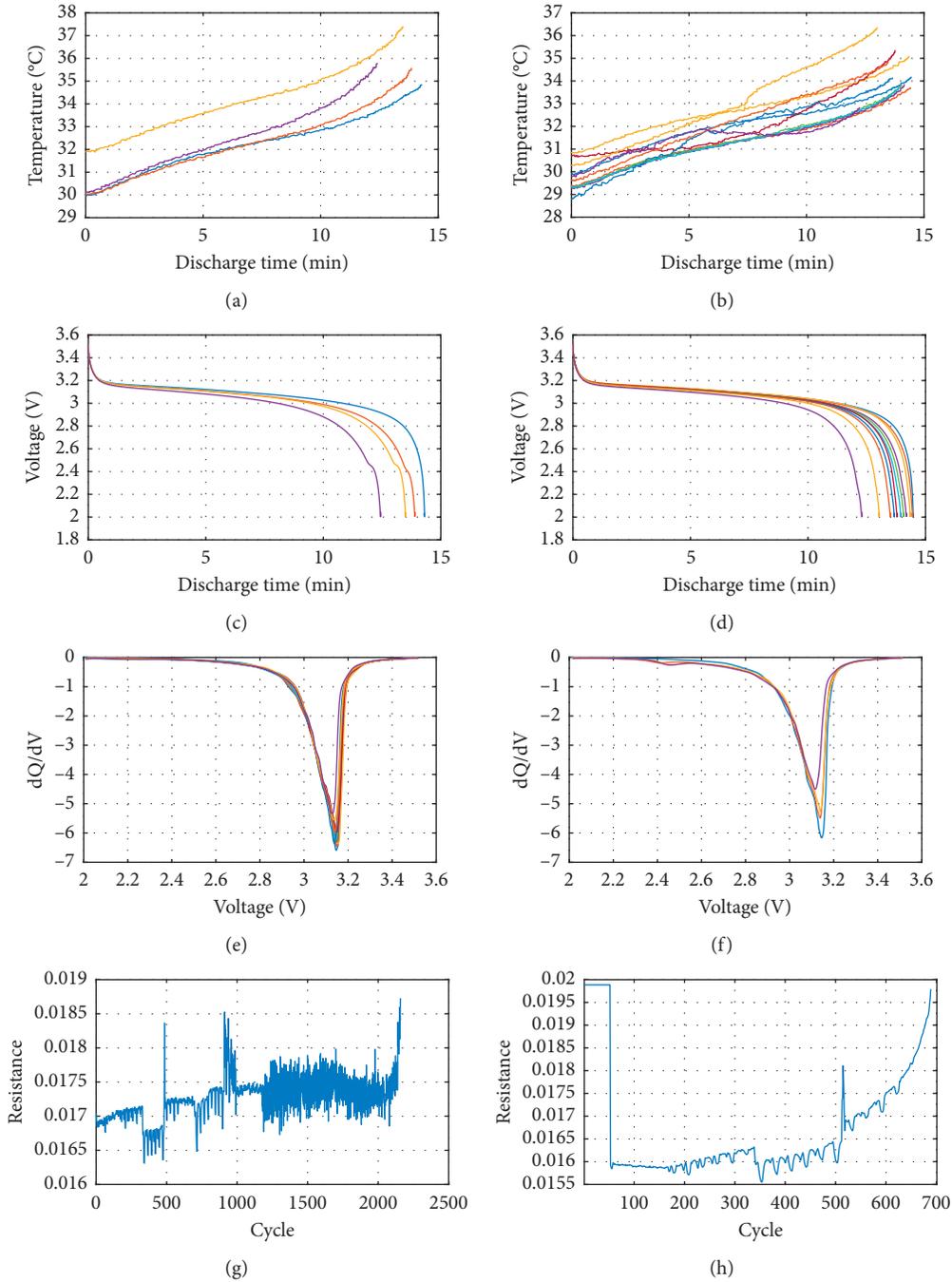
FIGURE 4: (a) Discharge temperature for cell with low lifetime, every 200 cycles; (b) discharge temperature for cell with high lifetime, every 200 cycles; (c) discharge voltage curves for cell with low lifetime, every 200 cycles; (d) discharge voltage curves for cell with high lifetime, every 200 cycles; (e) IC curve for cell with low lifetime, every 200 cycles; (f) IC curve for cell with high lifetime, every 200 cycles; (g) IR curve for cell with low lifetime; (h) IR curve for cell with high lifetime.

