

## Research Article

# A Novel Method for Lithium-Ion Battery Fault Diagnosis of Electric Vehicle Based on Real-Time Voltage

Fang Li , Yongjun Min , and Ying Zhang 

College of Automobile and Traffic Engineering, Nanjing Forestry University, Nanjing 210037, China

Correspondence should be addressed to Yongjun Min; yjmin@njfu.edu.cn

Received 2 April 2022; Revised 8 May 2022; Accepted 11 May 2022; Published 25 May 2022

Academic Editor: Jian Su

Copyright © 2022 Fang Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The cell faults of lithium-ion batteries will lead to the atypical deterioration of battery performance and even thermal runaway. In this paper, a novel fault diagnosis method for lithium-ion batteries of electric vehicles based on real-time voltage is proposed. Firstly, the voltage distribution of battery cells is confirmed in electric vehicles, and the reasons are analyzed. Furthermore, kurtosis is utilized to discover cell faults for the first time. After the kurtosis-based strategy alarm, the faulty cells in the battery pack are identified through multidimensional scaling and density-based spatial clustering of applications with noise. This method reduces the computational load of the data platform due to the characteristics of the sequential structure. Finally, the strategies to quantify the level of faulty cells and evaluate the safety of electric vehicles are presented. Through the real-time data collected by electric vehicles, it has been proven that this method can warn and locate faulty cells earlier than the original system method and has better robustness than other unsupervised fault diagnosis methods.

## 1. Introduction

To alleviate the energy crisis and deteriorating environmental pollution, lithium-ion batteries are widely used in electric vehicles (EVs) because of their long cycle life, cleanliness, high energy density, and high-power density [1, 2]. EVs will be the development trend of future automobiles and the focus of competition in the global automobile industry. Nowadays, China regards the EV industry as one of the key strategic emerging industries. As the core component of an EV, lithium-ion batteries are assembled from many cells in series or parallel to provide drivers with sufficient range and power performance [3]. Therefore, the state of the battery directly determines the overall performance of an EV. Potential battery cell failure will lead to the decline of the comprehensive performance of EVs and even battery short circuits. In extreme cases, it will lead to safety accidents such as thermal runaways [4, 5]. In recent years, EV safety accidents occur frequently, which hinders the development of the EV industry [6]. Lithium-ion battery cell voltage is the most intuitive and effective dynamic information in the operation of EVs. Early and precise detection of voltage faults is helpful to take measures

to avoid safety accidents [7]. Early warning and isolation of battery failure units based on real-time battery parameters are of great importance to improve the safety of EVs.

To enhance the reliability and safety of lithium-ion batteries, many scholars have proposed different methods for lithium-ion battery fault diagnosis. Current fault diagnosis methods can be divided into three categories: experience-based methods, model-based methods, and data-driven methods [5, 8, 9].

The experience-based method is based on the existing prior knowledge, using logical analysis and reasoning the relationship between events to achieve battery fault diagnosis. It can be divided into the expert system [10], fuzzy logic [11], and graph theory [12]. Experience-based methods have no learning ability, resulting in limited generalization ability, which makes this kind of method less applied in lithium-ion battery fault diagnosis. The model-based method establishes the mathematical or chemical model of lithium-ion batteries. Residual signals are obtained to detect and identify faults by comparing measurable signals from the model outputs [13]. Xiong et al. [14] estimated the state of charge (SOC) based on unscented Kalman filter (UKF) and

recursive least squares (RLS) to locate voltage sensor fault location by the residuals. Wei et al. [15] presented a second-order equivalent circuit model (ECM) and strong extended tracking Kalman filter (ST-EKF) to estimate terminal voltage online for battery fault diagnosis. Song et al. [16] designed a fractional-order Luenberger observer to accomplish the task of fault detection. Dey et al. [17] utilized a partial differential equation (PDE) model-based diagnosis scheme for thermal faults in lithium-ion batteries. Kong et al. [18] developed an electrochemical model for lithium-ion batteries to detect early internal short circuit cells. The advantage of the model-based method is that the model mechanism is clear and easy to modify. However, the model-based method is difficult to be applied in practice because of many parameters or complex equation calculations, and sensor errors with many parameters can easily affect their performance.

With the development of machine learning and artificial intelligence, as well as the advancement of computer software and hardware. In recent years, data-driven methods have received a lot of attention in battery fault diagnosis. The data-driven method can be regarded as a “black box,” which overcomes the problems of complicated modeling and parameter identifications [8, 19]. Jiang et al. [11] established the early warning model for charging safety of EVs by detecting the cell voltage based on the back propagation neural network (BPNN). Yao et al. [20] decomposed and reconstructed the voltage data through the reconstructing discrete wavelet transform (RDWT) and then inputs it into the general regression neural networks (GRNN) to divide the fault into four levels. Li et al. [21] located potential faulty cells by combining ECM and long short-term memory (LSTM) neural networks. Jiang et al. [19] proposed a support vector machine (SVM) for the study of temperature rise caused by short circuit failure of batteries. Yang et al. [22] applied random forests (RF) to the research of external short circuit fault diagnosis. The above methods require a large number of training sets to have good results, resulting in excessive memory consumption and unsuitable for new vehicles. It is used in lithium-ion battery fault diagnosis and detection by converting basic information into characteristic information, such as entropy theories. Duan et al. [23] designed a method for evaluating the inconsistency of cells based on Shannon entropy. Liu et al. [24] evaluated each cell based on Shannon entropy to determine the potential faulty cells. The Shannon entropy theory has also been applied in the detection of battery pack connection faults [23]. The improved Shannon entropy [7], sample entropy [25], and multiscale entropy [6] have also been utilized for battery fault diagnosis. Some improved entropy theories have an excessive number of iterations, which increases the running time of the method. The method based on statistical knowledge and other methods has great prospects in practical fault diagnosis. Xia et al. [26] proposed a method to detect internal short circuit fault by calculating the correlation coefficient between cell voltages. Zhao et al. [4] presented a 3  $\sigma$  multilevel screening strategy and neural network to screen abnormal cells and used a clustering algorithm to verify the effectiveness of its screening strategy. Sun et al. [27] used

a correlation coefficient and  $K$ -means clustering method to diagnose faulty cells. Wang et al. [28] established a data-driven abnormal cell capacity diagnosis model based on statistics and a tree-based model. She et al. [29] developed a battery state of health SOH estimation model synthesizing from battery state of health based on incremental capacity analysis (ICA) and considering the inconsistency of cells. Chen et al. [30] developed the method of local outlier factor (LOF) to detect the voltage fault of the battery cells. Some of the above studies were carried out in the laboratory. Zhang et al. [31] optimized the multiobjective design of the hybrid energy storage system for EVs to extend the battery life and reduce the failure rate. However, in practice, the battery faces a more complex environment due to the impact of the environment and the different driving modes of drivers.

In this paper, the data obtained are not for the charging and discharging of lithium-ion batteries in the laboratory. All of the data collected comes from the real world to ensure a more practical fault diagnosis model. Based on mileage, the cell voltage distribution during charging and discharging of EV is analyzed. Firstly, kurtosis is used as an early warning indicator for faulty cells for the first time. Secondly, the method combines multidimensional scaling (MDS) with density-based spatial clustering of applications with noise (DBSCAN) to diagnose faulty cells. The subsequent calculation is performed only after the alarm of the kurtosis-based detection method, which reduces the computing load of the platform. Furthermore, the fault type of the abnormal cell is determined and quantified by using statistical knowledge, and the voltage consistency of the whole vehicle is quantified. Finally, the validity of the method is verified by the real-time data collected from EVs and compared with other unsupervised methods.

The rest of the paper is organized as follows. Section 2 describes the data acquisition and analysis. Section 3 introduces the methodology for voltage fault diagnosis. Section 4 discusses the diagnostic results in detail. Section 5 concludes this paper.

## 2. Data Acquisition and Analysis

*2.1. The Data Source of EVs.* The EV has the characteristics of informatization and networking. Therefore, China has established data centers based on big data technology. For the data acquisition specification of EVs, China published the technical specifications of remote service and management systems for EVs in 2016. The format, range, and frequency of collected data are also stipulated in this specification. In current, the EV big data centers have formed three levels of structure: Enterprise Service and Monitoring Center (ESMC), Local Service and Management Center (LSMC), and National Service and Management Center (NSMC). The structure of the EV data center is shown in Figure 1. During the operation of an EV, the vehicle terminal transmits the data collected by the sensor to the data center through the general packet radio service (GPRS) wireless network according to the specification. All data is collected by the ESMC for the first time and reported to other centers following the communication protocol.

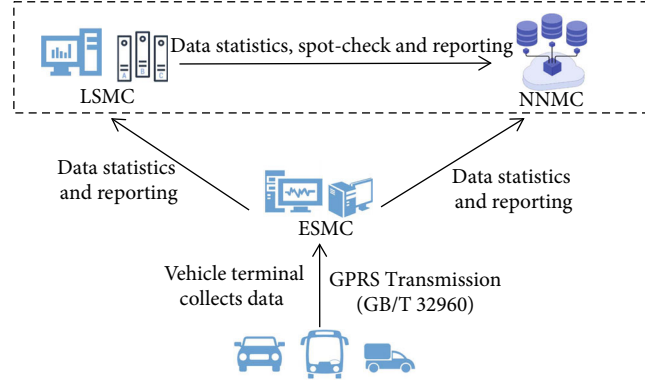


FIGURE 1: Three levels structure of electric vehicle data centers.

TABLE 1: Acquisition systems and contents.

System of EV	Acquisition content
Battery system	Battery voltage (V), battery current (A), insulation resistance (k $\Omega$ ), SOC (%), cell voltage (V), and probe temperature ( $^{\circ}$ C).
Vehicle control system	Accumulated mileage (km), vehicle status (start/stop), gear, speed (km/h), accelerator pedal travel (%), brake pedal travel (%), DC-DC status (work/disconnect), and location information (GPS).
Motor drive system	Motor speed (r/min), motor torque (N·m), motor temperature ( $^{\circ}$ C), motor controller voltage (V), and motor controller current (A).

