

# Fault Diagnosis and Abnormality Detection of Lithium-ion Battery Packs Based on Statistical Distribution

Qiao Xue<sup>1</sup>, Guang Li<sup>2</sup>, Yuanjian Zhang<sup>3</sup>, Shiquan Shen<sup>1</sup>, Zheng Chen<sup>1,2\*</sup>, and Yonggang Liu<sup>4\*</sup>

<sup>1</sup>Faculty of Transportation Engineering, Kunming University of Science and Technology, Kunming, 650500, China

<sup>2</sup>School of Engineering and Materials Science, Queen Mary University of London, London, E1 4NS, United Kingdom

<sup>3</sup>Sir William Wright Technology Center, Queen's University Belfast, Belfast, BT9 5BS, United Kingdom

<sup>4</sup>State Key Laboratory of Mechanical Transmissions & School of Automotive Engineering, Chongqing University, Chongqing, 400044, China

Email: strxue\_qiao@163.com, g.li@qmul.ac.uk, y.zhang@qub.ac.uk, shiquan219@gmail.com, chen@kust.edu.cn, andylyg@umich.edu.

Corresponding author: Zheng Chen (chen@kust.edu.cn), and Yonggang Liu (andylyg@umich.edu)

**Abstract:** Lithium-ion battery packs are widely deployed as power sources in transportation electrification solutions. To ensure safe and reliable operation of battery packs, it is of critical importance to monitor operation status and diagnose the running faults in a timely manner. This study investigates a novel fault diagnosis and abnormality detection method for battery packs of electric scooters based on statistical distribution of operation data that are stored in the cloud monitoring platform. According to the battery current and scooter speed, the operation states of electric scooters are clarified, and the diagnosis coefficient is determined based on the Gaussian distribution to highlight the parameter variation in each state. On this basis, the K-means clustering algorithm, the Z-score method and  $3\sigma$  screening approach are exploited to detect and locate the abnormal cells. By analyzing the abnormalities hidden beneath the external measurement and calculating the fault frequency of each cell in pack, the proposed algorithm can identify the faulty type and locate the faulty cell in a timely manner. Experimental results validate that the proposed method can accurately diagnose faults and monitor the status of battery packs. This theoretical study with practical implications shows the promising research direction of combining data mining technologies with machine learning methods for fault diagnosis and safety management of complex dynamical systems.

**Key Words:** Electric scooters, battery pack, fault diagnosis, abnormality detection, Gaussian distribution.

## I. INTRODUCTION

Global warming, environmental pollution and oil crisis have raised worldwide concerns [1], and transportation electrification can effectively mitigate their passive influences [2]. Because of lightness,

convenience as well as abilities of alleviating urban traffic congestion and exhaust pollution, electric scooters (E-scooters) have become attractive solutions to travels in a short distance (such as less than 5 Km) [3]. Lithium-ion batteries are normally assembled into packs [4] to supply the driving power for E-scooters due to their advantages of high energy density, long cycle life, no memory effect and wide operation temperature [5]. Compared to battery systems for electric vehicles (EVs) [6], E-scooters only deploy a smaller power battery pack which may be composed of dozens of cells structured in a series/parallel topology [7]. Nevertheless, it is still imperative to employ the so-called battery management systems (BMS) to ensure safe and reliable operation of battery packs. [8]. As a key function of BMS, the real-time fault diagnosis and safety control of battery packs are always indispensable to mitigate occurrence of harmful operations [9]. Currently, fault diagnosis technologies have been substantially developed and applied based on advanced models and algorithms. However, different from other mechanical or electrical systems, lithium-ion battery packs form a quite complex system consisting of a variety of sub-systems, such as cells, thermal-control unit and BMS [10]. In recent years, increased failure risks of battery systems promote research on faster fault diagnosis and higher safety management [11].

Common electrical faults of battery packs can be divided into three categories: abuse [12], sensor faults [13] and connection faults [14]. Battery abuse faults mainly refer to external short circuit (ESC), internal short circuit (ISC), overcharge and over-discharge. Sensor faults usually indicate abnormal operation of current transducers as well as voltage and temperature sensors, and connection faults are usually caused by loose contact between neighboring cells. Abuse faults and connection faults may lead to irreversible battery damages, such as thermal runaway and even catastrophic accidents [15]. Generally, on-board fault diagnosis of battery packs is attained through setting the safe operation thresholds, by which the electrical faults can be judged by comparison with the actual parameter values. Although the threshold judgement is reliable and easy to implement, it does not allow timely, efficient and secure management of battery packs in practice, as all the fault types may not be reported synchronously, and potential abnormality cannot be detected only by simple rules. To overcome this bottleneck, model-based observation, machine learning and data mining technologies emerge to tackle the hidden abnormality and help prognose unexpected forthcoming behaviors [16]. Ref. [17] conducts a comprehensive investigation and review of model-based and signal-based fault diagnosis techniques for lithium-ion batteries. Ref. [18] summarizes the approaches and applications of fault diagnosis,

which can be classified into the knowledge-based and hybrid/active methods. In [19], a two-layer model-based ESC diagnosis algorithm is proposed via investigating the fault characteristics of lithium-ion batteries. In [13], a model-based sensor fault diagnosis and isolation scheme based on adaptive extended Kalman filter (EKF) is employed to detect the current or voltage sensor fault with low computational efforts. In [20], an adaptive fault diagnosis method is proposed based on multiple nonlinear models and EKF, where the EKF is leveraged to estimate the terminal voltage of each model and generate the residual signals, thereby calculating the probabilities of fault signatures. In [21], a lithium-ion battery fault diagnosis system suitable for high-power scenarios is designed, and it can evaluate the degradation of lithium-ion batteries and conduct diagnosis with the full knowledge of internal fault mechanism. Ref. [22] proposes an online multi-fault diagnosis method based on a non-redundant crossed-style measurement circuit and an improved correlation coefficient approach. By these approaches, the cell faults can be detected by identifying the correlation coefficient of neighboring voltages with fault marks. In [23], a multi-fault detection method is proposed based on the improved sample entropy for real-time fault detection even when no obvious voltage variation appears in an early stage. The proposed method can effectively identify the fault type without the dependence on battery model. However, the inconsistency among cells is not taken into consideration. In [24], a fractional first-order resistance-capacitance (RC) model is constructed to characterize the electrical characteristics of battery, and the random forest model is proposed to identify the behavior of electrolyte leakage when the ESC fault occurs. Nonetheless, due to complexity and nonlinearities of battery systems, the fault cause is difficult to identify or judge in practical operations [25]. Although these works have advanced a body of fault diagnosis algorithms, most of them are simply based on static experiments data, which may be not qualified for complex time-varying operating scenarios [26].

Presently, statistical methods have been widely employed for fault diagnosis [27] and also applied to safety management of battery systems [28]. Through analysis and training of a large number of historical operation data, an early fault warning algorithm using long short-term memory (LSTM) network is proposed for the prediction of voltage abnormality [29]. Ref. [30] proposes a statistical method to detect the abnormal variations of cell terminal voltages based on the Gaussian distribution and a multi-level screening strategy. Ref. [31] presents a voltage fault diagnosis method based on the improved Shannon entropy, in which the forthcoming voltage fault can be effectively predicted and prevented. In [32], an interclass correlation

coefficient (ICC) method is investigated to calculate the correlation values of voltages of adjacent battery cells for fault diagnosis and fault location. However, the proposed methods in these works [29-32] are mainly based on the voltage data of a single cell in battery packs, and they cannot accurately diagnose faults and anomalies incurred by variation of other parameters, such as current, temperature and even power demand.

In short, the conventional fault diagnosis methods for lithium-ion battery packs, to the authors' knowledge, are inefficient for detecting the faults and abnormalities and locating faulty cells of battery packs. To address this issue, a systemic faults diagnosis method and a voltage abnormality detection approach are mainly investigated and developed for the battery pack based on the E-scooter monitoring and service platform. Through comprehensive analysis of operation data of the battery pack in E-scooters, we use the statistical technology to analyze the distribution characteristics of each parameter in battery packs and design the abnormal state detection coefficients. For the systemic fault diagnosis, by incorporating a data visualization technique, the proposed algorithm can classify the operation status of battery packs and effectively diagnose the potential faults. The operating state of E-scooters is classified and identified according to the current of battery packs and the speed of E-scooters. Since the parameters vary dramatically in different modes, such as running, charging and rest, an effective classification is employed to avoid the occurrence of misdiagnosis. The abnormal state detection coefficient is comprehensively designed according to the distribution characteristics of parameters' variation. The systematic faults of battery pack and possible abnormal state can be diagnosed by one coefficient. For the voltage abnormality, an accurate detection and location algorithm of the abnormal cell voltage are attained by combining the data analysis method and the visualization technique. Firstly, the faulty or abnormal battery cells' voltage is roughly identified and classified using the K-means clustering algorithm [33]. Secondly, the abnormal cell voltage is located based on the designed coefficient that is calculated according to the Z-score theory [34]. Finally, for the sake of diagnosing health status of cells in battery packs, the fault occurrence frequency of cells voltage is calculated based on the  $3\sigma$  screening strategy [30]. Through the analysis of fault frequency for cells voltage in time domain, it can contribute to prognostics and then provide early warning of potential faults. In summary, this research paper is a theoretical study with practical implications for online fault and abnormality diagnosis of lithium-ion batteries, and the main innovations and contributions of this study can be attributed to the following two aspects: 1) A diagnosis coefficient is designed based on the distribution characteristics of each parameter in battery packs. By applying

the designed coefficient, the systematic faults of battery pack and possible abnormal state can be timely diagnosed. 2) The t-SNE technique, The K-means clustering and Z-score methods are exploited to detect and accurately locate the abnormal cell voltage. In addition, the fault frequency of cells voltage in time domain is calculated based on the  $3\sigma$  screening strategy. According to the fault frequency, it can prognose the potential faults of cells in a timely manner.