TABLE 1: Metrics to evaluate classification results.

	Actual positive	Actual negative
Predicted positive	True positive (TP)	False positive (FP)
Predicted negative	False negative (FN)	True negative (TN)

the first 128 cycles, the accuracy of $F1$ is 50.41%. The sensitivity and specificity of $F1$ are 53.01% and 45%, respectively. Here, it is not easy to distinguish the battery quality from the capacity because capacity differs little during the early cycles. Other features we extracted, such as $F4$, $F5$, $F7$, and $F8$, performed much better than the capacity, with classification accuracies exceeding 90%.

We also considered the classification effect using data from the first 256 cycles (Table 3). Using training data from the first 256 cycles, the specificity was much higher, but still insufficient. The classification sensitivity was 96.39% and the

TABLE 2: Classification results using data from first 128 cycles.

Features	Sensitivity (%)	Specificity (%)	Accuracy (%)
$F1$	53.01	45.00	50.41
$F2$	79.52	70.00	76.42
$F3$	65.06	65.00	65.04
$F4$	96.39	85.00	92.68
$F5$	98.80	85.00	94.31
$F6$	80.72	77.50	79.67
$F7$	93.98	92.50	93.50
$F8$	96.39	87.50	93.50
$F9$	93.98	92.50	93.50

TABLE 3: Classification results using data from first 256 cycles.

Features	Sensitivity (%)	Specificity (%)	Accuracy (%)
$F1$	96.39	25.00	73.17
$F2$	84.34	50.00	73.17
$F3$	86.75	80.00	84.55
$F4$	87.95	95.00	90.24
$F5$	78.31	90.00	82.11
$F6$	73.49	80.00	75.61
$F7$	97.59	92.50	95.93
$F8$	89.16	90.00	89.43
$F9$	97.59	92.50	95.93

specificity is 25%, which means that high-lifetime batteries could be accurately distinguished, but low-lifetime batteries could not be distinguished correctly. We therefore considered using other features for the classification. The experimental results using data from the first 128 cycles and the first 256 cycles, plotted in Figures 5 and 6, respectively, showed that increasing the amount of data improved the classification specificity of $F3$–$F8$ to a certain extent, but the sensitivity of $F4$, $F5$, $F6$, and $F8$ decreased. Because our experimental data had more high-quality batteries, the overall accuracy rate decreased slightly. Overall, increasing the amount of data only improved the classification accuracy (specificity) of low-quality batteries, whereas the classification of high-quality batteries did not improve. Also, the cost due to longer test times often includes not only production delay costs but also opportunity costs. If we can judge the quality of a battery in fewer cycles, the maintenance and replacement of the battery will be timely. If we use the first 256 cycles of data to make predictions, although the accuracy rate is improved, the overall improvement is limited, and the life expectancy of low-life batteries is less than half at this time. Therefore, it is preferable to use the first 128 cycles of data to make predictions.As shown in Figure 4, the $F3$, $F4$, $F6$, and $F8$ features achieved higher classification accuracy when using the first 128 cycles' data. Features derived from early cycles (such as discharge voltage and temperature) had great predictive performance, even before capacity decay began. We therefore studied the degenerative patterns that did not immediately cause capacity decay, but still showed up in other characteristic curves. Of all the features, the highest classification accuracy was obtained by using the maximum temperature ($F8$) during discharging in the first 128 or 256 cycles, as shown in Figures 5 and 6, respectively. The reason is that, for high-lifetime batteries, the oscillation frequency of the average temperature is lower when discharging, whereas for low-lifetime batteries, the frequency is much higher. Figure 7 shows the values of the different features of the battery in the first 256 cycles. For batteries with low lifetimes, features 4, 6, and 8 tended to fluctuate greatly, and these differences were captured by the proposed method because we compared the time-frequency decomposition maps. For batteries with low lifetimes, feature 3 showed a downward trend, whereas batteries with high lifetimes were relatively stable. This difference was also captured by the proposed model because the SLEX transform was carried out with a window Fourier transform.
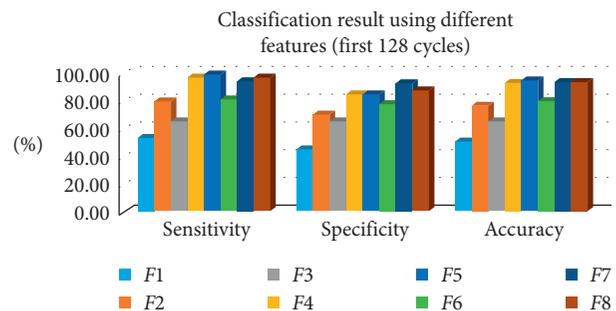