The data platform architecture is mainly based on the Linux system and Java programming language and is built with the Hadoop system. By April 2022, the NSMC has connected 8 million new energy vehicles. Simple data statistics cannot meet the current data analysis needs of the new energy automotive industry. The data mining and law exploration behind the data deserve more attention, such as EVs safety warnings and vehicle technology subsection analysis. In order to solve the above problems, it is necessary to establish relevant models with the powerful power of big data platform or targeted cloud platform in close combination with the actual operation of new energy vehicles, so as to achieve the effect of analyzing the operation of new energy vehicles and early warning [32].

**2.2. Data Content and Cleaning.** To ensure the acquisition of real and sufficient data, this paper collects the data of 10 electric buses running for one year from the ESMC of an electric vehicle enterprise in Suzhou, China. The main content of the collection is that each system during vehicle driving includes the battery system, motor drive system, and vehicle control system as shown in Table 1 [24].

In this paper, the cell voltages of EVs are extracted for analysis. The voltage acquisition accuracy is 0.01 V, and the format obtained from the ESMC is a matrix. The row vector of the matrix is the voltage data of all cells at a certain moment, and each column vector is the voltage data of each cell. The acquisition frequency of each collected data is 0.1 Hz.

Due to the influence of buildings and the environment during driving, the data may be lost in the steps of data acquisition, transmission, and decoding, resulting in null values. Sensor failure also causes the collected data to be obviously abnormal or null, so it is necessary to clean the data.

The data cleaning strategies of this paper are as follows: The extreme abnormal data is removed through the box plots, such as 65,535 V caused by signal loss. Box plots are a common method for handling outliers in data. The feature of box plots is that there are no strict requirements on the distribution of data processed during the use of box plots. The box plot eliminates outliers as shown in Figure 2.

In Figure 2, the  $Q_1$  and  $Q_3$  are the upper and lower quartiles of the data, respectively. The interquartile range (IQR) is the length of box plot, which is the blue square in Figure 2.

The calculation formula of each symbol is as follows:

$$\text{IQR} = Q_3 - Q_1,$$

$$f_{\text{low}} = Q_1 - k \times 1.5\text{IQR}, \quad (1)$$

$$f_{\text{up}} = Q_3 + k \times 1.5\text{IQR}.$$

To prevent the fault values from being eliminated, the variable  $k$  is used to expand the  $[f_{\text{low}}, f_{\text{up}}]$  and the  $k = 3$ . The interval between the normal values is  $[f_{\text{low}}, f_{\text{up}}]$ . The range of outliers is  $(-\infty, f_{\text{low}}) \cup (f_{\text{up}}, +\infty)$ .

The missing data with more than 3 consecutive rows missing is discarded, and the remaining missing data is supplemented by the previous value.

**2.3. Voltage Distribution Analysis.** For some battery cell fault researchers, the cell voltage distribution is regarded as a normal distribution in some literature, and the Z-score [33] or  $3\sigma$  [4] method is proposed to diagnose voltage fault. However, the conclusion that the cell voltages conform to the normal distribution is not verified. A series capacity degradation

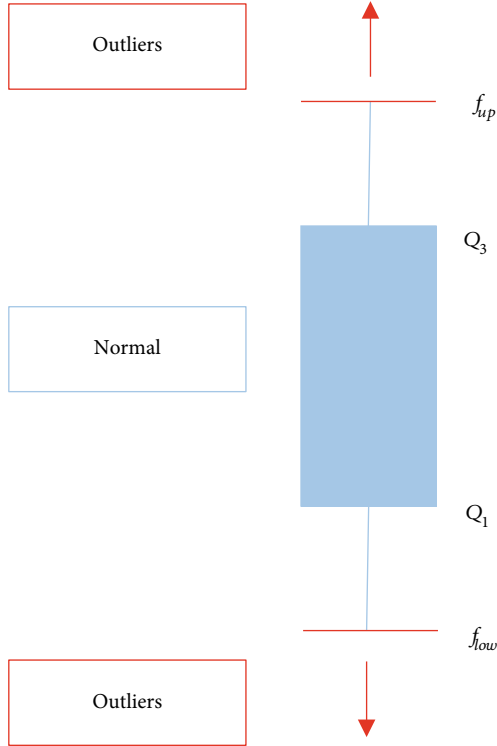


FIGURE 2: Rules for eliminating outliers from box plot.

TABLE 2: The stress-1 evaluation standard.

Stress-1	Goodness of fit
0.2	Poor
0.1	Fair
0.05	Good
0.025	Excellent
0	Perfect

model is established in the laboratory environment. Through the accelerated aging test, it can be concluded that the degradation of batteries is independent of each other and conforms to the normal distribution [34]. It is difficult to approximate the normal distribution of battery cells in the real world because of the complicated operating conditions of EVs. The reasons are as follows, and some reasons are verified in 4.

- (1) The number of battery cells varies from dozens [4] to thousands [8] of different vehicles. So the number of cells is insufficient to support the normal distribution theory for some EVs
- (2) Some EVs have battery packs installed in different locations of the EV. The battery packs are installed on the side and back of the EV in this study. Due to the different location temperatures and other conditions of the EVs, it will affect all the cells in the whole battery pack. Even the performance of all the cells in the battery packs is inferior to other battery packs

- (3) The cells are connected in series and parallel. In the process of charge and discharge, the cells affect each other, resulting in each cell not being completely independent. In addition, the different control strategies of the balancing system in the battery management system (BMS) also influence the parameters of the battery cells

The normal distribution has some simple and efficient methods for screening outliers, but forcing data to be considered as normal distribution will degrade the accuracy of the model. Therefore, battery cell distribution should be tested and analyzed when developing faulty cell diagnosis methods.

### 3. Methodology

**3.1. Kurtosis.** Kurtosis can be defined as the standardized fourth population moment about the mean, which is a dimensionless parameter [35]. It was originally used to represent impulsive characteristics in signal processing and quantify the waveform peaks. Kurtosis has been widely used in fault diagnosis [36], signal processing [37], and other fields. The formula of kurtosis is shown in formula (2). The  $\sigma$  is the standard deviation (SD) as shown in formula (3).

$$k = \frac{E(X - \mu)^4}{(E(X - \mu)^2)^2} = \frac{\mu_4}{\sigma^4}, \quad (2)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}, \quad (3)$$

where  $k$  is the kurtosis,  $E$  is the expected value,  $\mu$  is the sample mean, and  $\mu_4$  is the fourth moment about the mean.

**3.2. Multidimensional Scaling.** Multidimensional scaling (MDS) is a multivariate data analysis method that can show “distance” in low-dimensional space. The principle is that when determining the similarity of  $n$  objects, MDS can maximize the similarity of  $n$  objects in two-dimensional or three-dimensional space [38]. After optimization by Kruskal [39] et al., it has been widely used in fields such as medical analysis [40] and environmental research [41]. The specific calculation steps are as follows:

There are  $n$  objects in the set, and the similarities between each object are calculated with the correlation coefficient in the statistics. The  $\hat{d}_{ij}$  is the similarity between  $i$  and  $j$ . So the generalized distance matrix  $\hat{D} = \hat{d}_{ij}(n \times n)$  is constructed, and the nondiagonal elements of the matrix are arranged from small to large as follows:

$$\hat{d}_{i_1 j_1} \leq \hat{d}_{i_2 j_2} \leq \dots \leq \hat{d}_{i_l j_l} \dots \leq \hat{d}_{i_m j_m}, \quad m = \frac{1}{2} n(n-1), \quad i_l < j_l; \quad l = 1, 2, \dots, m. \quad (4)$$

To indicate ranking, the  $\hat{d}_{i_j j_i}$  is the similarity of  $i_i$  and  $j_i$

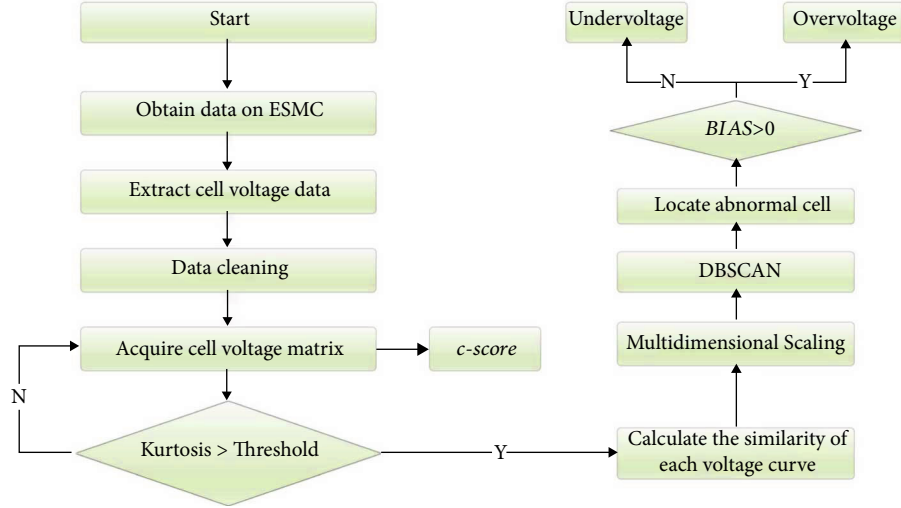


FIGURE 3: The flowchart of the voltage fault diagnosis method.

calculated by the correlation coefficient, and  $l$  is the similarity arranged in the  $l$ -th.