The remainder of this study is structured as follows. The monitoring platform and the data sources of this work are briefly introduced in Section II. Section III gives the detailed derivation for the proposed fault diagnosis method. In Section IV, the results of fault diagnosis for battery packs are presented and discussed. Section V investigates the abnormal detection of cell voltage, and the conclusions are given in Section VI.

## II. THE MONITORING PLATFORM AND DATA ANALYSIS

The operation safety of battery systems is one of the main issues hindering application and market penetration of E-scooters and EVs. In addition to the built-in fault diagnosis system in BMS of battery packs, a real-time management platform that can monitor battery operation and provide decision-making reference for end-users and manufacturers is also a feasible solution, as shown in Fig. 1. The platform can not only acquire and diagnose real-time operating data, but also predict operation conditions of E-scooters by means of analyzing historical and current operation data. Note that the E-scooter investigated in this study is equipped with a battery pack with the rated voltage of 48 V, and its maximum speed is 25 Km/h. The whole operation data are transmitted to the monitoring platform between the BMS and central platform through wireless communication, and the original data are compiled via the communication protocol. In this paper, we mainly investigated the faults diagnosis of E-scooter's battery system, and the selected data in this paper include the E-scooter's speed, battery pack voltage, current, SOC, temperature and the voltage of cells. Through these data, we can evaluate the operation state of battery pack in E-scooters. The specification and dimension parameters of the battery pack are shown in Table I.

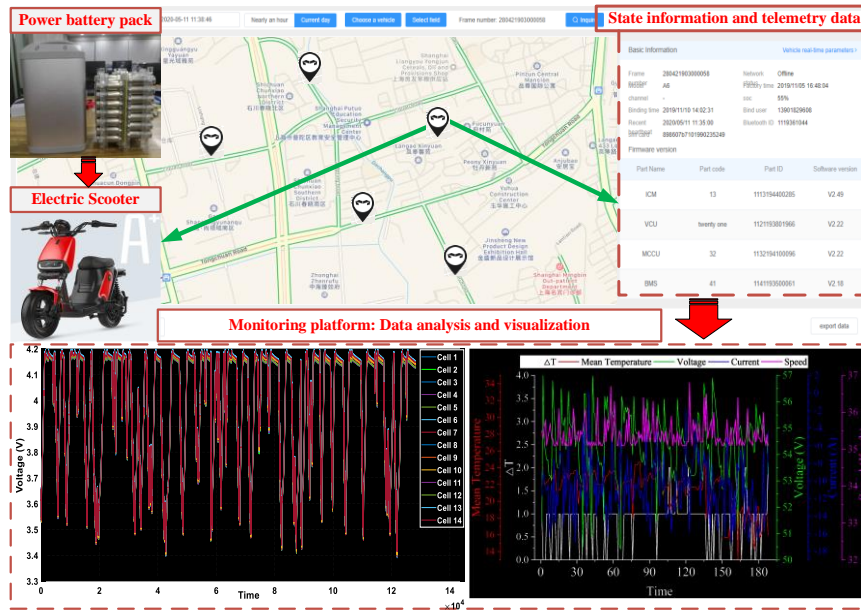


Fig. 1. The architecture of monitoring platform.

Table I. The specification parameters of the battery pack

Parameters	Value
Nominal Voltage (V)	48
Nominal Capacity (Ah)	24
Pack Weight (kg)	8±0.05
Size (mm)	145.5×178.5×283

To intuitively analyze the evolution characteristics of vehicle parameters, the monitoring curves of typical measurements in one month are shown in Fig. 2. Note that the negative current represents the discharging mode, and in contrast, the positive current indicates the charging mode. In addition, there exist three temperature sensors deployed to record the surface temperature of battery cells in a pack, allowing for a rough judgment of the consistency of battery pack based on the three different measurements. As can be seen from the partial figure of Fig. 2, when the speed is greater than zero and the current is less than zero, the driving power provided by the motor increases, and consequently the voltage and SOC decreases. However, a sudden change of one parameter may lead to a chain reaction. For instance, when the battery pack is being charged, an abnormal voltage signal may indicate over-voltage or under-voltage faults, even other parameters look normal. From this point of view, one can conclude that the fault type needs to be determined according to not only the immediate measure, but the variation range of different parameters. Table II shows the normal range of several parameters and the typical faults. It is noteworthy that we mainly focus on these electrical and thermal faults in this study.

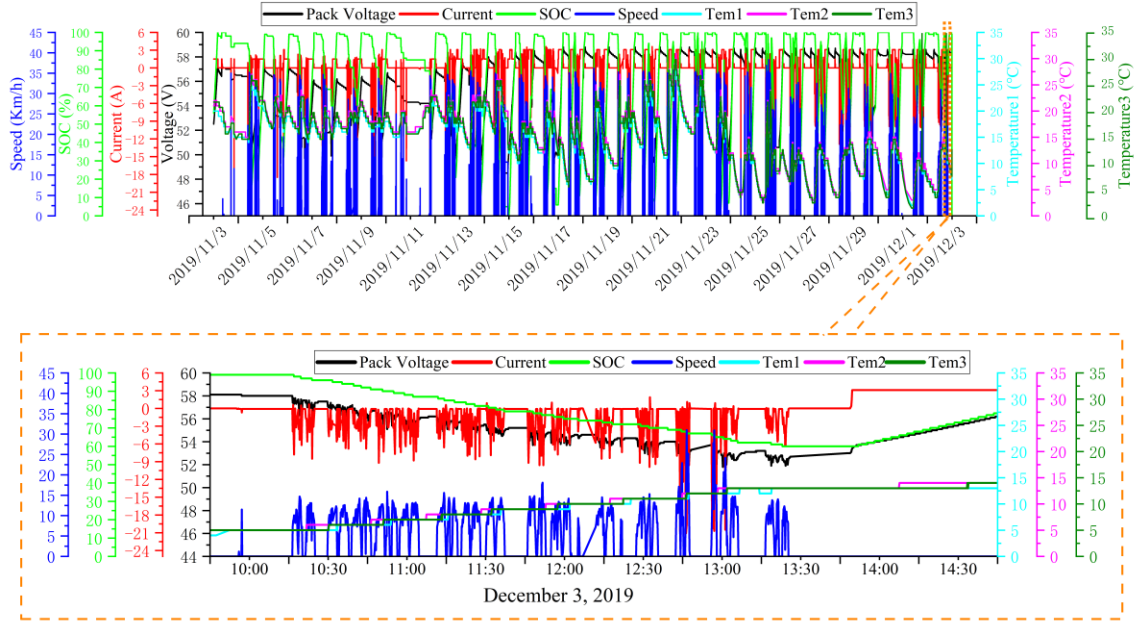


Fig. 2. The measurement of E-Scooters.

Table II. Normal range and possible failure of several typical parameters

Parameters	Normal range	Possible faults
Pack voltage (V)	[38.5, 58.8]	Over-voltage/Under-voltage
Speed (Km/h)	[0, 25]	Over-speed
Current (A)	Charge (0, 8)	Open circuit/Short circuit/Discharge over-current
	Discharge [-24, 0]	
Temperature (°C)	[-10, 50]	Temperature anomaly/Excessive temperature difference
Cell voltage (V)	[2.75, 4.20]	Over-voltage/Under-voltage/ Large voltage difference

In the next step, the diagnosis algorithms are devised based on the measurement variation and the characteristics of different faults.

### III. METHODOLOGY

In practical applications, failures seldom occur, and the fault data only account for a small portion of all the operation data. If the variation of measurement is only influenced by random errors, the statistical distribution is supposed to comply with the Gaussian distribution, which is consequently employed to describe the distribution characteristics of each measurement, as:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (1)$$

#### A. The Status Division of E-Scooter

From Fig. 2, the related parameters of the studied battery pack vary dramatically in different states and obviously are not in line with the Gaussian distribution. To cope with it, the E-scooter operating state needs to

be firstly identified and classified. Usually, the operation states of E-scooters can be divided into four categories, namely, discharging, charging, rest/dormancy and offline state, as listed in Table III. For simplicity, the current and speed are hereinafter referred to as  $C$  and  $S$ , respectively. The judgement conditions for cases 1 and 3 are explained as below. Intuitively, one of determination conditions for case 1 is  $C < 0$  and  $S = 0$ . When a sudden stop occurs, the motor control module continuously sends current request to the BMS; and in this case,  $S$  equals zero but the discharge operation continues.  $C = 0$  and  $S = 0$  can easily identify the operating status of case 3; and however, another condition of  $-1 < C < 0$  and  $S = 0$  can also refer to the occurrence of case 3, as when the E-Scooter is still (such as waiting for a traffic light), i.e.,  $S = 0$ ; small discharge current will be incurred due to the self-consumption of on-board control and driving units. Thus, the judgement condition can be reformulated to  $-1 < C \leq 0$  and  $S = 0$ . Actually, the parameters of a battery pack vary within a certain range case by case, and the statistical principle complies with the previously mentioned Gaussian distribution. The next step is to perform a mathematical modeling process for diagnosis and detection based on the Gaussian distribution characteristics.

