

Received January 18, 2022, accepted January 27, 2022, date of publication January 31, 2022, date of current version February 10, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3147802

Fault Identification of Lithium-Ion Battery Pack for Electric Vehicle Based on GA Optimized ELM Neural Network

LEI YAO¹, SHIMING XU¹, YANQIU XIAO¹, JUNJIAN HOU¹, XIAOYUN GONG¹, ZHIJUN FU¹, (Member, IEEE), AND AIHUA TANG², (Member, IEEE)

¹Henan Engineering Research Center of New Energy Vehicle Lightweight Design and Manufacturing, Zhengzhou University of Light Industry, Zhengzhou 450000, China

²Vehicle Engineering Institute, Chongqing University of Technology, Chongqing 400054, China

Corresponding author: Yanqiu Xiao (y104410003@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 62073298; in part by the Department of Science and Technology of Henan Province under Grant 212102210008, Grant 202102210289, and Grant 212102210237; in part by Zhengzhou Science and Technology Bureau under Grant 2020CXZX0046; and in part by the Scientific Research Foundation of Chongqing University of Technology.

ABSTRACT The battery system is one of the core technologies of the new energy electric vehicle, so the frequent occurrence of safety accidents seriously limits the large-scale promotion and application. An innovative extreme learning machine optimized by genetic algorithm (GA-ELM)-based method is proposed to estimate the current system status, which improves the accuracy and timeliness of fault identification. It is feasible in the application of electric vehicles. To ensure the effectiveness of the signal, the proposed method is adopted using the simple mean filter to clean the data with eliminate wrong points. After the variance analysis, covariance, a horizontal variance of the filtered data, a modified feature parameters matrix is presented. The dimension is reduced by principal component analysis to improve the engineering application ability. Furthermore, a comprehensive GA-ELM-based identification method is proposed to reduce the resulting identification error of extreme learning machines due to the initial value change. More importantly, the sensitivity and accuracy of different solutions are compared and verified, which shows the technique has great potential in battery fault diagnosis based on the voltage signal.


INDEX TERMS Lithium-ion battery, fault diagnosis, extreme learning machines, feature parameters.

I. INTRODUCTION

With the improvement of human living environment requirements, countries worldwide have paid great attention to the deterioration of the global environment and climate warming. The zero-emission electric vehicle industry has become an essential field of competition and development among countries, and battery-driven electric vehicles and hybrid electric vehicles have entered a period of rapid evolution [1]–[4]. As the energy storage carrier of electric cars, battery safety, and reliability significantly affect the performance of electric vehicles [5]. Compared with other batteries, lithium-ion power battery has been widely used in the electric vehicle industry due to the advantages of wide working temperature range, high energy density, high power density, and low

self-discharge rate [6]. However, many electric vehicle fire incidents have been caused mainly by battery failure [7]–[9]. The batteries themselves are a highly complex system with interdependent features. A feasible solution is urgently needed to achieve the reliable classification of characteristic battery parameters [10] to diagnose battery faults accurately. Batteries are primarily in complex turbulence, vibration, and impact conditions. Lithium batteries may have a virtual connection and short circuits without detecting and diagnosing their fault state, resulting in hidden safety risks [11]. Therefore, good fault detection and management system in engineering applications are essential [12]–[14]. Battery Management System (BMS) usually collects relevant data information for the battery system through sensors and then carries out state identification and fault diagnosis [15]–[17].

The BMS is critical in the whole operation of electric vehicles, which can respond to any battery fault with the

The associate editor coordinating the review of this manuscript and approving it for publication was Jon Atli Benediktsson .

fastest speed, determine the fault location and cause, and give reasonable treatment methods [18]. Failures that the BMS cannot detect cause safety problems in electric vehicles. Common battery faults can be divided into four categories: overcharging [19], excessive discharge [20], internal short circuits [21], and external connection faults [22]. Excessive charge and discharge, internal short courses, or connection failure can lead to abnormal voltage fluctuations and overheating. Excessive charge and discharge can be avoided by monitoring parameters such as voltage, while external connection faults are the most serious, which will lead to a sharp rise in battery temperature and induce thermal runaway [23]. Many scholars put forward two methods based on temperature and voltage to avoid related faults according to different battery faults.