Euclidean distance matrix  $D$  is defined in  $r$ -dimensional space:

$$D = d_{ij}(n \times n), \quad d_{ij} \geq 0, \quad d_{ij} = 0, \quad i, j = 1, 2, \dots, n. \quad (5)$$

The  $d_{ij}$  is the Euclidean distance between elements  $i$  and  $j$  in  $r$ -dimensional space.

The matrix  $D$  satisfies the following conditions:

$$d_{i_1 j_1} \leq d_{i_2 j_2} \leq \dots \leq d_{i_l j_l} \leq \dots \leq d_{i_m j_m}, \quad m = \frac{1}{2}n(n-1), \quad i_l < j_l; \quad l = 1, 2, \dots, m. \quad (6)$$

The  $d_{i_l j_l}$  is the between elements  $i_l$  and  $j_l$ , and  $l$  is the distance arranged in the  $l$ -th.

To find a suitable point to make  $D$  and  $\hat{D}$  as similar as possible, MDS changes the position of  $n$  objects in  $r$ -dimensional space by iteration. To measure this similarity, Kruskal's normalized stress-1 criterion is defined as follows [42]:

$$\text{Stress} - 1 = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^n (d_{ij} - \hat{d}_{ij})^2}{\sum_{i=1}^n \sum_{j=1}^n d_{ij}^2}}. \quad (7)$$

The evaluation standard for stress-1 is shown in Table 2 [39].

In order to reduce dimensions, the center matrix  $B = b_{ij}$  ( $n \times n$ ) based on the distances of objects is constructed as follows:

$$b_{ij} = -\frac{1}{2} \left( d_{ij}^2 - \frac{1}{n} \sum_{j=1}^n d_{ij}^2 - \frac{1}{n} \sum_{i=1}^n d_{ij}^2 + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^2 \right), \quad i, j = 1, \dots, n. \quad (8)$$

The eigenvalues  $\Lambda$  and eigenvectors  $\Gamma$  of the  $B$  are calcu-

lated and arranged as follows:

$$\Lambda = \text{diag}(\lambda_1, \lambda_2 \dots \lambda_c \dots \lambda_r), \quad (9)$$

$$\Gamma = (e_1, e_2 \dots e_c \dots e_r),$$

where the  $\lambda_c$  is the  $c$ -th eigenvalue of matrix  $B$  and the  $e_c$  is the eigenvalue corresponding to the eigenvalue  $\lambda_c$  of matrix  $B$ . The  $r$  is the dimension of the MDS output.

The coordinates of each object in the low dimension space could be expressed as follows:

$$X = \Gamma \Lambda^{1/2}. \quad (10)$$

**3.3. Density-Based Spatial Clustering of Applications with Noise.** The density-based spatial clustering of applications with noise (DBSCAN) algorithm is a density-based spatial clustering method proposed by Ester et al. [43]. It is one of the most representative density-based clustering methods. The algorithm defines the following concepts: Define  $\epsilon$  as radius parameter and  $Minpts$  as neighborhood density threshold. Within a set,  $p$  is considered a core point if it contains objects within the radius of  $\epsilon$  that exceeds the  $Minpts$ . If the  $q$  is within a radius of  $p$ , the  $p$  and  $q$  are called direct density reachable. For the objects  $p_1, p_2, \dots, p_i, \dots, p_n$ , if all  $p_i$  and  $p_{i+1}$  are direct density reachable, the  $p_1$  and  $p_n$  are called density reachable. For  $w$  in the set, if both  $p_i$  and  $q_i$  are density reachable with  $w$ , the  $p_i$  and  $q_i$  are called density connection.

The process of the DBSCAN is as follows: All objects in the set are traversed through direct density reachable, density reachable, and density connection. When a cluster is clustered, the next point is selected to start the cluster until no new point is added to any one cluster. Finally, the above objects are clustered into several clusters, and the objects not included in any cluster are noise data [44].

**3.4. Diagnosis Method.** Firstly, the real-time operation data of EV is obtained by the ESMC. Through data cleaning, the obvious abnormal values of cell voltages are removed

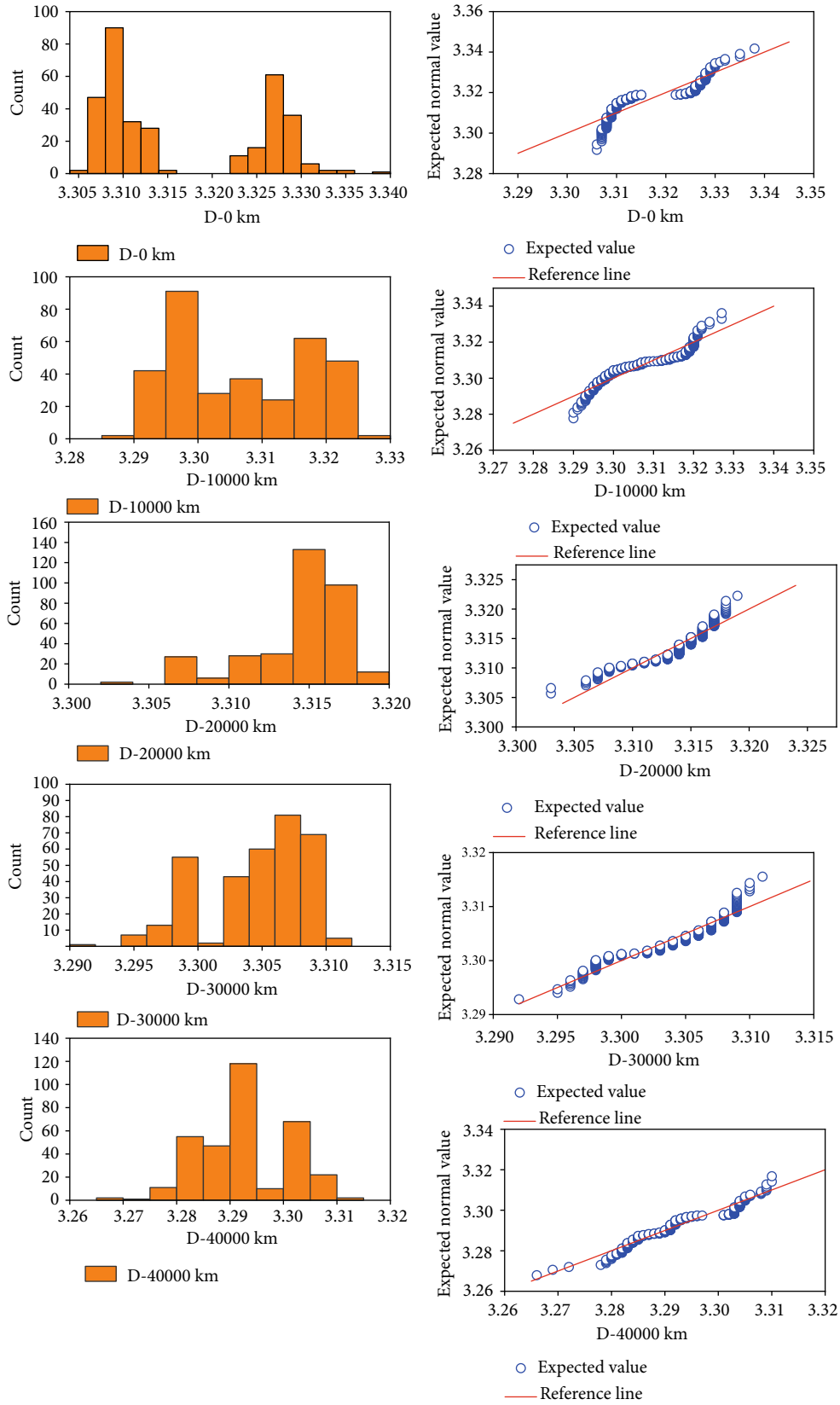


FIGURE 4: Voltage distribution in discharging conditions.