Table III. The classification of states for E-Scooters

Classification	Description	Judgement conditions
Case 1	Discharge	$C < 0$ & $S \geq 0$
Case 2	Charge	$C > 0$ & $S = 0$
Case 3	Rest or dormancy	$-1 < C \leq 0$ & $S = 0$
Case 4	Offline	Else

### B. Mathematical Modeling of Systemic Faults Diagnosis Method

The statistical model of fault diagnosis and anomaly detection are established based on the statistical algorithms, of which the diagnostic criteria are designed with the consideration of variation of each parameter. The specific mathematical model is derived as follows. The operating data are acquired from a monitoring platform, including the battery pack voltage  $V$ , battery current  $C$ , SOC, vehicle speed  $S$  and temperature  $T$ . By calculating the mean value and the difference between the maximum and the minimum temperature, we have:

$$\begin{cases} T_t = [T_1^t & T_2^t & T_3^t] \\ \Delta T_t = \text{Max}(T_t) - \text{Min}(T_t) \\ \hat{T}_t = 1/3 \cdot (T_1^t + T_2^t + T_3^t) \end{cases} \quad (2)$$

Then, a matrix for E-scooter state can be yielded at time  $t$ , as:

$$D_t = [\Delta T_t \quad \hat{T}_t \quad V_t \quad C_t \quad S_t] \quad (3)$$

Next, a matrix of states over a period can be calculated by repeating (3):



$$D_n = [\Delta T_n \hat{T}_n V_n C_n S_n] \quad (4)$$

The matrix  $D_i^n$  can be divided into the matrices corresponding to the four operating states of E-scooter, as:

$$\begin{cases} D_1^n = [\Delta T_n \hat{T}_n V_n C_1^n S_1^n] \\ D_2^n = [\Delta T_n \hat{T}_n V_n C_2^n] \\ D_3^n = [\Delta T_n \hat{T}_n V_n C_3^n] \\ D_4^n = [\Delta T_n \hat{T}_n V_n] \end{cases} \quad (5)$$

Consequently, the mean value of the matrix corresponding to each state can be calculated, as:

$$\begin{cases} \bar{D}_i^j = \sum_{t=1}^n D_i^{jt}, \quad i=1,2,3,4, \quad t=1,2,\dots,n \\ \bar{D}_i = [\bar{D}_i^1 \quad \bar{D}_i^2 \quad \dots \quad \bar{D}_i^j] \end{cases} \quad (6)$$

where  $j$  represents the  $j$ th variable in the  $i$ th vector. Similarly, the standard deviation  $\sigma$  and variance  $\sigma^2$  of each numerical matrix can be yielded, as:

$$\begin{cases} \sigma_i = [\sigma_i^1 \quad \sigma_i^2 \quad \dots \quad \sigma_i^j] \\ \sigma_i^2 = [(\sigma_i^1)^2 \quad (\sigma_i^2)^2 \quad \dots \quad (\sigma_i^j)^2] \end{cases} \quad (7)$$

Then, the corresponding coefficient matrix for each operation state can be constructed, as:

$$f_t = \begin{bmatrix} \bar{D}_i^1 & \bar{D}_i^2 & \dots & \bar{D}_i^j \\ \sigma_i^1 & \sigma_i^2 & \dots & \sigma_i^j \\ (\sigma_i^1)^2 & (\sigma_i^2)^2 & \dots & (\sigma_i^j)^2 \end{bmatrix} \quad (8)$$

Next, the coefficient matrix of the  $t$ th operation state can be simplified, as:

$$f_t = \begin{bmatrix} \bar{D}_i \\ \sigma_i \\ \sigma_i^2 \end{bmatrix} \quad (9)$$

According to (1), the diagnostic criterion can be defined, as:

$$P_t = \prod_{j=1}^k \left[ \frac{1}{\sqrt{2\pi} f_{t2}^j} \cdot \exp\left(-\frac{(D_t^j - f_{t1}^j)^2}{2 \cdot f_{t3}^j}\right) \right] \quad (10)$$

where  $f_{t1}^j$ ,  $f_{t2}^j$  and  $f_{t3}^j$  represent the first, second and third vector of the coefficient matrix  $f_t$ . Based on (9), the coefficient matrix  $f$  can be calculated from a large number of normal operating data to determine  $P$  at each moment. The threshold  $\varepsilon$  of  $P$  needs to be set for fault diagnosis.

### C. The Method of Voltage Abnormality Detection for Cells

To ensure the operation security and reliability of the battery pack, it is imperative to monitor the operation

status and diagnose the state of health (SOH) for each battery cell. However, the diagnosis and analysis of each cell's voltage imposes significant computational burden and reduces the real-time performance of diagnosis. To address this issue, the high-dimensional monitoring data (each cell data is regarded as one-dimension) of all cells' voltage is compressed to two-dimension data using the t-distributed stochastic neighbor embedding (t-SNE) technique [35], which is an effective data visualization tool that produces better visualization by converting the tendency to crowd points in the center of map. The dimensionality reduction process based on the t-SNE can improve the readability of data and meanwhile reserve the crucial information contained in the original data. After the dimension reduction, the K-means clustering algorithm is exploited to perform the cluster analysis for screening the abnormal cells voltage in the battery pack. The abnormal cell is then located based on the detection criterion by combining the Gaussian distribution principle and the Z-score method. The K-means clustering algorithm and the Z-score method are employed to achieve the two-step fault diagnosis and location for cell voltage. The K-means clustering algorithm is an unsupervised division clustering segmentation method, which can be effective to eliminate singular sample points that are not consistent with the actual class labels [36]. The Z-score, referred to as the standard score, stems from the standard deviation, and features the function of risk prediction. Therefore, it is often employed as a measurement of discrete degree between individuals in a group [31]. The mathematical modeling of voltage anomaly diagnosis is presented as follows.

Similar as before, a diagnostic coefficient of voltage anomaly is designed based on the probability density function. According to (11), the probability density value for each cell voltage can be formulated, as:

$$P(i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(V_i - \mu)^2}{2\sigma^2}\right) \quad (11)$$

Due to the inconsistency of battery cells, the voltage measurement of cells is different most of time. In particular, the voltage fluctuation of fault cells may be the most or least in a certain period of time; and thus, the voltage probability density calculated from the standard deviation and mean values may not be able to accurately detect the voltage faults. Therefore, we calculate the voltage anomaly coefficient according to the Z-score principle, as:

$$Z = \frac{x - \mu}{\sigma} \quad (12)$$

where  $x$  is the target sample,  $\mu$  is the mean value of all samples, and  $\sigma$  is the standard deviation of

samples. The value of  $Z$  represents the distance between the target sample and the mean value of the original samples based on the standard deviation. When the target sample is lower than the mean value,  $Z$  is negative; otherwise, it is positive. Based on the  $Z$ -score method, the voltage anomaly coefficient is calculated, as:

$$Z_i(t) = \frac{\left| P_i(t) - \frac{1}{k} \sum_{i=1}^k P_i(t) \right|}{\sigma_p(t)} \quad (13)$$

where  $P_i(t)$  is the probability density value of  $i$ th cell voltage at time  $t$ ,  $k$  is the number of cells and  $k=14$  in this work, and  $\sigma_p(t)$  is the standard deviation of probability density for all the cells' voltage at time  $t$ . When the  $Z$  value of voltage for each cell is obtained, and it can be then used to determine which cell is abnormal. To analyze the characteristics and evolutions of fault states for each cell voltage in the battery pack, a fault diagnosis model is established in time domain. Generally, there are two reasons for the differences among cell voltages: systematic fault and random error. We assume that the voltage of cell is only affected by random error, and the diagnosis criteria are built by the Gaussian distribution to detect the fault cell voltage. Currently, the  $3\sigma$  screening strategy has been widely adopted based on the Gaussian distribution probability for fault-free data centralized screening [30], and for this reason, we continue to employ the  $3\sigma$  principle to screen the fault data in time domain. By calculating the mean value  $\mu$  and the standard deviation  $\sigma$  of cell voltage for all the sample data, the difference value between the cell voltage  $V_i^t$  and  $\mu$  can be obtained, as:

$$\Delta V_i^t = V_i^t - \mu \quad (14)$$

By comparing the absolute value of  $\Delta V_i^t$  with  $3\sigma$ , we have:

$$\begin{cases} \text{if } 3\sigma \leq |\Delta V_i^t| \leq 3\sigma, & F_{t,i} = 0 \\ \text{else,} & F_{t,i} = 1 \end{cases} \quad (15)$$

According to (15), the fault matrix at time  $t$  can be constructed, as:

$$F_t = (F_{t,1}, \dots, F_{t,n}) \quad (16)$$

where  $F_{t,i} = 0$  or 1. The fault matrix for a specific period of  $(t_0 \sim t_1)$  can be yielded, as:

$$F = (F_{t_0}^T, \dots, F_{t_1}^T)^T \quad (17)$$

Then, we can get:

$$FH_{day}^j = \frac{1}{n} \sum_{i=1}^n F_{day}^{i,j} \quad (18)$$

where  $i$  denotes the fault vector of the day  $i$ ;  $j$  is the number of cell in battery pack;  $FH_{day}^j$  is the fault frequency of  $j$ th cell for a day; and  $F_{day}^{i,j}$  is the fault state of  $j$ th cell at time  $i$ .