Because of the closed structure of the lithium battery, the internal temperature changes with the operation of the lithium battery. The research shows that the battery short circuit fault is crucial for thermal runaway and ignition. The early detection of the battery short circuit has become an essential task of battery management. When the temperature difference between individual cells increases by 5° , the battery capacity will decrease by 1.5-2%. The uneven temperature will also reduce the accuracy of battery pack fault diagnosis [24]. The temperature sensor is generally uniformly arranged on the surface of the battery based on the temperature method. The battery operation can be analyzed in real-time according to the collected temperature data [25]–[27]. Literature [28] proposed that the capacity and internal resistance differences were calculated to represent the abnormality of capacity and temperature by combining temperature and voltage parameters. In Literature [29], thermal power generation was taken as one of the fault indicators. The fault diagnosis of lithium battery was carried out using the electrochemical-thermal coupling model, which provided a new idea for the online detection algorithm. Although the temperature measured on the battery surface can infer the current operating state of the battery, the temperature has the problem of lag reaction. Other researchers compared the preset threshold value with voltage, current, and temperature rise rate [30] to identify the external connection faults of the battery. Because surface temperature is a nonlinear system determined by power, internal resistance, sensor position, and other influencing factors, fault diagnosis based on temperature alone may lead to misjudgment.

On the other hand, the battery voltage value is easier to be accurately collected than the temperature and can better reflect the current operating state of the lithium battery. A particular voltage signal is organized based on the voltage method to diagnose and locate battery faults, evaluate battery status [31]–[33]. Xiong *et al.* [34] proposed to combine the least square way and unscented Kalman filter to determine the error between the estimated value and the actual value of the cell SOC and identify the fault location by comparing with the set threshold. In reference [35], a multi-scale thermoelectric-chemical model was adopted to detect

excessive charging-discharge faults combined with terminal voltage signals generated by extended Kalman filtering. Yang *et al.* [36] proposed a diagnostic algorithm based on the first-order equivalent circuit model and random forest algorithm combined with the root mean square error of voltage, which can effectively predict the electrolyte leakage problem of lithium batteries. Liu *et al.* [37] proposed an integrated learning method for battery fault diagnosis based on Ruboost and reformulated three indicators to characterize essential electrode qualities. Their test was carried out using capacity information from lithium titanate and lithium iron phosphate batteries. This model can deal with the fault problem well. Yu *et al.* [38] proposed using the ammeter method to correct SOC and compared and studied the main features of various open-circuit voltage models in an application, which provided a reference for the direction of fault diagnosis of lithium batteries based on voltage method.

Both temperature-based and voltage-based fault diagnosis methods are data-driven by collecting many original data signals, extracting characteristic parameters for data analysis to achieve the purpose of fault diagnosis. With the progress of artificial intelligence technology, data-driven methods have achieved rapid development [39], [40], and commonly used techniques can be divided into traditional machine learning and deep learning. Sbarufatti *et al.* [41] proposed an adaptive fault diagnosis model based on radial basis function neural network (RBFNN). In this method, the particle filter (PF) is used to identify parameters, and the terminal voltage is measured in real-time to provide fault diagnosis for a lithium battery. The method has good accuracy. However, the data set used in the modeling process is measured in a specific environment, different from the data generated by the actual vehicle operation or the data generated by a test standard of electric vehicle operating conditions. Therefore, the application ability of this method in actual working conditions remains to be verified. Zhao *et al.* [42] introduced a fault diagnosis method driven by a 3σ multi-level screening strategy and machine learning using actual vehicle operation data for several months. This method uses a neural network to fit the battery faults in the battery pack. It can detect and screen the faults of the lithium battery pack according to the abnormal terminal voltage by using the 3σ strategy. Compared with other methods, the reliability of this model is verified, but the reliability of this method is based on high time cost and a large amount of data.

With improved computer computing power and data acquisition channels, deep learning methods based on big data acquisition have been widely used. Wang *et al.* [43] proposed transforming signals into images and achieving classification and recognition of different types of faults through an optimized convolutional neural network (CNN) model. Compared with the traditional neural network algorithm, the model can eliminate the complicated parameter tuning process and diagnose the fault directly, with an accuracy of 97%. However, the pooling layer of the convolutional

neural network will lose a lot of valuable information, and it is easy to ignore the overall and local relevance of data. The sequence structure of long short-term memory (LSTM) neural network makes it more suitable for lengthy sequence data processing. Hong *et al.* [44] first used the term LSTM neural network for fault diagnosis of the battery system. In this paper, many actual taxi operation data are used to conduct off-line training of LSTM recursive neural network, and voltage anomalies are predicted and diagnosed during lithium battery operation. The results show that this method can accurately diagnose the battery power supply fault. Although the deep learning algorithm can save the tedious steps of feature extraction, it has high requirements on processor hardware, the large amount of data required for model building, and engineering practicability to be improved.