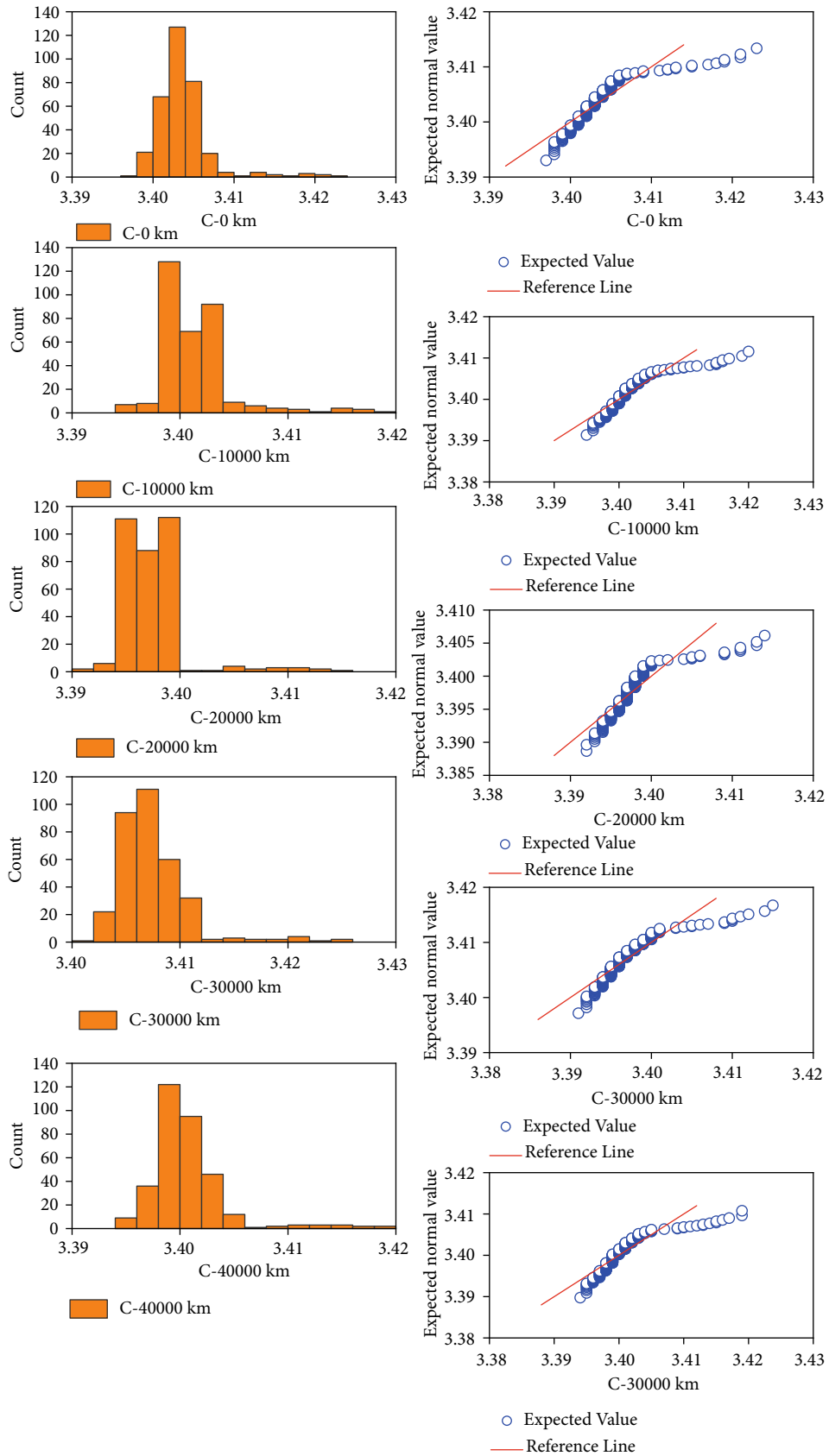


FIGURE 5: Voltage distribution in charging conditions.

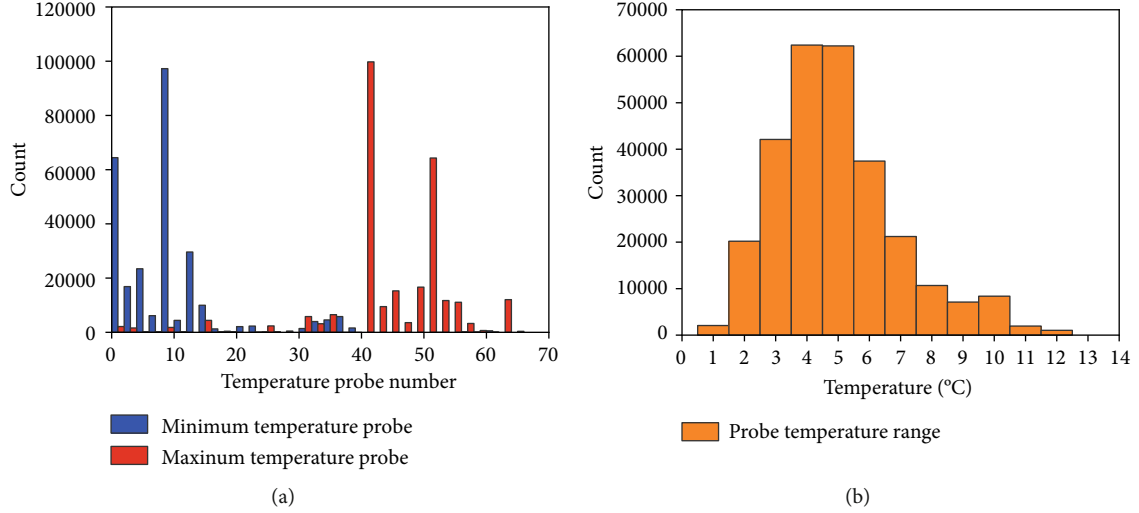


FIGURE 6: Temperature probe distribution and range. (a) The number distribution of the maximum temperature probe and the minimum temperature probe. (b) The range of the temperature of the probe.

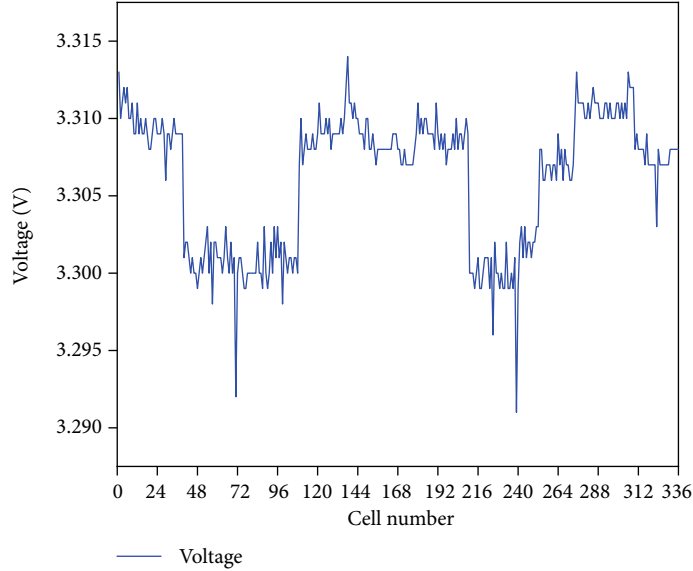


FIGURE 7: Cell voltage distribution at one moment.

and the missing values are filled. The cell voltage matrix  $U$  with  $t$  interval is collected.

The matrix  $U$  is as follows:

$$U = \begin{bmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,n} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,n} \\ \vdots & \vdots & u_{k,j} & \vdots \\ u_{t,1} & u_{t,2} & \cdots & u_{t,n} \end{bmatrix}_{t \times n}, \quad (11)$$

where  $t$  is the sampling time and  $n$  is the number of cells. The  $u_{k,j}$  is the voltage of the  $j$ -th cell in the  $k$ -th sampling time.

The sliding window without repeating sample is used for the voltage matrix  $U$ . The kurtosis of the matrix at each time is calculated. If the kurtosis exceeds the set threshold  $k_{\text{Thres}}$  for 3 consecutive moments ( $k_{i-1} > k_{\text{Thres}}, k_i > k_{\text{Thres}}, k_{i+1} > k_{\text{Thres}}$ ), it is determined as the battery cell fault. The above conditions will trigger the faulty cell alarm. The matrix  $K$  is shown below:

$$K = \begin{bmatrix} k_1 \\ \vdots \\ k_i \\ \vdots \\ k_t \end{bmatrix}_{t \times 1}, \quad (12)$$

where  $k_i$  is the kurtosis of the sampling at the time  $i$ .



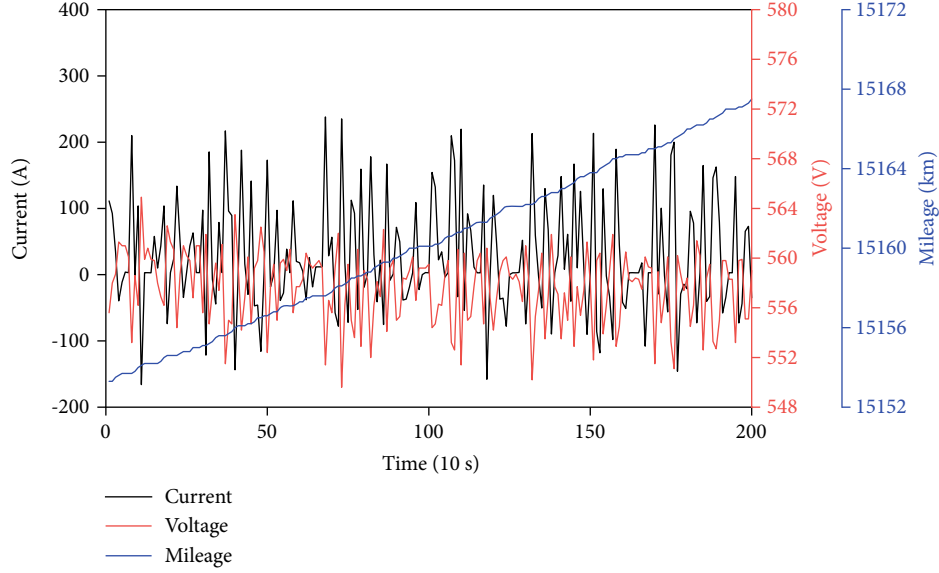


FIGURE 8: Some typical parameter curves of running EV.

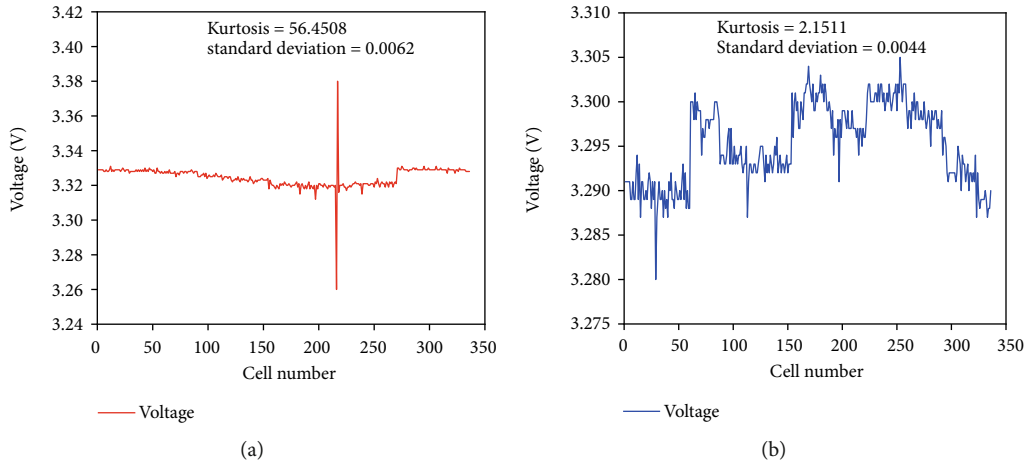


FIGURE 9: Kurtosis and SD in different cases. (a) The kurtosis and SD before the cell fault. (b) The kurtosis and SD during normal operation.