#### D. Algorithm Design

The parameters of the diagnosis model are identified by analyzing and training the off-line data. The monitoring platform can provide historical operation data, of which the frequency demanded by the monitoring platform is 0.1 Hz. Since the operation data are inevitably affected by noises, sensor errors, and BMS errors when acquiring and transmitting the measurement, the raw data exported from the monitoring platform include some inferior values, so that it cannot be analyzed directly by statistical methods. To cope with this issue, the original exported data are further processed to ensure that the data for modeling are valid. The framework and flow chart of faults diagnosis and abnormality detection is shown in Fig. 3. As can be seen, the whole framework includes three modules, namely the data cleaning and washing module, systemic faults diagnosis module and the abnormality detection module of cells voltage. In the systemic fault diagnosis module, the primary task is to properly process the raw data, then the operation status of E-Scooter needs to be identified, and thus the diagnosis accuracy can be improved. Finally, the model needs to be evaluated under different operating conditions to verify its feasibility, reliability and stability. In the abnormality detection module of cells voltage, the K-means clustering algorithm is firstly exploited to screen the abnormal voltage data. Then, the Z-score method is employed along with the Gaussian distribution to detect and locate the abnormal cells. Through the mathematical analysis and modeling based on the relevant historical operation data, the online fault diagnosis and abnormality detection scheme can be constructed to achieve the real-time status monitoring and fault diagnosis. During the analysis process of fault diagnosis modeling, one critical step is to determine the threshold  $\varepsilon$  of the fault diagnosis coefficient  $P$ , which typically requires a large amount of offline data for trial-and-error. When  $\varepsilon$  is determined,  $P$  can be calculated at each moment to determine the state of E-scooter.

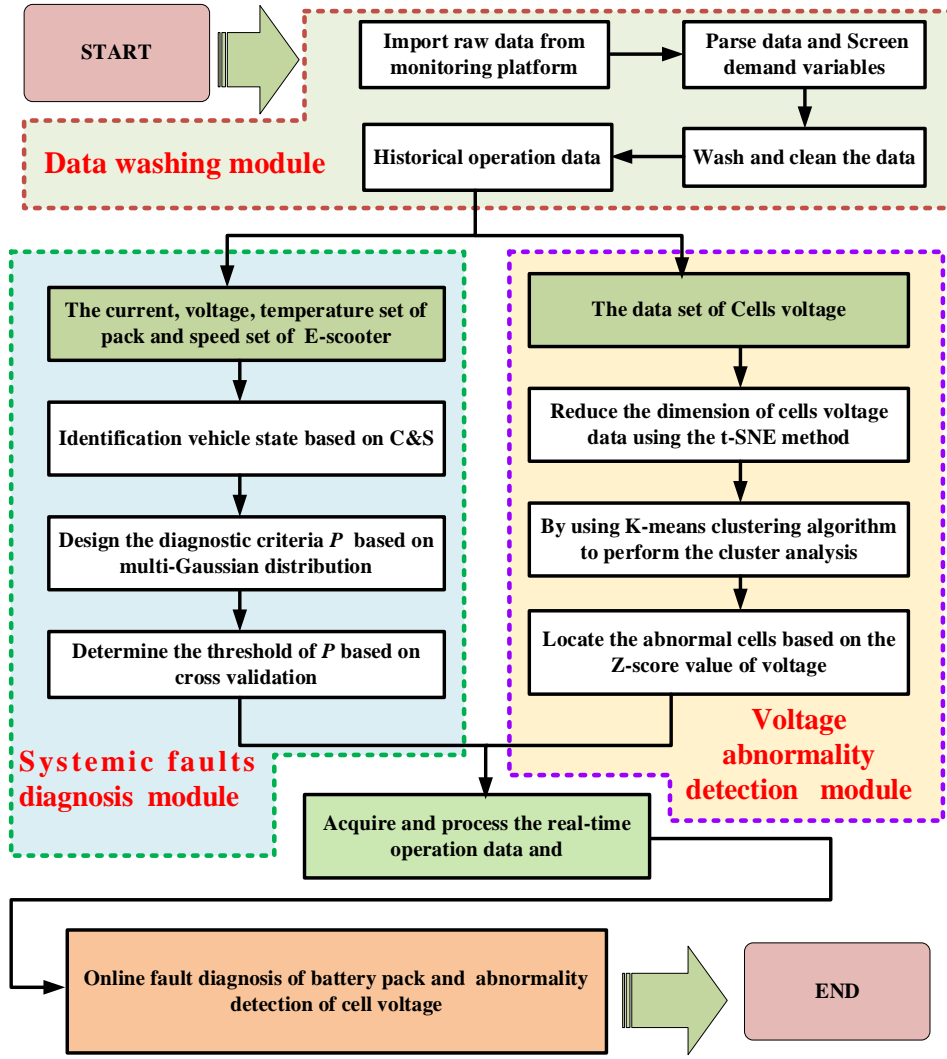


Fig. 3. The framework and flow chart of faults diagnosis and abnormality detection.

Next, a series of diagnosis results for systemic faults are presented, and the corresponding discussions are conducted.

#### IV. RESULTS AND DISCUSSION

For the systemic fault diagnosis, the operation status of E-scooter is firstly identified to improve the diagnosis accuracy. The diagnostic criterion  $P$  is then set to diagnose the systemic faults. Based on  $P$ , the faults caused by the voltage, current, temperature of battery pack and the speed of E-scooter can be diagnosed simultaneously. In this section, the threshold  $\varepsilon$  of  $P$  is firstly determined using the historical operation data of E-scooters.

##### A. Threshold Determination of the Diagnostic Coefficient

The coefficient matrix  $f$  in (9) is calculated using the processed data from 50 E-scooters after removing the abnormal values. Then, the data of other 30 E-scooters are divided into two parts: normal data

and abnormal data. The  $P$  corresponding to normal and abnormal part data is respectively calculated according to (10), and the threshold  $\varepsilon$  of  $P$  is determined through the comparative analysis. The calculated values of  $P$  for the normal part and abnormal part data are shown in Fig. 4. As can be seen from Fig. 4 (a),  $P$  varies dramatically. In this case, the coefficient corresponding to a single parameter is less than 1, it increases the change rate of  $P$  when multiplied by several coefficients. Furthermore, a slight change in one single coefficient may result in several orders of magnitude increase of  $P$ . Nonetheless, it can be seen from the enlarged view of Fig. 4 that  $P$  corresponding to the normal part of the data is greater than  $3 \times 10^{-5}$ , while that corresponding to the abnormal part of data is less than  $1 \times 10^{-5}$ . Hence, according to the calculation results, the threshold  $\varepsilon$  of  $P$  is set to  $1 \times 10^{-5}$  in this work. As  $P$  is sensitive to the abnormality of corresponding parameters,  $P$  may be far less than  $1 \times 10^{-5}$  when the fault occurs. Note that when  $P$  is less than  $1 \times 10^{-10}$ ,  $P$  is set to 0 for simplicity.