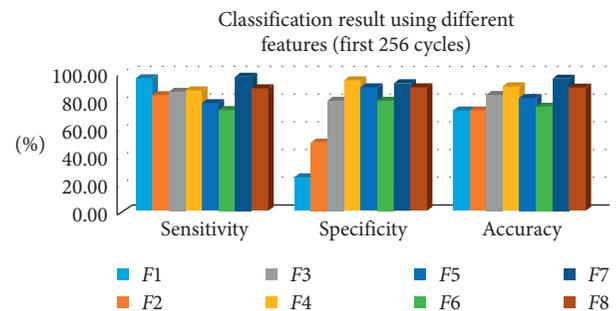FIGURE 5: Classification result with data from first 128 cycles.

FIGURE 6: Classification result with data from first 256 cycles.

Therefore, these four indicators performed well in the proposed classification method. These properties were displayed early in the cycle, before the onset of capacity fade, as shown in Figure 7. Therefore, these features presented much greater classification accuracies than the capacity.

4.2. Comparative Study. In this section, the classification accuracy of the proposed method is compared with that of methods such as KNN and SVM because these methods are classic methods for dealing with classification problems, and these two methods have been widely used in the battery field, as described in Section 1. The training and testing procedures (five folds crossvalidation) were the same as those used in the SLEX method, described in Section 3.3.