If the voltage alarm is triggered, the distance matrix is obtained by calculating the Euclidean distance of all cell voltage curves in the window. The Euclidean distance of cell voltage curves is calculated as follows:

$$d_{i,j} = \sqrt{\sum_{k=1}^t (u_{k,i} - u_{k,j})^2}, \quad (13)$$

where  $d_{i,j}$  is the similarity between the  $i$ -th cell voltage curve and the  $j$ -th cell voltage curve. The  $u_{k,i}$  is the  $i$ -th cell voltage value at the  $k$ -th moment.

The similarity between cell voltage curves is expressed by matrix  $D$ :

$$D = [d_{i,j}]_{n \times n} = \begin{bmatrix} 0 & d_{1,2} & \cdots & d_{1,n} \\ d_{2,1} & 0 & \cdots & d_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n,1} & d_{n,2} & \cdots & 0 \end{bmatrix}_{n \times n}, \quad (14)$$

If all cell voltage changes consistently during the operation of EV, matrix  $D$  should approach matrix 0.

The distance matrix  $D$  is mapped to two-dimensional space based on MDS. Normalize the data of the two dimensions, respectively, so that all data are between the interval  $[0,1]$ . DBSCAN is used to determine noise points in two-dimensional space, and the corresponding index numbers of noise points are faulty cell numbers.

When the faulty cell numbers are determined, it is judged that the fault cell is overvoltage or undervoltage by formula (15). The formula of  $\bar{u}_{k,j}$  is shown in (16)

$$\text{BIAS} = \frac{\sum_{k=1}^t (u_{k,j} - \bar{u}_{k,j})}{t}, \quad (15)$$

$$\bar{u}_{k,j} = \frac{\sum_{j=1}^n u_{k,j}}{n}, \quad (16)$$

where the  $u_{k,j}$  is the voltage of the  $j$ -th cell in the  $k$ -th

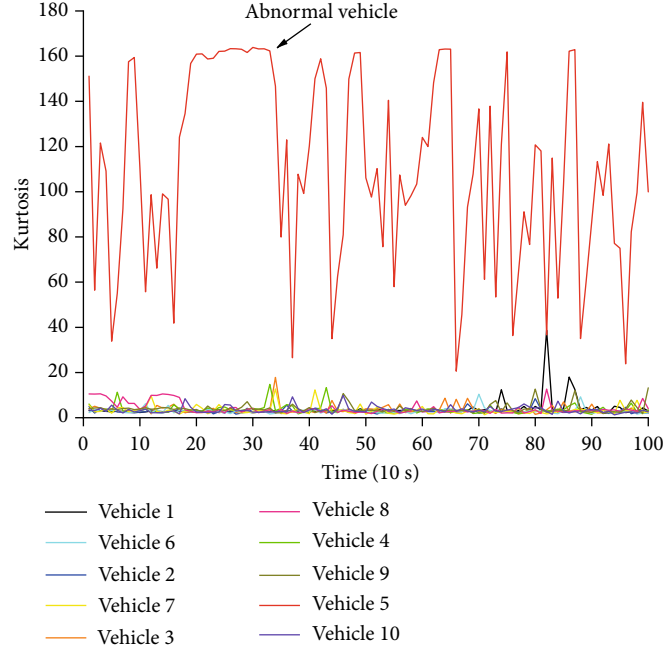


FIGURE 10: Kurtosis of the first window.

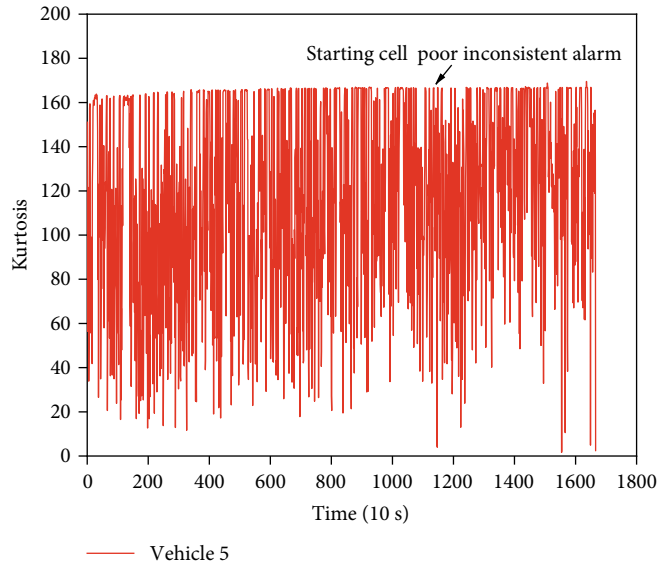


FIGURE 11: Kurtosis of all windows of vehicle 5.

sampling time. The  $\bar{u}_{k,j}$  is the average voltage of the cell at the  $k$ -th time.

The BIAS greater than 0 indicates cell overvoltage and less than 0 is cell undervoltage.

Define the concept of vehicle inconsistency score to quantify the safety of each vehicle as follows:

$$c\text{-score} = \frac{\sum_{i=1}^t k_i}{t}, \quad (17)$$

where the  $t$  is the total sampling time in a window and  $k_i$  is the kurtosis of  $i$  time. The higher the  $c$ -score of inconsis-

tency, the worse the consistency of the battery cells in the EV.

The characteristic of this diagnostic method is that the following calculation is performed to locate the faulty unit after the prealarm based on the kurtosis. This method reduces the computational load of the data centers and has practical application capabilities. The flowchart of the method is illustrated in Figure 3.

## 4. Results and Discussion

**4.1. Distribution Results.** To test whether the cell voltages of EVs in this paper conform to the normal distribution, cell

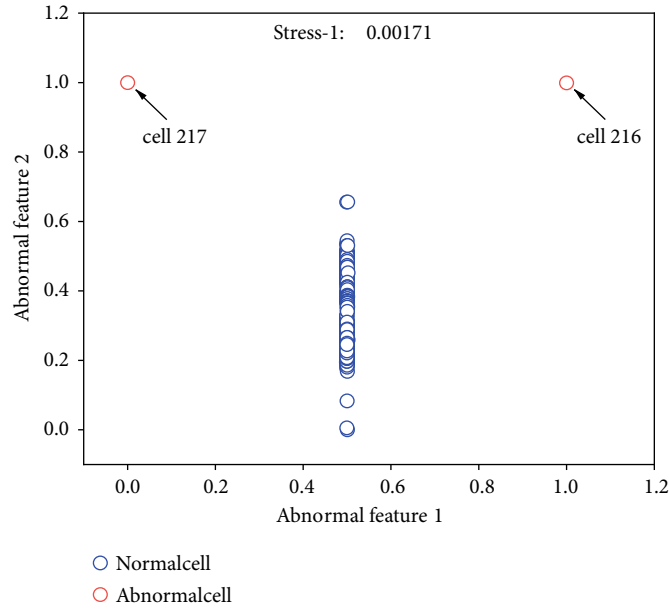


FIGURE 12: DBSCAN results for vehicle 5 before the alarm.

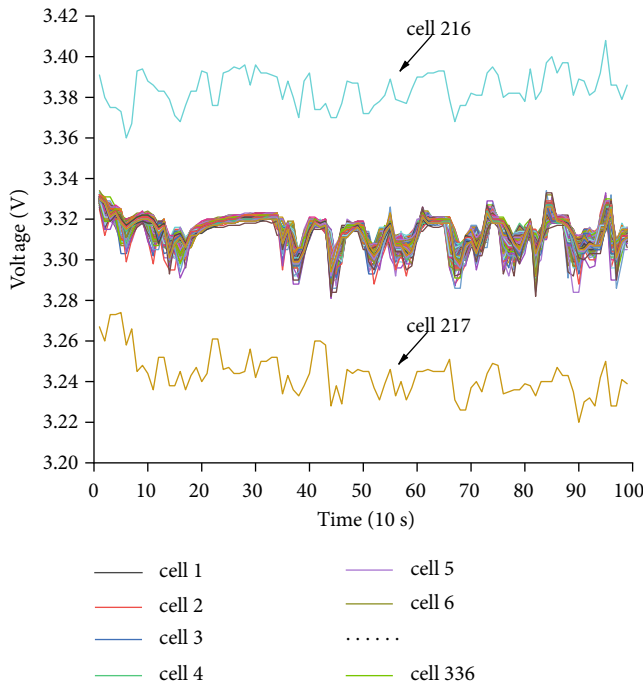


FIGURE 13: Cell voltage curves of vehicle 5 before the alarm.

voltages of five vehicles are selected for analysis with the change of mileage. The cell voltage distribution of EVs at 0 km, 10, 000 km, 20, 000 km, 30, 000 km, and 40, 000 km is extracted. Restrictions were added to ensure that it is not disturbed by other factors:

- (1) The charging and discharging conditions of EVs are distinguished, and the voltage distribution during charging and discharging is recorded for each vehicle at different mileage

- (2) For discharge case, the total discharge current is between 23.5 A and 26.5 A, and SOC is between 78% and 82%. For charging case, SOC is between 78% and 82%, and charging current is platform period with an error no more than 5 A

Based on the above restrictions, 50 samples of cell voltage distribution were obtained. The distribution of the five vehicles is similar, taking vehicle 1 as an example. The voltage distribution and Q-Q diagram of vehicle 1 during discharge are shown in Figure 4 and during charging are shown in Figure 5. The *D* is the discharge state of EVs in Figure 4 and the *C* is the charge state of EVs in Figure 5.