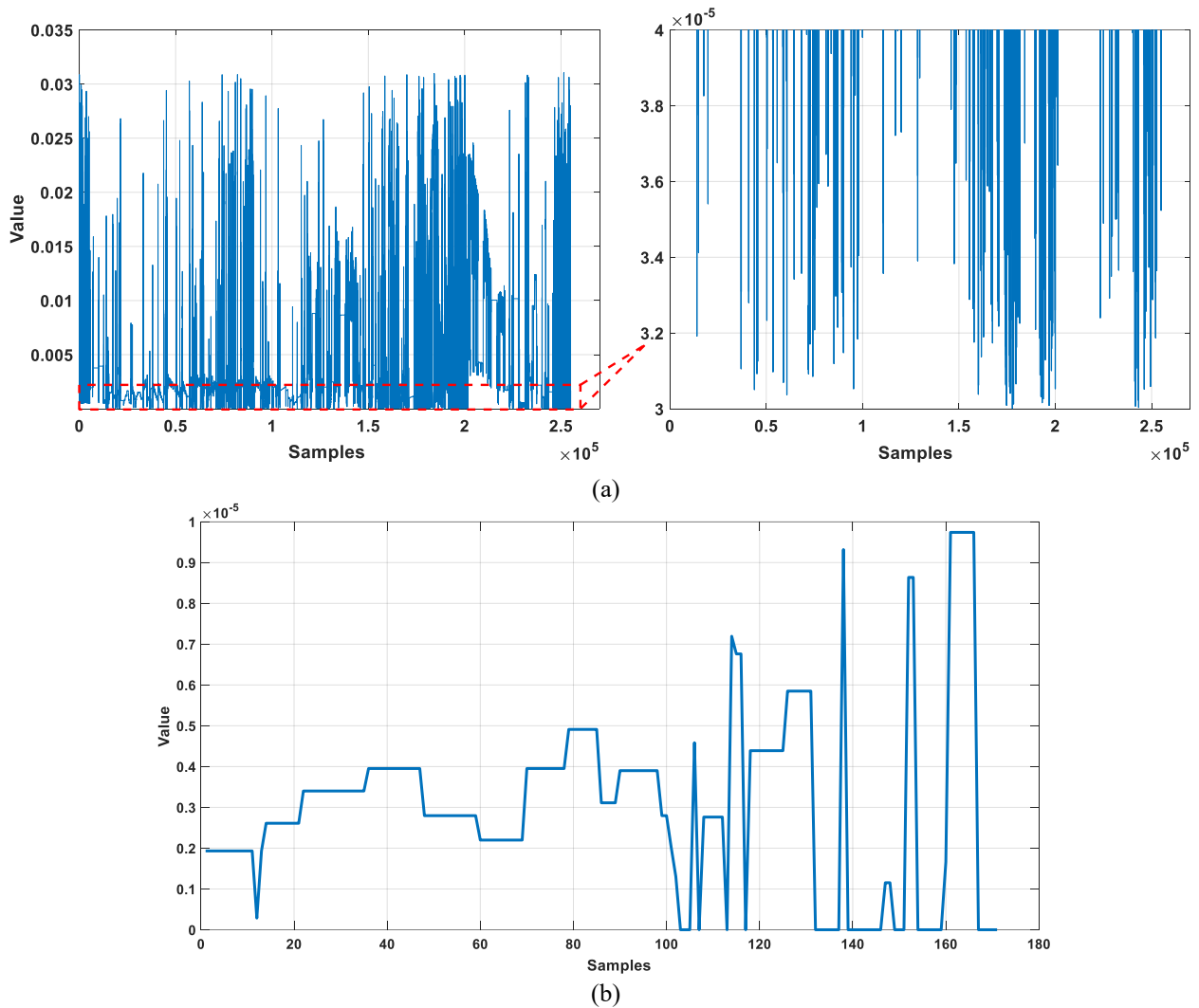


Fig. 4. The calculated results of  $P$  for different samples. (a) Normal data. (b) Abnormal data.

### B. Faults Diagnosis for Battery Pack

To verify the fault diagnosis ability of proposed method for battery packs, the partial random historical operation data of one E-scooter is considered for validation. Note that the E-scooter selected for validation is excluded from the data of 86 E-scooters for training. Based on the calculated results of  $P$ , it can be found that  $P$  is equal to 0 (less than  $1 \times 10^{-10}$ ) at three moments, and in this case the systematic state shows a malfunctional alarm. The calculated  $P$  and corresponding parameters are listed in Table IV, and the parameters curves are shown in Fig. 5. According to  $P$  and other parameters, it can be judged that excessive temperature difference fault, discharging over-current fault and abnormal temperature fault occur at 66921 s, 72338 s and 114251 s, respectively. It can be seen from Fig. 5 (a) that the abnormal drop of the second temperature value at 66921 s results in a temperature increase of 5 °C, and  $P$  is far less than  $1 \times 10^{-10}$  at this moment. The temperature difference among different locations in the battery pack can represent the consistency of cells to some extent. As can be obviously observed from Fig. 5 (b), the discharge current of the battery pack reaches 48.2 A at 72338 s, which is much more than the upper boundary (24 A) of the normal range listed in Table II. However, the corresponding speed of E-scooter is only 13.6 km/h at this moment, and it should be within the normal range [0, 25]. From initial analysis, the over-current fault may be raised by the extreme operation of E-scooter, such as emergency uphill. Fig. 5 (c) shows that the value of temperature rises sharply at 114251 s in the third time, and then quickly returns to the original value. Nevertheless, the voltage, current and SOC of battery pack at this time are normal according to the criteria shown in Table IV, and the speed of E-scooter is only 1.1 km/h, indicating that the abnormal temperature fault may be caused by the acquisition error of the temperature sensor. Therefore, the diagnosis results highlight that the proposed method can quickly and accurately diagnose the over-current fault, excessive temperature difference fault and sensor sampling fault caused by the parameters abnormality of battery pack.

Table IV. The value of corresponding parameters and  $P$  for faults state.

Time (s)	Voltage (V)	Current (A)	Speed (Km/h)	$\Delta T$ (°C)	$\hat{T}$ (°C)	$P$	System state
66921	54.4	-0.1	11.8	5	16	0	Fault Alarm
72338	49.7	-48.2	13.6	0	20	0	Fault Alarm
114251	51.3	0	1.1	44	26.67	0	Fault Alarm

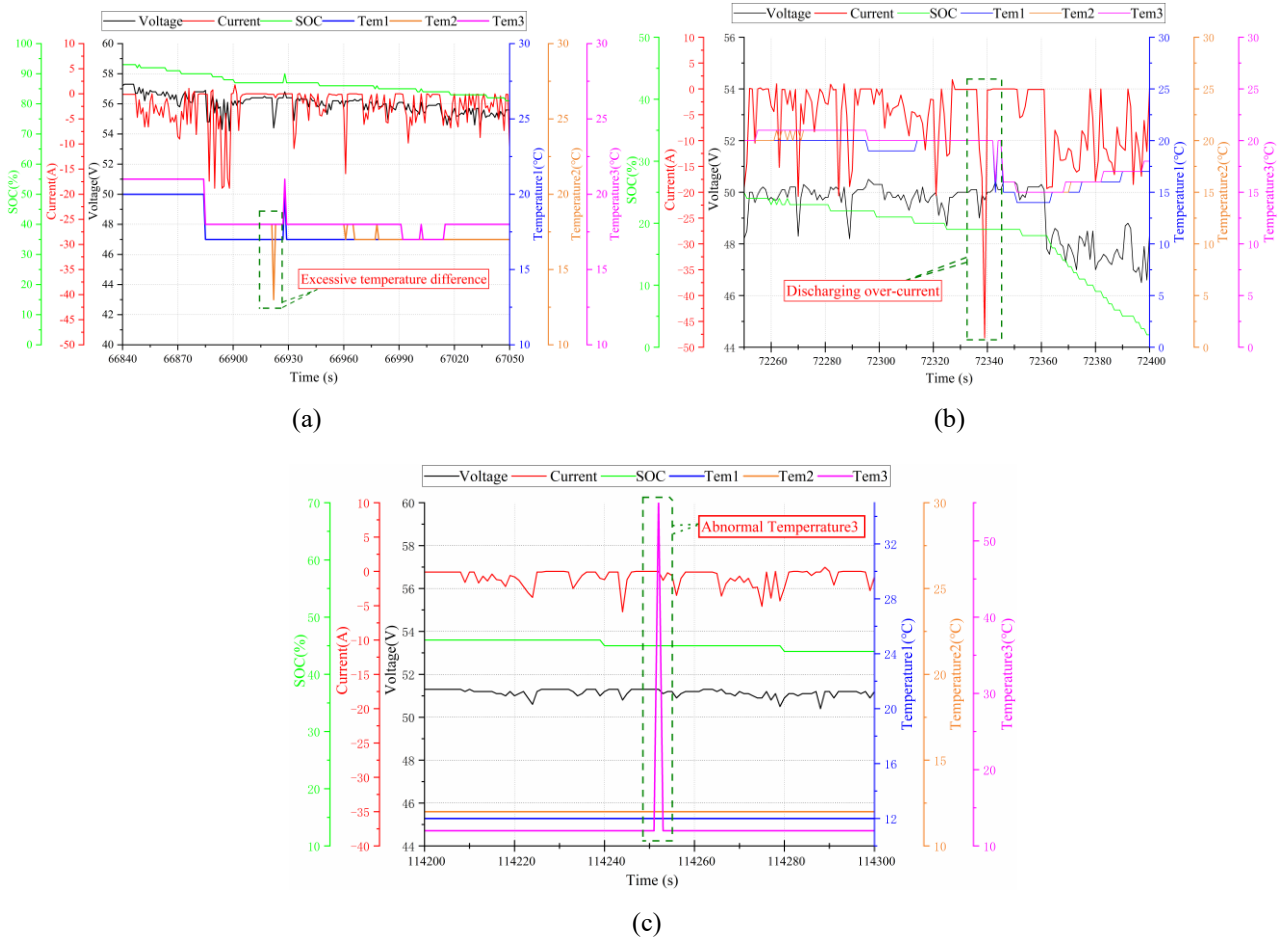


Fig. 5. Partial faults data of the battery pack.