Considering the diversity of features, classifications based on $F1$, $F4$, and $F8$ were compared and the results are presented in Table 4. For $F1$ (capacity), the classification
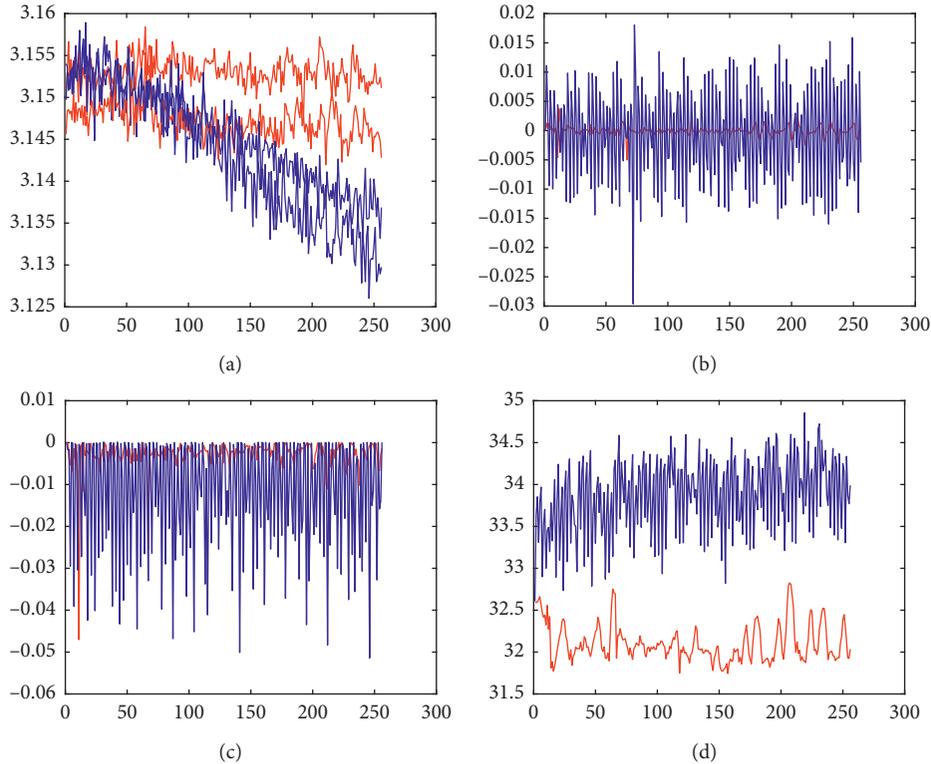
FIGURE 7: (a) $F3$; (b) $F4$; (c) $F6$; (d) $F8$. High lifetime (red line) and low lifetime (blue line).

accuracy using SLEX was 50.41%, which was not as good as the classification accuracies of KNN and SVM. However, the specificity was 45%, which was greater than that of KNN and SVM. Overall, none of these methods performed well on this feature. For F4 and F8, the SLEX methods achieved a classification accuracy above 90%, which was better than the classification accuracies of KNN and SVM. For low-lifetime batteries, the accuracy of the SLEX method was much higher than that of the other two methods, which is reflected in the value of specificity. Because the high-lifetime batteries samples in datasets are about twice as high as low-lifetime batteries (83 vs. 40), KNN and SVM were more likely to classify unknown batteries as high-lifetime batteries, resulting in lower specificity values for both methods. With the SLEX method, we took the average of the calculated periodograms from both groups according to equations (7) and (8). Therefore, the influence of sample size on the classification problem is avoided.

From the comparative analysis, we reached the following conclusions. First, we learned from the results that little difference in capacity was seen during the early cycle of a battery, so it was difficult to distinguish battery quality from this function, as shown in Table 4. Second, from Table 4, the specificities of KNN and SVM were low because, during the experiment, many low-lifetime batteries were misclassified as high-lifetime types, whereas the SLEX method achieved much better results. Third, because the SLEX method could discern the variations of the nonstationary time series in the variance by spectrum analysis, better classification effect was achieved.

TABLE 4: Classification performance compared with KNN and SVM.

Feature	Metrics	KNN (%)	SVM (%)	SLEX (%)
$F1$	Sensitivity	90.36	87.95	53.01
	Specificity	22.5	12.5	45
	Accuracy	68.29	63.41	50.41
$F4$	Sensitivity	93.98	84.34	98.80
	Specificity	77.5	75	85
	Accuracy	88.62	81.30	94.31
$F8$	Sensitivity	85.54	98.80	96.39
	Specificity	65	22.5	87.5
	Accuracy	78.86	73.98	93.5

4.3. Multivariate Case. The SLEX method can only be applied to univariate time series. To take multiple features into consideration at the same time, we proposed the following rules to further improve the classification accuracy. As discussed in Section 4.1, we only considered using data from the first 128 cycles. First, we imposed the univariate classification method on each feature, including features 2, 4, 5, 6, 7, 8, and 9 in Table 2, because the classification accuracy using these features is the highest among all features (taking into account the diversity of features, we selected seven dissimilar features). Then, if four or more features classify the battery as high lifetime, the final decision was made to classify the battery into the high-lifetime class; otherwise, the battery was considered to be in the low-lifetime class. The classification results using all seven features together were shown in Table 5 and Figure 8. By using features such as

TABLE 5: Classification results using data from first 128 cycles.