The voltage distribution is similar to the long tail distribution when the battery is discharged, and the Q-Q plots are similar to the inclined “S.” The reason for the long tail distribution is that there are double peaks, indicating that more data deviate from the normal distribution, which tends to weaken with the increase of mileage. But the voltage distribution approximates the left-skewed distribution when the battery is charged, and the Q-Q plots are similar to the inverted “U.” The reason for the left-skewed distribution is that more data are centered on the left. The cell voltages are more concentrated when charging because of the balance system of the BMS.

Based on the Shapiro-Wilk test, none of the above 50 samples could be accepted as normally distributed at the 0.05 level, which proves that the battery voltage of EVs does not conform to the normal distribution regardless of the mileage change or whether it is charged or discharged. To increase the test sample size, 2,000 driving segments (whether charging or discharging) were randomly selected from multiple vehicles for the Shapiro-Wilk test, and the results did not conform to the normal distribution.

For the vehicles in this study, 64 temperature probes were installed in 12 battery packs. Based on the one-year driving data of a vehicle, the number of distribution of the

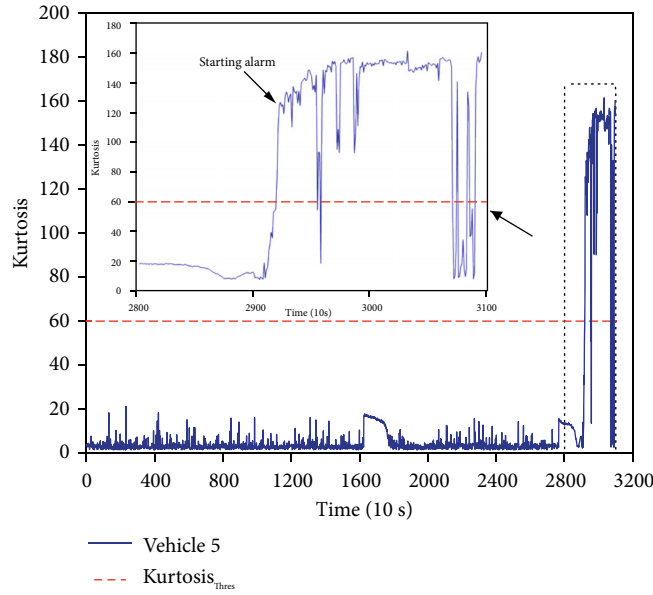


FIGURE 14: Vehicle 5 kurtosis of all windows one day earlier.

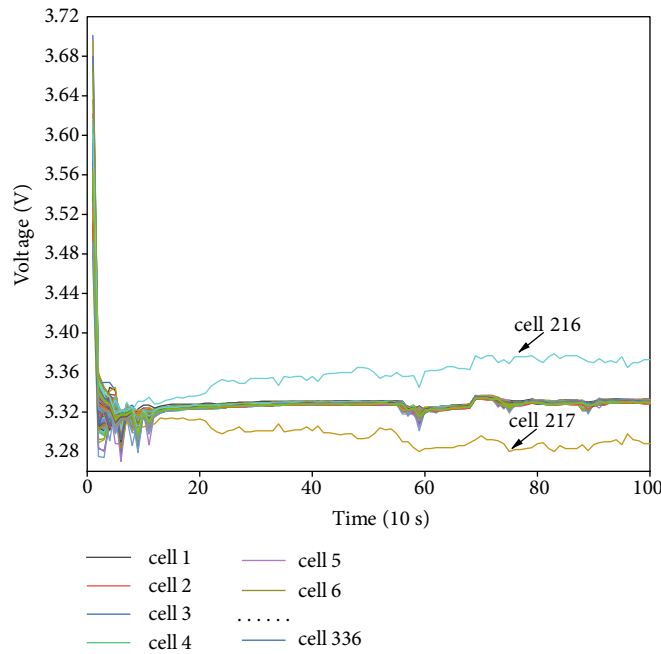


FIGURE 15: Cell voltage curves of vehicle 5 one day earlier.

maximum temperature probe and the minimum temperature probe are shown in Figure 6(a). The range is the difference between the maximum and minimum probe temperatures, and the range of the temperature of the probe is shown in Figure 6(b). This means that the cells in different locations are in different temperatures, and the temperature range of the probe is greater than 3°C many times, which will affect the performance of some battery packs. The difference in battery packs leads to the distribution of cell voltage like a square wave at a moment as shown in Figure 7.

**4.2. Cell Fault Early Alarm.** The Figure 8 shows the battery voltage, battery current, and mileage of an EV, while it is driving. When the driver accelerates the vehicle, the current of the battery increases, and the voltage decreases slightly. When the driver brakes the vehicle, the current is negative because of the braking energy feedback.

The vast majority of cell voltages is normal during the running time of an EV. Only a few cells are abnormal, and the voltage differences with other normal cells are subtle, especially at the beginning of the cell fault. SD is commonly used

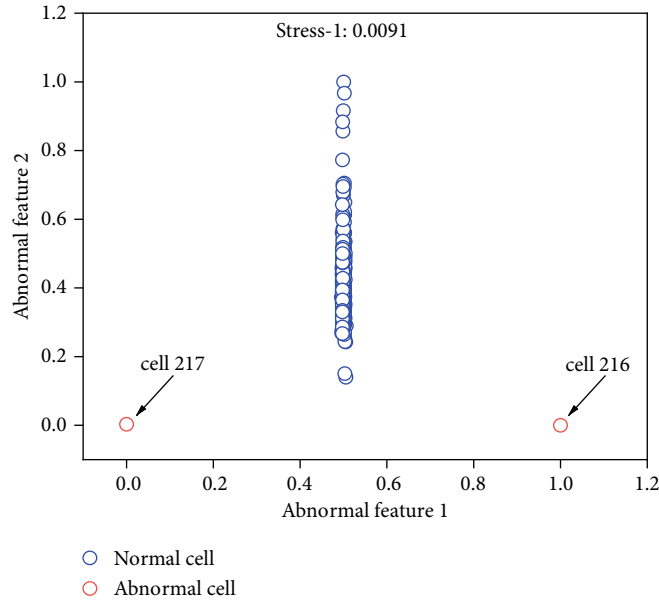


FIGURE 16: DBSCAN results for vehicle 5 before alarm a day earlier.

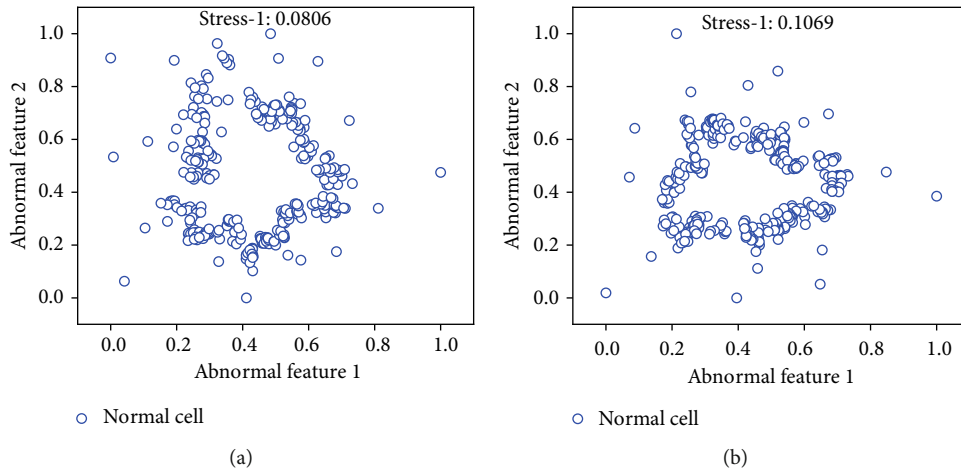


FIGURE 17: DBSCAN results for the normal vehicles. (a) DBSCAN results of vehicle 1. (b) DBSCAN results of vehicle 2.

TABLE 3: BIAS of fault cells.

Cell number	BIAS
216 (2020/5/24)	-3.039%
217 (2020/5/24)	2.957%
216 (2020/5/25)	-7.052%
217 (2020/5/25)	7.026%

in many EVs as an early warning indicator for detecting cell faults [6]. However, traditional detection indicators such as SD cannot detect cell faults and alarms at the beginning of failure. Kurtosis can sensitively capture peaks in voltage curves.

The following figures compare the warning capabilities of SD and kurtosis for cell fault. The cell voltage distribution of one hour before the cell fault alarm on that day is shown in Figure 9(a), and the voltage distribution curve of a vehicle

during normal operation is shown in Figure 9(b). The SD of the normal moment and the moment before the fault changes slightly. But the kurtosis before the cell fault is much greater than that at the normal moment, which can effectively identify the fault and avoid the problem of false alarms caused by setting a small threshold.

The operation data of 10 EVs is selected. The cell voltages are extracted based on the sliding window, and the window size is 100. The kurtosis for each window of 10 vehicles is calculated. In the first window, the kurtosis of vehicle 5 is significantly higher than that of other vehicles as shown in Figure 10. However, at this time, vehicle 5 has no faulty cell alarm, and the cell fault alarm is not triggered until 3 hours later as shown in Figure 11. But the fault is detected by the kurtosis at the first window. The above research proves that kurtosis can be used as an indicator of the faulty cell early warning.

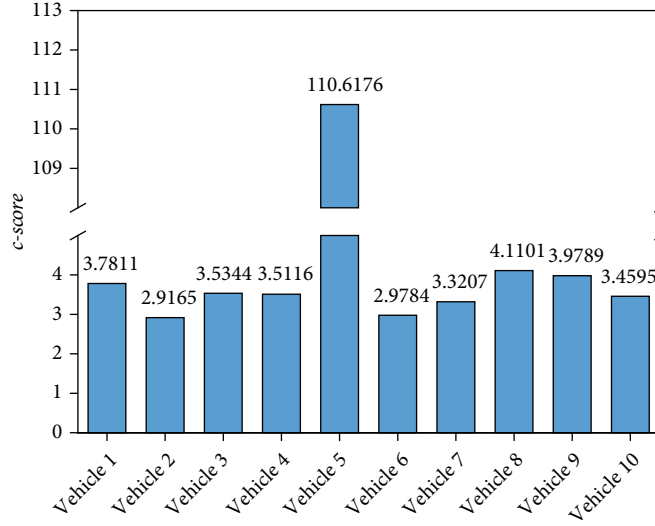


FIGURE 18:  $c$ -score of the 10 vehicles.