### C. The Status Monitoring of System in Real Time

Based on the operation data recorded by the E-scooter in real time, the system status can be determined by calculating  $P$  at each moment. Fig. 6 (a) shows the variation of  $P$  for an E-scooter that is selected randomly within a certain period, and Fig. 6 (b) to (d) depicts the variation curves of different parameters including battery pack's voltage, current, SOC, temperature and the speed of E-scooter during this period. It can be observed from Fig. 6 (a) that  $P$  is less than  $1 \times 10^{-5}$  occasionally but more than  $1 \times 10^{-6}$  in all the remaining time, indicating that the E-Scooter operates normally, and no fault occurs during this period. It can be seen from Fig. 6 (b) that the voltage, current and SOC are all within the normal range. As shown in Fig. 6 (c), the three temperature values are equal most of the time during this period and the temperature difference reaches 2 °C occasionally. As can be found from Table II, the speed does not exceed 25km/h, and it is obvious from Fig. 6 (d) that the speed is slightly above 25 km/h for some timeslots. Although the temperature difference and the E-scooter speed are only slightly beyond the normal range, the calculated  $P$  varies significantly and occasionally becomes less than the preset threshold. Since the probability density values are less than 1,  $P$



will become smaller when several numbers are multiplied. Hence, as long as one parameter is slightly abnormal, the diagnosis coefficient  $P$  will be much smaller or even decrease by several orders of magnitude. Therefore, based on the designed fault diagnosis coefficient, the proposed method can not only effectively diagnose the system fault, but also detect the abnormality of the battery pack in real time.

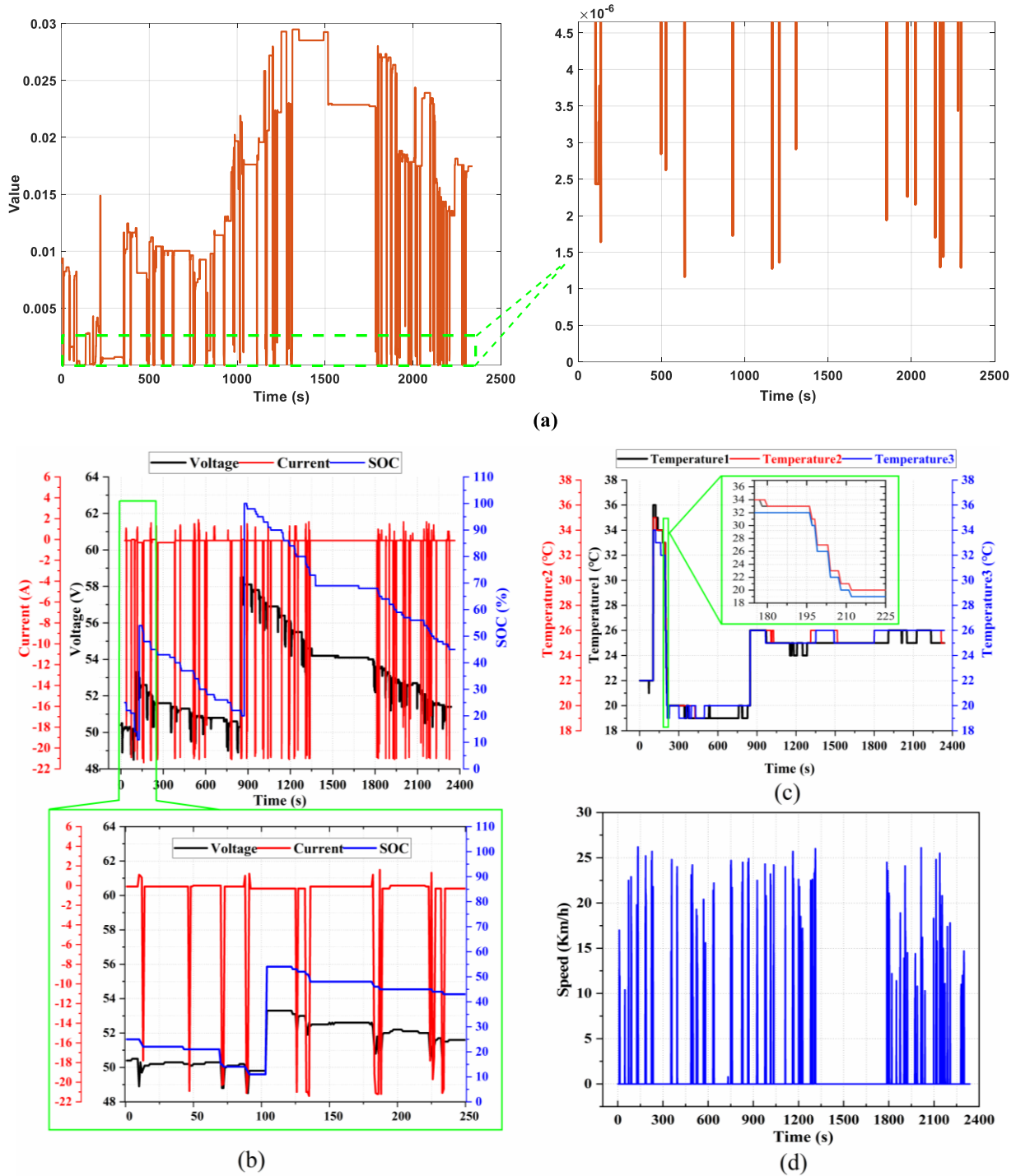


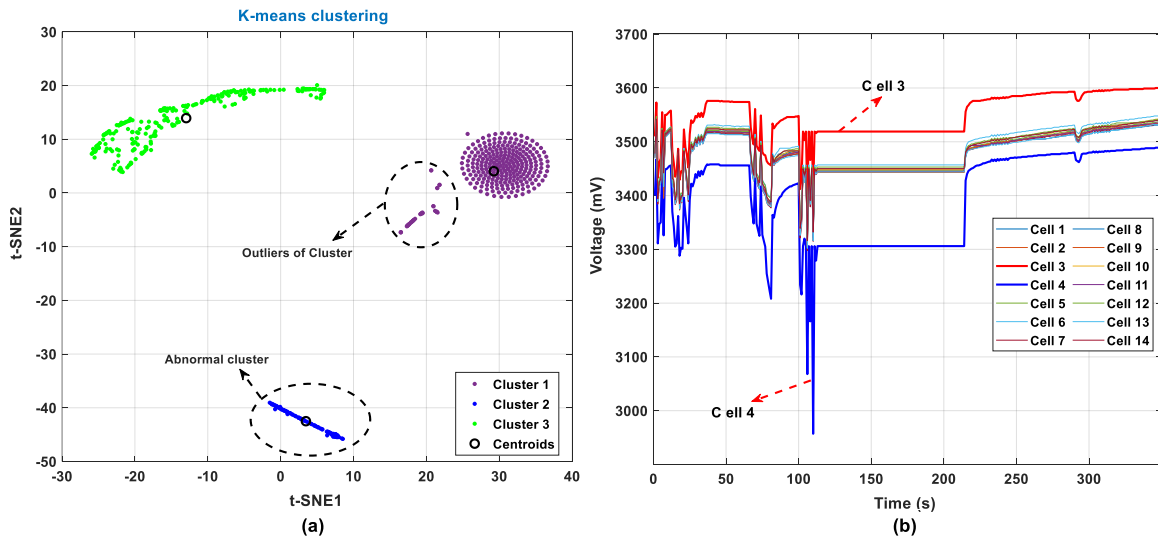
Fig. 6. The diagnosis results and variation curves of corresponding parameters for battery pack in real time. (a) The value of diagnosis coefficient  $P$  for battery pack; (b) The variation curves of voltage, current and SOC; (c) The temperature variation curves; (d) The speed of E-scooter at each moment.

## V. FAULTS DIAGNOSIS AND PROGNOSIS OF CELL VOLTAGE

As discussed above, the faults diagnosis and abnormality of battery pack can be detected in real time. In addition, timely detection and positioning of faults and defects of cells can improve the health and safety of the whole battery pack.

### A. Abnormal Voltage Diagnosis and Location

In this study, the voltage data of 14 cells in a battery pack is firstly applied to conduct the dimensionality reduction based on the t-SNE, and the result of dimensionality reduction is then analyzed according to the K-means clustering algorithm. The corresponding clustering result is shown in Fig. 7 (a). As can be seen, the compressed data are divided into several clusters according to their similarity and distance. In the same cluster, a fraction of outliers locating far away from the clustering center highlight some probable abnormal/fault tendency. This is likely to be caused by the abnormal data, such as those in the dotted circle of cluster 1. There also exist a small portion of independently clustered data, such as cluster 2 in Fig. 7 (a), which differ clearly from other clusters. Typically, this type of cluster is caused by the fault data. In general, the data visualization and clustering analysis contribute to visualization of abnormal cells by classifying the abnormal cells data. Then, the  $Z$  value calculated by (13) for each cell is employed to identify the abnormal voltage of cell. Next, the calculation results obtained from the real voltage data of cells are illustrated in detail.