Features	Sensitivity (%)	Specificity (%)	Accuracy (%)
$F2$, $F4$, $F5$, $F6$, $F7$, $F8$, and $F9$	96.39	92.50	95.12

FIGURE 8: Classification results using data from first 128 cycles.

features that we extracted from each cycle, the sensitivity was 96.39%, which was higher than other features except for F5. In this study, we mainly focused on the values of specificity and accuracy because, in actual application, the accurate identification of low-lifetime batteries is more practical. The classification accuracy improved to 95.12% and the specificity was 92.5%, values that were the best among all results obtained from the first 128 cycles. Therefore, by combining these features, we obtain better classification results overall.

5. Conclusions

In this paper, a classification method based on the SLEX model is proposed to process battery capacity data and monitor battery quality at early stage. Our proposed model aims to classify batteries with high lifetimes and with low lifetimes using only data from the first 128 charge-discharge cycles. The proposed method constructs a classifier that can self-adaptively maximize the discrepancy between two groups of data, which makes it easier to classify data with unknown labels. In Section 4.2, the experimental results showed that the proposed method achieved higher classification accuracy than commonly used method such as KNN and SVM. In addition, the experimental results show that a classification accuracy larger than 90% is achieved by using some features. Moreover, the proposed method performed much better on features extracted from temperature, voltage, and IC curves than on capacity. Among these features, batteries with high lifetime and low lifetime can exhibit different properties early in the charge-discharge cycle, which can be captured by our proposed classification model, as discussed in Section 4.1. Therefore, our proposed features

and classification model can match perfectly and get accurate classification results. Because the classification model is only applicable to one-dimensional time series, a voting model based on each factor is proposed in practical application to extend the classification model to the case of multidimensional time series problems. The experimental results show a classification accuracy and specificity of 95.12% and 92.5%, respectively, by using effective features we extracted from the battery discharging process. In practice, assessments using early-cycle battery data (the first 128 cycles) will bring new opportunities for battery optimization and production.

Data Availability

The data used to support the findings of the study are given in [16] for a reference. Supplementary information is available for the previously reported studies at https://doi.org/10.1038/s41560-019-0356-8.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] B. Ribeiro, "Support vector machines for quality monitoring in a plastic injection molding process," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 35, no. 3, pp. 401–410, 2005.

[2] O. A. Omitaomu, M. K. Jeong, A. B. Badiru, and J. W. Hines, "Online support vector regression approach for the monitoring of motor shaft misalignment and feedwater flow rate," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 5, pp. 962–970, 2007.

[3] J. I. Park, S. H. Baek, M. K. Jeong, and S. J. Bae, "Dual features functional support vector machines for fault detection of rechargeable batteries," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 39, no. 4, pp. 480–485, 2009.

[4] F. Peng, X. Zhou, H. Liu, H. Tan, Q. Luo, and J. Hu, "A time series classification method for battery event detection," in *Proceedings of the 2017 IEEE 23rd International Conference on Parallel and Distributed Systems (ICPADS)*, pp. 17–24, IEEE, Shenzhen, China, December 2017.

[5] X. Hu, J. Jiang, D. Cao, and B. Egardt, "Battery health prognosis for electric vehicles using sample entropy and sparse Bayesian predictive modeling," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 4, pp. 2645–2656, 2015.

[6] M. Dubarry and B. Y. Liaw, "Identify capacity fading mechanism in a commercial LiFePO$_4$ cell," *Journal of Power Sources*, vol. 194, no. 1, pp. 541–549, 2009.

[7] Y. Zhou, M. Huang, Y. Chen, and Y. Tao, "A novel health indicator for on-line lithium-ion batteries remaining useful life prediction," *Journal of Power Sources*, vol. 321, pp. 1–10, 2016.