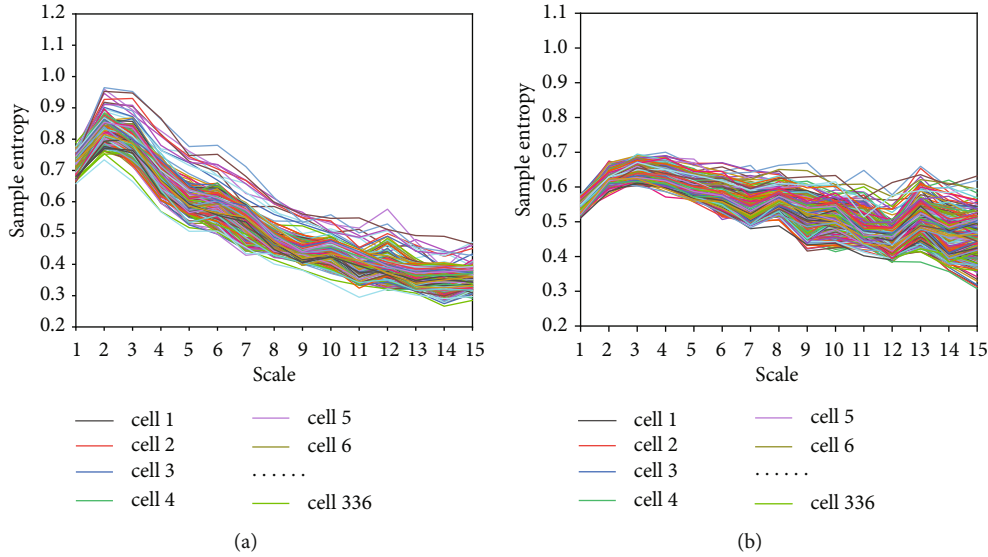


FIGURE 19: Result of multiscale entropy for cells in different vehicles. (a) Result of multiscale entropy for the normal vehicles. (b) Result of multiscale entropy for vehicle 5.

**4.3. Fault Cell Identification.** The similarity matrix between the voltage curves of 336 cells of vehicle 5 is calculated. MDS is used for distance matrices, and the results are mapped in two-dimensional space. The stress-1 is 0.00171 indicating that DMS fits well. The DBSCAN in two-dimensional space shows that cells 216 and 217 are distinct outliers. The result of DBSCAN is illustrated in Figure 12. Observing the voltage curves, it is found that the voltage curves of cells 216 and 217 are offset. The voltage curves of 336 cells corresponding to the first window are shown in Figure 13.

Although the actual offset is small, this difference can be magnified based on MDS and DBSCAN for cell fault diagnosis.

Through a large number of sampling windows of the 10 vehicles, the kurtosis alarm threshold of 60 is set by a trial and error method. The  $\epsilon$  is 0.3 and  $Minpts$  is 5 in DBSCAN.

The data of the day before vehicle 5 triggers the alarm is selected. According to the fault diagnosis method proposed in this paper, the kurtosis exceeds the set value in sampling from 2,900 to 3,000 samples as shown in Figure 14.

With the analysis of the cell voltage curves, the voltage curves of cells 216 and 217 begin to deviate gradually as shown in Figure 15. The higher cell voltage in the first few minutes is due to the charging of the vehicle. The early warning of this method is one day earlier than the actual alarm and can identify abnormal cells as shown in Figure 16.

Predagnosis of battery cells based on kurtosis is necessary. Outlier cells were not found to use the MDS and DBSCAN for normal vehicles. The DBSCAN results of normal vehicles are shown in Figure 17(a), 17(b). More importantly, the MDS and DBSCAN for each window will increase the data center computing load.

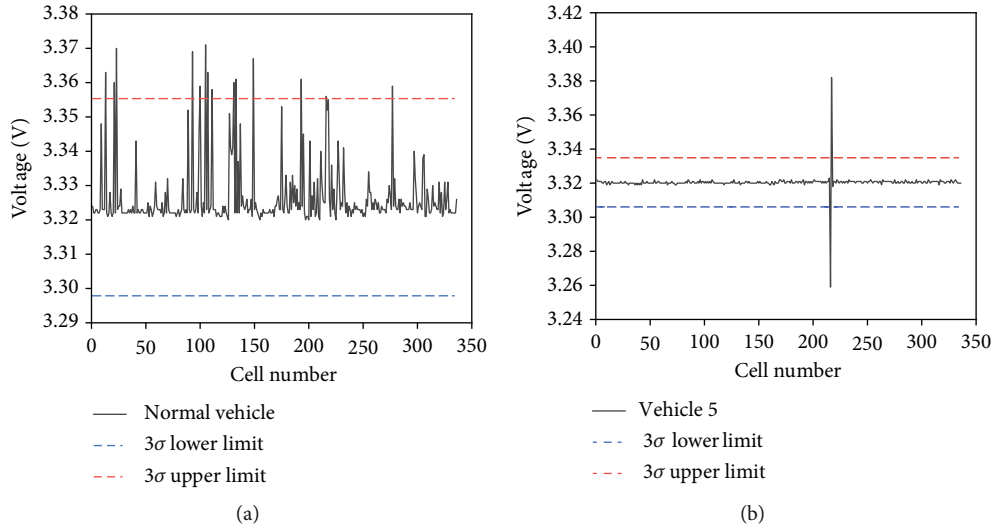


FIGURE 20: Result of  $3\sigma$  method for cells in different vehicles. (a) Result of  $3\sigma$  method for the normal vehicles. (b) Result of  $3\sigma$  method for the vehicle 5.

**4.4. Fault Quantification.** Based on the BIAS, cells 216 and 217 are quantified to determine overvoltage or undervoltage, respectively. The cell consistency of the vehicle deteriorates from the day before the alarm. The decrease in cell 216 is almost equal to the increase in cell 217 as shown in Table 3. The reason is that cells 216 and 217 are connected in parallel, and changes in the internal resistance or capacity of one cell affect the other.

According to the formula (17), the  $c$ -score of the 10 vehicles is shown in Figure 18. The higher the  $c$ -score, the worse the consistency of the vehicle. The cell voltage consistency of nine vehicles is fair, while the inconsistency  $c$ -score of vehicle 5 is significantly higher than other vehicles.

**4.5. Method Comparison.** This method is compared with common unsupervised algorithms such as  $3\sigma$  [4] and entropy-based [6]. In this paper, the ability of multiscale sample entropy is tested in diagnosing abnormal cells, and the multiscale entropy of 336 cells is calculated. Generally, the template of length  $m$  points is 2. The range of tolerance  $r$  is generally between  $0.1 \times SD$  and  $0.25 \times SD$ . In this paper, the  $r = 0.15 \times SD$  by trial and error method and the scale = 15. In practice, it is found that more data is needed for multiscale sample entropy; otherwise, matching vectors will not be found in the calculation process, which will lead to the absence of a definition of sample entropy. Each cell of vehicle 5 is calculated with multiscale sample entropy as shown in Figure 19(a). Figure 19(b) is the multiscale sample entropy of a normal vehicle. The sample entropies of each individual are very small, so it is difficult to identify the fault cells.

The  $3\sigma$  method is used for cell fault voltage diagnosis, which can accurately identify faults in advance as shown in Figure 20(a). But some normal cells are diagnosed as faulty cells as shown in Figure 20(b). Because the cell voltage data are not normally distributed, this conclusion was validated before the method was developed.

## 5. Conclusions

In this paper, the novel method for lithium-ion battery fault diagnosis of EV based on real-time voltage is presented. The effectiveness of the method is verified based on the real-time data collected by EVs. The related conclusions are drawn as follows:

- (1) The cell voltage distribution of some types of EVs is verified not to be a normal distribution in this work. The reason is the limitation of the number of cells and the interaction of the cells in series and parallel. Other reasons include the influence of the balance system of BMS, and the location of the battery pack is different
- (2) Kurtosis is used as an early warning indicator to find faulty cells for the first time. It is verified that kurtosis has a better ability to detect cell faults than SD. The fault cells are identified through MDS and DBSCAN. The faulty cells were detected one day earlier than the previous method. Furthermore, the BIAS judges the type of cell fault and quantifies the degree of fault, and the  $c$ -score evaluates the cell inconsistency of EVs
- (3) Comparisons of the entropy-based method and  $3\sigma$  method verify the superiority of this method. It is difficult to find faulty cells in advance based on the entropy-based method. The  $3\sigma$  method can accurately identify the cell faults in advance, but some normal cells are diagnosed as faulty cells

Future work can be carried out in the following areas:

- (1) The existing data preprocessing methods do not combine the operation principle of EVs. In the future, some battery parameters will be combined with machine learning to propose a more efficient and accurate filling strategy

- (2) In this paper, this battery fault diagnosis model was established based on the cell voltage. A fault diagnosis model based on multivariate information fusion will be established combining temperature, voltage, and capacity data of the cells in the future
- (3) The charge and discharge states of EVs should be discussed separately in subsequent studies, which can better distinguish the differences between the static and dynamic processes of EVs

## Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## References

- [1] R. Yang, R. Xiong, S. Ma, and X. Lin, "Characterization of external short circuit faults in electric vehicle Li-ion battery packs and prediction using artificial neural networks," *Applied Energy*, vol. 260, article 114253, 2020.
- [2] J. Hu, Z. Wei, and H. He, "An online adaptive internal short circuit detection method of lithium-ion battery," *Automotive Innovation*, vol. 4, no. 1, pp. 93–102, 2021.
- [3] Y. Huang, Y. Tang, W. Yuan et al., "Challenges and recent progress in thermal management with heat pipes for lithium-ion power batteries in electric vehicles," *SCIENCE CHINA Technological Sciences*, vol. 64, no. 5, pp. 919–956, 2021.
- [4] Y. Zhao, P. Liu, Z. Wang, L. Zhang, and J. Hong, "Fault and defect diagnosis of battery for electric vehicles based on big data analysis methods," *Applied Energy*, vol. 207, pp. 354–362, 2017.
- [5] J. Hong, Z. Wang, and Y. Yao, "Fault prognosis of battery system based on accurate voltage abnormality prognosis using long short-term memory neural networks," *Applied Energy*, vol. 251, article 113381, 2019.
- [6] J. Hong, Z. Wang, W. Chen, and L. Wang, "Multi-fault synergistic diagnosis of battery systems based on the modified multi-scale entropy," *International Journal of Energy Research*, vol. 43, no. 14, pp. 8350–8369, 2019.
- [7] Z. Wang, J. Hong, P. Liu, and L. Zhang, "Voltage fault diagnosis and prognosis of battery systems based on entropy and Z-score for electric vehicles," *Applied Energy*, vol. 196, pp. 289–302, 2017.
- [8] R. Xiong, W. Sun, Q. Yu, and F. Sun, "Research progress, challenges and prospects of fault diagnosis on battery system of electric vehicles," *Applied Energy*, vol. 279, article 115855, 2020.
- [9] M. Adaikkappan and N. Sathiyamoorthy, "Modeling, state of charge estimation, and charging of lithium-ion battery in electric vehicle: a review," *International Journal of Energy Research*, vol. 46, no. 3, pp. 2141–2165, 2022.
- [10] L. Y. Wang, L. F. Wang, W. Liu, and Y. W. Zhang, "Research on fault diagnosis system of electric vehicle power battery based on OBD technology," *Proceedings of the 2017 International Conference on Circuits, Devices and Systems (ICCDs)*, pp. 95–99, 2017.
- [11] J. Jiang, X. Cong, S. Li, C. Zhang, W. Zhang, and Y. Jiang, "A hybrid signal-based fault diagnosis method for lithium-ion batteries in electric vehicles," *IEEE Access*, vol. 9, pp. 19175–19186, 2021.
- [12] C. Wu, C. Zhu, and Y. Ge, "A new fault diagnosis and prognosis technology for high-power lithium-ion battery," *IEEE Transactions on Plasma Science*, vol. 45, no. 7, pp. 1533–1538, 2017.
- [13] X. Hu, K. Zhang, K. Liu, X. Lin, S. Dey, and S. Onori, "Advanced fault diagnosis for lithium-ion battery systems: a review of fault mechanisms, fault features, and diagnosis procedures," *IEEE Industrial Electronics Magazine*, vol. 14, no. 3, pp. 65–91, 2020.
- [14] R. Xiong, Q. Yu, W. Shen, C. Lin, and F. Sun, "A sensor fault diagnosis method for a lithium-ion battery pack in electric vehicles," *IEEE Transactions on Power Electronics*, vol. 34, no. 10, pp. 9709–9718, 2019.
- [15] J. Wei, G. Dong, and Z. Chen, "Model-based fault diagnosis of lithium-ion battery using strong tracking extended Kalman filter," *Energy Procedia*, vol. 158, pp. 2500–2505, 2019.
- [16] S. Kong, M. Saif, and G. Cui, "Estimation and fault diagnosis of lithium-ion batteries: a fractional-order system approach," *Mathematical Problems in Engineering*, vol. 2018, 12 pages, 2018.
- [17] S. Dey, H. E. Perez, and S. J. Moura, "Model-based battery thermal fault diagnostics: algorithms, analysis, and experiments," *IEEE Transactions on Control Systems Technology*, vol. 27, pp. 576–587, 2017.
- [18] X. Kong, G. L. Plett, M. S. Trimboli et al., "Pseudo-two-dimensional model and impedance diagnosis of micro internal short circuit in lithium-ion cells," *Journal of Energy Storage*, vol. 27, article 101085, 2020.
- [19] L. Jiang, Z. Deng, X. Tang, L. Hu, X. Lin, and X. Hu, "Data-driven fault diagnosis and thermal runaway warning for battery packs using real-world vehicle data," *Energy*, vol. 234, p. 121266, 2021.
- [20] L. Yao, Y. Xiao, X. Gong, J. Hou, and X. Chen, "A novel intelligent method for fault diagnosis of electric vehicle battery system based on wavelet neural network," *Journal of Power Sources*, vol. 453, article 227870, 2020.
- [21] D. Li, Z. Zhang, P. Liu, Z. Wang, and L. Zhang, "Battery fault diagnosis for electric vehicles based on voltage abnormality by combining the long short-term memory neural network and the equivalent circuit model," *IEEE Transactions on Power Electronics*, vol. 36, pp. 1303–1315, 2020.
- [22] R. Yang, R. Xiong, H. He, and Z. Chen, "A fractional-order model-based battery external short circuit fault diagnosis approach for all-climate electric vehicles application," *Journal of Cleaner Production*, vol. 187, pp. 950–959, 2018.
- [23] B. Duan, Z. Li, P. Gu, Z. Zhou, and C. Zhang, "Evaluation of battery inconsistency based on information entropy," *Journal of Energy Storage*, vol. 16, pp. 160–166, 2018.
- [24] P. Liu, Z. Sun, Z. Wang, and J. Zhang, "Entropy-based voltage fault diagnosis of battery systems for electric vehicles," *Energies*, vol. 11, no. 1, p. 136, 2018.
- [25] X. Li, K. Dai, Z. Wang, and W. Han, "Lithium-ion batteries fault diagnostic for electric vehicles using sample entropy analysis method," *Journal of Energy Storage*, vol. 27, article 101121, 2020.
- [26] B. Xia, Y. Shang, T. Nguyen, and C. Mi, "A correlation based fault detection method for short circuits in battery packs," *Journal of Power Sources*, vol. 337, pp. 1–10, 2017.



- [27] Z. Sun, Y. Han, Z. Wang et al., "Detection of voltage fault in the battery system of electric vehicles using statistical analysis," *Applied Energy*, vol. 307, p. 118172, 2022.
- [28] Z. Wang, C. Song, L. Zhang, Y. Zhao, P. Liu, and D. G. Dorrell, "A data-driven method for battery charging capacity abnormality diagnosis in electric vehicle applications," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 1, pp. 990–999, 2022.
- [29] C. She, L. Zhang, Z. Wang, F. Sun, P. Liu, and C. Song, "Battery state of health estimation based on incremental capacity analysis method: synthesizing from cell-level test to real-world application," *Electronics*, p. 1, 2021.
- [30] Z. Chen, K. Xu, J. Wei, and G. Dong, "Voltage fault detection for lithium-ion battery pack using local outlier factor," *Measurement*, vol. 146, pp. 544–556, 2019.
- [31] L. Zhang, X. Hu, Z. Wang, F. Sun, J. Deng, and D. G. Dorrell, "Multiobjective optimal sizing of hybrid energy storage system for electric vehicles," *IEEE Transactions on Vehicular Technology*, vol. 67, pp. 1027–1035, 2017.
- [32] A. Ali-Eldin and E. Elmroth, "Data management, communication systems and the edge," *Challenges for the future of transportation*, vol. 1, article 100024, 2021.
- [33] Q. Xue, G. Li, Y. Zhang, S. Shen, Z. Chen, and Y. Liu, "Fault diagnosis and abnormality detection of lithium-ion battery packs based on statistical distribution," *Journal of Power Sources*, vol. 482, article 228964, 2021.
- [34] W. Zhen-Po, L. Peng, and W. Li-Fang, "Analysis on the capacity degradation mechanism of a series lithium-ion power battery pack based on inconsistency of capacity," *Chinese Physics B*, vol. 22, no. 8, article 088801, 2013.
- [35] L. T. DeCarlo, "On the meaning and use of kurtosis," *Psychological Methods*, vol. 2, no. 3, pp. 292–307, 1997.
- [36] X. Song, H. Wang, and P. Chen, "Weighted kurtosis-based VMD and improved frequency-weighted energy operator low-speed bearing-fault diagnosis," *Measurement Science and Technology*, vol. 32, no. 3, article 035016, 2021.
- [37] M. Elforjani and E. Bechhoefer, "Analysis of extremely modulated faulty wind turbine data using spectral kurtosis and signal intensity estimator," *Renewable Energy*, vol. 127, pp. 258–268, 2018.
- [38] I. Borg and P. J. Groenen, *Modern Multidimensional Scaling: Theory and Applications*, Springer Science & Business Media, 2005.
- [39] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.
- [40] E. Alnazly, O. M. Khraisat, A. M. Al-Bashaireh, and C. L. Bryant, "Anxiety, depression, stress, fear and social support during COVID-19 pandemic among Jordanian healthcare workers," *PLoS One*, vol. 16, no. 3, article e0247679, 2021.
- [41] M. Das, S. Samanta, D. Sudheesan et al., "Fish diversity, community structure and ecological integrity of river," *Brahmani*, 2021.
- [42] I. Borg, P. J. Groenen, P. Mair, I. Borg, P. J. F. Groenen, and P. Mair, "The goodness of an MDS solution," in *Applied Multidimensional Scaling*, pp. 21–26, Springer, 2013.
- [43] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the kdd*, pp. 226–231, 1996.
- [44] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "DBSCAN revisited, revisited: why and how you should (still) use DBSCAN," *ACM Transactions on Database Systems (TODS)*, vol. 42, no. 3, pp. 1–21, 2017.