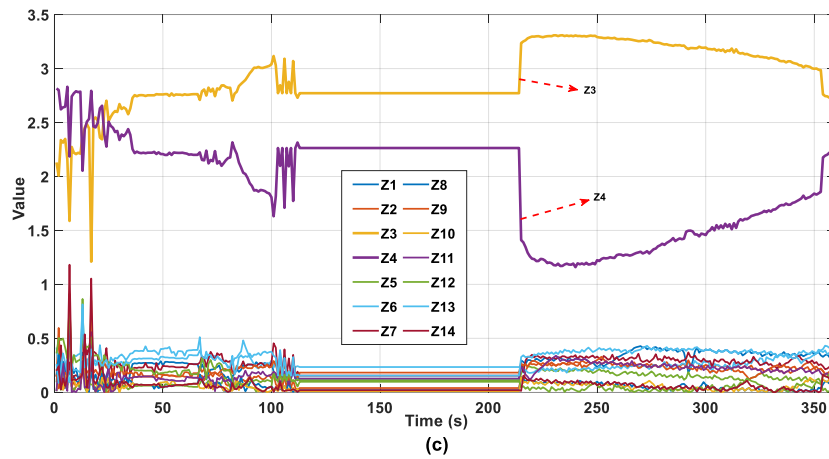


Fig. 7. The diagnosis results and voltages of a battery pack cells. (a) The results of K-means Clustering. (b) The voltage curves of all cells. (c) The values of  $Z$  for all cells.

To illustrate the diagnosis performance and evaluate the accuracy of abnormality location of the proposed method, all the cells' voltage in a short period (3600 s) is considered as an example, and the corresponding voltage variation curves and diagnosis results are shown in Fig. 7 (b) and (c), respectively. As can be found, the voltage of cells 3 and 4 is maximum and minimum in this period, deviating significantly from the other voltage curves. In particular, the voltage curve of cell 4 fluctuates obviously, and the voltage drops rapidly at 81 s and 110 s, reaching the maximum difference of 453 mV. The voltage abnormality of cell 4 results in the degraded electrical performance and leads to the fault of excessive voltage difference during the discharging stage. As shown in Fig. 7 (c), the  $Z$  values of cells 3 and 4 are much larger than those of other cells and fluctuate more significantly, manifesting that the proposed voltage anomaly coefficient can accurately diagnose the voltage anomaly and locate the fault cells. Therefore, the proposed method for voltage fault diagnosis can detect the aberrant battery cell accurately in a timely manner, thereby enabling great significance to prognosis and safety management of future battery failures.

#### B. Prognosis of Voltage Faults in Time Domain

In this study, a large amount of voltage data are analyzed based on the Gaussian distribution. For the sake of intuitive analysis of fault diagnosis results and potential faults, the fault frequency is investigated to describe the health status of cells. We need to respectively count the normal and abnormal amount of data for the cell voltage over a period. Taking the data of a battery pack within one month as an example, the fault frequency matrix is calculated according to (14) to (18), and the fault frequency distribution is shown in Fig. 8 (a). As can be seen, on the 12th day, except for cells 3 and 6, the voltage measurements of other cells are abnormal for a short period, which may be caused by the improper operating behavior of E-scooter. To further analyze

the cause of this fault, the corresponding cells voltage data and the voltage difference value  $\Delta V$  calculated based on (14) are shown in Fig. 8 (b) and (c), respectively. It can be observed from Fig. 8 (b) that the cell voltage consecutively declines and is less than 3.4 V during the period of 5560 s to 6620 s, indicating the battery pack is continuously discharged. In particular, during the short time period of 6600 s to 6610 s, although all the cells are already in a low voltage state, the battery pack is still discharged with a high rate current at this time, thus leading to a rapid drop in voltage. Through corresponding calculation, the  $3\sigma$  value of all cells historical voltage is 630.02, and at 6607 s on the 12th day, except the  $\Delta V$  of Cell 3 and Cell 6 are less than the  $3\sigma$ , the  $\Delta V$  of other cells is greater than 630.02. From Fig. 8 (b), we can also find that the voltage of all the cells drops sharply at 6607 s, and then quickly returns to the original value. From this point of view, the experimental results highlight that the proposed method can accurately detect the voltage abnormality. From the detection results and the voltage variation trajectories of cells, it can be concluded that the detected abnormality is a rapid descent of voltage caused by the battery pack that is discharged with a high rate current in a low voltage stage. This kind of improper behavior that the E-scooter operates in the extreme environment may trigger the over-current and under-voltage faults of cell. As can be seen from Fig. 8 (b), the voltage of Cell 4 is minimum during both the charging and discharging process, and it even drops to 2.942V at 6607 s, which is far lower than that of other cells. The voltage of Cell 4 obviously deviates from the normal voltage, and the extreme operation of E-scooter generates a significant impact on Cell 4 at 6607 s. Thus, we can summarize that the abnormality of Cell 4 is mainly caused by the deterioration of its own electrical performance. To sum up, it is reasonable that we attribute the detection fault of most of cells on 12th day to the improper scooter operation and ascribe the abnormality of Cell 4 to the deterioration of its electrical performance. By means of the discussion, we can conclude that the proposed statistical method can calculate the fault frequency by constructing the abnormality detection model of the cell voltage in time domain, and then it can be employed to further analyze the reason of the faults. The fault frequency of cell voltage can also be used for quantitatively evaluating the cells' health status in battery packs. Thus, the proposed method is of essential contribution to safety and health management of battery system for E-scooters.

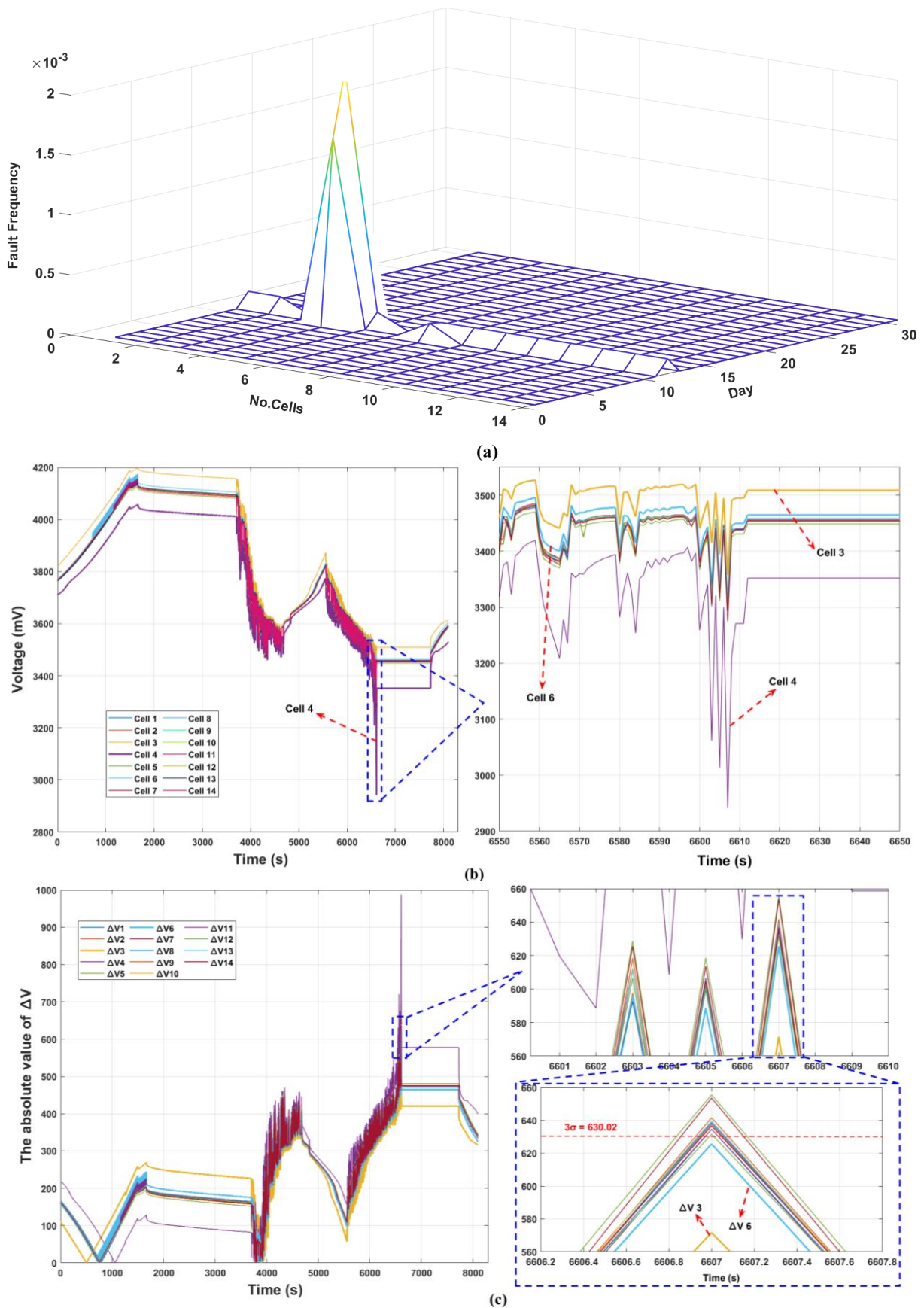


Fig. 8. The frequency distribution of voltage fault and partial diagnosis results of voltage abnormality. (a) Fault frequency distribution of cells voltage in a month; (b) The voltage curves of all cells on the 12th day; (c) The values of  $\Delta V$  for all cells on the 12th day.