[8] C. Weng, X. Feng, J. Sun, and H. Peng, "State-of-health monitoring of lithium-ion battery modules and packs via incremental capacity peak tracking," *Applied Energy*, vol. 180, pp. 360–368, 2016.

[9] L. Wang, C. Pan, L. Liu, Y. Cheng, and X. Zhao, "On-board state of health estimation of lifepo4 battery pack through

differential voltage analysis," *Applied Energy*, vol. 168, pp. 465–472, 2016.

[10] A. W.-C. Fu, E. Keogh, L. Y. H. Lau, C. A. Ratanamahatana, and R. C.-W. Wong, "Scaling and time warping in time series querying," *The VLDB Journal*, vol. 17, no. 4, pp. 899–921, 2008.

[11] J. Lines, L. M. Davis, J. Hills, and A. Bagnall, "A shapelet transform for time series classification," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 289–297, ACM, Beijing, China, August 2012.

[12] J. J. Pulli, "Extracting and processing signal parameters for regional seismic event identification," in *Monitoring a Comprehensive Test Ban Treaty*, pp. 743–754, Springer, Berlin, Germany, 1996.

[13] H.-Y. Huang, H. Ombao, and D. S. Stoffer, "Discrimination and classification of nonstationary time series using the SLEX model," *Journal of the American Statistical Association*, vol. 99, no. 467, pp. 763–774, 2004.

[14] H. Ombao, J. Raz, R. Von Sachs, and W. Guo, "The SLEX model of a non-stationary random process," *Annals of the Institute of Statistical Mathematics*, vol. 54, no. 1, pp. 171–200, 2002.

[15] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 713–718, 1992.

[16] K. A. Severson, P. M. Attia, N. Jin et al., "Data-driven prediction of battery cycle life before capacity degradation," *Nature Energy*, vol. 4, no. 5, p. 383, 2019.

[17] F. Yang, D. Wang, F. Xu, Z. Huang, and K.-L. Tsui, "Lifespan prediction of lithium-ion batteries based on various extracted features and gradient boosting regression tree model," *Journal of Power Sources*, vol. 476, p. 228654, 2020.

[18] F. Xu, F. Yang, Z. Fei, Z. Huang, and K.-L. Tsui, "Life prediction of lithium-ion batteries based on stacked denoising autoencoders," *Reliability Engineering & System Safety*, vol. 208, Article ID 107396, 2020.

[19] M. Dubarry, C. Truchot, and B. Y. Liaw, "Synthesize battery degradation modes via a diagnostic and prognostic model," *Journal of Power Sources*, vol. 219, pp. 204–216, 2012.

[20] X. Han, M. Ouyang, L. Lu, J. Li, Y. Zheng, and Z. Li, "A comparative study of commercial lithium ion battery cycle life in electrical vehicle: aging mechanism identification," *Journal of Power Sources*, vol. 251, pp. 38–54, 2014.

[21] C. Weng, J. Sun, and H. Peng, "A unified open-circuit-voltage model of lithium-ion batteries for state-of-charge estimation and state-of-health monitoring," *Journal of Power Sources*, vol. 258, pp. 228–237, 2014.

[22] F. Yang, D. Wang, Y. Zhao, K.-L. Tsui, and S. J. Bae, "A study of the relationship between coulombic efficiency and capacity degradation of commercial lithium-ion batteries," *Energy*, vol. 145, pp. 486–495, 2018.

[23] M. Kassem, J. Bernard, R. Revel, S. Pélissier, F. Duclaud, and C. Delacourt, "Calendar aging of a graphite/lifepo4 cell," *Journal of Power Sources*, vol. 208, pp. 296–305, 2012.

[24] Z. Huang, F. Yang, F. Xu, X. Song, and K. L. Tsui, "Convolutional gated recurrent unit-recurrent neural network for state-of-charge estimation of lithium-ion batteries," *IEEE Access*, vol. 7, pp. 93139–93149, 2019.