## VI. CONCLUSION

This paper presents a statistical method for fault diagnosis and abnormality detection of battery systems of electric scooters based on the data collected from the central monitoring platform. The work is conducted primarily in the whole system troubleshooting and internal cell voltage anomaly detection. For the faults diagnosis of the whole system, firstly, the electric scooter operation state is divided into four cases according to the battery current and speed. Then, the probability density of data for each moment is calculated by taking the advantage of the characteristic that each parameter variation is in line with the Gaussian distribution. Next, the probability density values corresponding to all parameters are multiplied as the fault diagnosis coefficient. The validation results manifest that the proposed method can accurately diagnose the system faults and abnormalities and can be applied to the real-time safety management of electric scooters. In addition, the combination of data analysis and visualization methods enable timely and accurate detection of the cell voltage abnormality. Furthermore, the fault frequency analysis in time dimension provides further insight into the occurrence time and location of abnormality.

Since the designed diagnosis system does not require additional equipment and energy consumption, it has been applied to the fault diagnosis and state monitoring of E-scooters. In the next step, we will continue to optimize the proposed algorithm and develop advanced methods for fault diagnosis to reduce the computation burden imposed on the host monitoring platform. In addition, although this work mainly focuses on small battery packs, the proposed method is also applicable to fault diagnosis of high-power battery system, such as those in electric vehicles. Therefore, the method of combining data mining with machine learning will be further applied to security management of complex dynamic systems due to its great application perspective.

## ACKNOWLEDGMENTS

This work was supported in part by the National Key R&D Program of China (No. 2018YFB0104000), in part by the National Natural Science Foundation of China (No. 61763021 and 51775063), and in part by the EU-funded Marie Skłodowska-Curie Individual Fellowships Project under Grant 845102-HOEMEV-H2020-MSCA-IF-2018. In addition, the authors would like to thank the Shenzhen Mamotor Technology Ltd. for their hardware and program training support. Moreover, and most importantly, the authors would like to express deep gratitude to the anonymous reviewers for their corrections and helpful suggestions.

## REFERENCES

- [1] Y. Liu, J. Li, Z. Chen, D. Qin, and Y. Zhang, "Research on a multi-objective hierarchical prediction energy management strategy for range extended fuel cell vehicles," *Journal of Power Sources*, vol. 429, pp. 55-66, 2019.
- [2] X. Shu, G. Li, J. Shen, W. Yan, Z. Chen, and Y. Liu, "An adaptive fusion estimation algorithm for state of charge of lithium-ion batteries considering wide operating temperature and degradation," *Journal of Power Sources*, vol. 462, 2020.
- [3] P. Wells and X. Lin, "Spontaneous emergence versus technology management in sustainable mobility transitions: Electric bicycles in China," *Transportation Research Part A: Policy and Practice*, vol. 78, pp. 371-383, 2015.
- [4] Y. Shang, N. Cui, B. Duan, and C. Zhang, "Analysis and Optimization of Star-Structured Switched-Capacitor Equalizers for Series-Connected Battery Strings," *IEEE Transactions on Power Electronics*, vol. 33, pp. 9631 - 9646, 2017
- [5] L. Lu, X. Han, J. Li, J. Hua, and M. Ouyang, "A review on the key issues for lithium-ion battery management in electric vehicles," *Journal of Power Sources*, vol. 226, pp. 272-288, 2013.
- [6] Y. Wang, J. Tian, Z. Chen, and X. Liu, "Model based insulation fault diagnosis for lithium-ion battery pack in electric vehicles," *Measurement*, vol. 131, pp. 443-451, 2019.
- [7] Y. Wu, Q. Xue, J. Shen, Z. Lei, Z. Chen, and Y. Liu, "State of Health Estimation for Lithium-Ion Batteries Based on Healthy Features and Long Short-Term Memory," *IEEE Access*, vol. 8, pp. 28533-28547, 2020.
- [8] Z. Chen, Q. Xue, R. Xiao, Y. Liu, and J. Shen, "State of Health Estimation for Lithium-ion Batteries Based on Fusion of Autoregressive Moving Average Model and Elman Neural Network," *IEEE Access*, vol. 7, pp. 102662-102678, 2019.
- [9] X. Kong et al., "Signal synchronization for massive data storage in modular battery management system with controller area network," *Applied Energy*, vol. 197, pp. 52-62, 2017.
- [10] C. Wu, C. Zhu, Y. Ge, and Y. Zhao, "A review on fault mechanism and diagnosis approach for Li-ion batteries," *Journal of Nanomaterials*, vol. 2015, pp. 1-9, 2015.
- [11] D. Ren, X. Feng, L. Lu, X. He, and M. Ouyang, "Overcharge behaviors and failure mechanism of lithium-ion batteries under different test conditions," *Applied Energy*, vol. 250, pp. 323-332, 2019.
- [12] Z. Chen, R. Xiong, J. Lu, and X. Li, "Temperature rise prediction of lithium-ion battery suffering external short circuit for all-climate electric vehicles application," *Applied Energy*, vol. 213, pp. 375-383, 2018.
- [13] Z. Liu and H. He, "Sensor fault detection and isolation for a lithium-ion battery pack in electric vehicles using adaptive extended Kalman filter," *Applied Energy*, vol. 185, pp. 2033-2044, 2017.
- [14] M. Ma, Y. Wang, Q. Duan, T. Wu, J. Sun, and Q. Wang, "Fault detection of the connection of lithium-ion power batteries in series for electric vehicles based on statistical analysis," *Energy*, vol. 164, pp. 745-756, 2018.
- [15] Y. Kang, B. Duan, Z. Zhou, Y. Shang, and C. Zhang, "A multi-fault diagnostic method based on an interleaved voltage measurement topology for series connected battery packs," *Journal of Power Sources*, vol. 417, pp. 132-144, 2019.
- [16] S. K. Ibrahim, A. Ahmed, M. A. E. Zeidan, and I. E. Ziedan, "Machine Learning Techniques for Satellite Fault Diagnosis," *Ain Shams Eng. J.*, 2019.
- [17] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3757-3767, 2015.
- [18] C. Zhiwei Gao, Carlo, Steven X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part II: Fault diagnosis with knowledge-based and hybrid/active approaches," *IEEE Transactions on Industrial Electronics*, vol. 62, 2015.
- [19] Z. Chen, R. Xiong, J. Tian, X. Shang, and J. Lu, "Model-based fault diagnosis approach on external short circuit of lithium-ion battery used in electric vehicles," *Applied Energy*, vol. 184, pp. 365-374, 2016.
- [20] A. Sidhu, A. Izadian, and S. Anwar, "Adaptive Nonlinear Model-Based Fault Diagnosis of Li-Ion Batteries," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 2, pp. 1002-1011, 2015.
- [21] 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.
- [22] Y. Kang, B. Duan, Z. Zhou, Y. Shang, and C. Zhang, "Online multi-fault detection and diagnosis for battery packs in electric vehicles," *Applied Energy*, vol. 259, p. 114170, 2020.
- [23] Y. Shang, G. Lu, Y. Kang, Z. Zhou, B. Duan, and C. Zhang, "A multi-fault diagnosis method based on modified Sample Entropy for lithium-ion battery strings," *Journal of Power Sources*, vol. 446, p. 227275, 2020.
- [24] 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.
- [25] S. Abada, G. Marlair, A. Lecocq, M. Petit, V. Sauvant-Moynot, and F. Huet, "Safety focused modeling of lithium-ion batteries: A review," *Journal of Power Sources*, vol. 306, pp. 178-192, 2016.
- [26] 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.

- [27] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137-3147, 2016.
- [28] S. Li, H. He, and J. Li, "Big data driven lithium-ion battery modeling method based on SDAE-ELM algorithm and data pre-processing technology," *Applied energy*, vol. 242, pp. 1259-1273, 2019.
- [29] 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, p. 113381, 2019.
- [30] 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.
- [31] 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.
- [32] X. Li and Z. Wang, "A novel fault diagnosis method for lithium-Ion battery packs of electric vehicles," *Measurement*, vol. 116, pp. 402-411, 2018.
- [33] J. L. Xu, J. W. Han, F. P. Nie, and X. L. Li, "Re-Weighted Discriminatively Embedded K-Means for Multi-View Clustering," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 3016-3027, Jun 2017.
- [34] S. S. Kim, M. L. Aprahamian, and S. Lindert, "Improving inverse docking target identification with Z-score selection," *Chemical Biology & Drug Design*, vol. 93, no. 6, pp. 1105-1116, Jun 2019.
- [35] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of machine learning research*, vol. 9, pp. 2579-2605, 2008.
- [36] J. Liu, Q. Li, W. Chen, and T. Cao, "A discrete hidden Markov model fault diagnosis strategy based on K-means clustering dedicated to PEM fuel cell systems of tramways," *International Journal of Hydrogen Energy*, vol. 43, no. 27, pp. 12428-12441, 2018.