A. MAJOR CONTENT

Through the above analysis, the main contribution of this paper can be divided into four aspects from the perspective of engineering application:

(1) The integrated test bench has been introduced to provide the environmental pressure of the battery system in the electric vehicle operation.

(2) Considering data mining principles and engineering practicability, it simply cleans the actual data and eliminates the bad points.

(3) The modified features parameters, which only depend on the fluctuation value of the voltage signal, are employed as the state identification vector, which can effectively avoid the hysteresis of the temperature signal and the difficulty of collecting the current signal.

(4) The GA-ELM is adopted to deal with the reduced dimension feature parameter matrix, which improves the reliability and practicability of the synthesis algorithm. The detailed test information is shown in Fig. 1.

The proposed method uses a genetic algorithm to optimize the initial weight and threshold value of an extreme learning machine. Compared with other traditional methods, it does not need iterative calculation, has faster operation speed, and has stronger engineering practicability. Compared with deep learning neural networks, this method has a short operation time, low requirements on hardware, and requires less data for model building and operation, so it does not need a large amount of data for training.

B. ORGANIZATION OF THE PAPER

In Section 2, the information about the test platform and battery are described in detail. The technique for a battery module in series is proposed in the third section, and the optimization principle and the modification process of feature quantity are explained systematically. The reliability verification of the synthesis algorithm and the overall conclusion is summarized in the last two sections.

II. TEST PLATFORM AND PRINCIPLES

A. COMPREHENSIVE TEST PLATFORM

The comprehensive test platform can firstly simulate the working state of an actual vehicle and then detect and record the processed signal of each cell in the variable loading mode, which provides the theoretical data for the subsequent fault identification. The platform consists of the following parts:

(1) A battery charge and discharge tester (Digtron-600) is adopted to simulate the current loaded value according to the varied demand output power in the electric vehicle operation.

(2) A vibration tester is introduced to load the vibration impact on the battery pack due to the uneven road surface and rapid acceleration or deceleration.

(3) A multi-channel information acquisition instrument is employed to obtain real-time signal data value through voltage sensor, temperature sensor, etc., record and save, and provide parameters for subsequent technical processing, as shown in Fig. 2.

The test object of this experiment is lithium-ion cells with the capacity of 50Ah, which Tianjin Lishen battery Co, LTD manufactures; the specific parameters are shown in the Table. 1. The cathode material is lithium manganite, and the negative electrode is graphite.

B. TEST PRINCIPLE

The experimental test steps in the manuscript are mainly based on China's automobile industry-standard "General Requirement of Traction Battery Enclosure for Electric Vehicles," which has been detailed provisions and instructions on the battery system design's reliability electric vehicles. To ensure the various experiment's comparability, the standard determines the environmental stress values, including frequency, sweep range, amplitude, temperature, and other parameters in various performance tests of the battery system. The sweep frequency range is 10 Hz ~55 Hz, the ambient temperature is 298 K, and the acceleration is 10 g.

III. DIAGNOSIS METHOD

A. EXTREME LEARNING MACHINE (ELM)

The extreme learning machine is an algorithm designed for a feed-forward neural network based on the traditional neural network [45]. However, the connection weights from the input layer to the hidden layer and the threshold value of the hidden layer are given manually or randomly, so only the connection weights between the hidden layer and the output value need to be calculated. Therefore, the time of error iteration calculation is reduced, and the speed of fault identification is improved, improving learning efficiency and optimizing parameter setting.

ELM neural network structure mainly consists of an input layer, an output layer, and a single hidden layer [46], as shown in Fig. 3.

For Q random distinct samples $(x_i, T_i; i = 1, \dots, Q)$, where $x \in R_n$ and $Y \in R_m$ represent the input matrix and output matrix of ELM, respectively, the output matrix can be

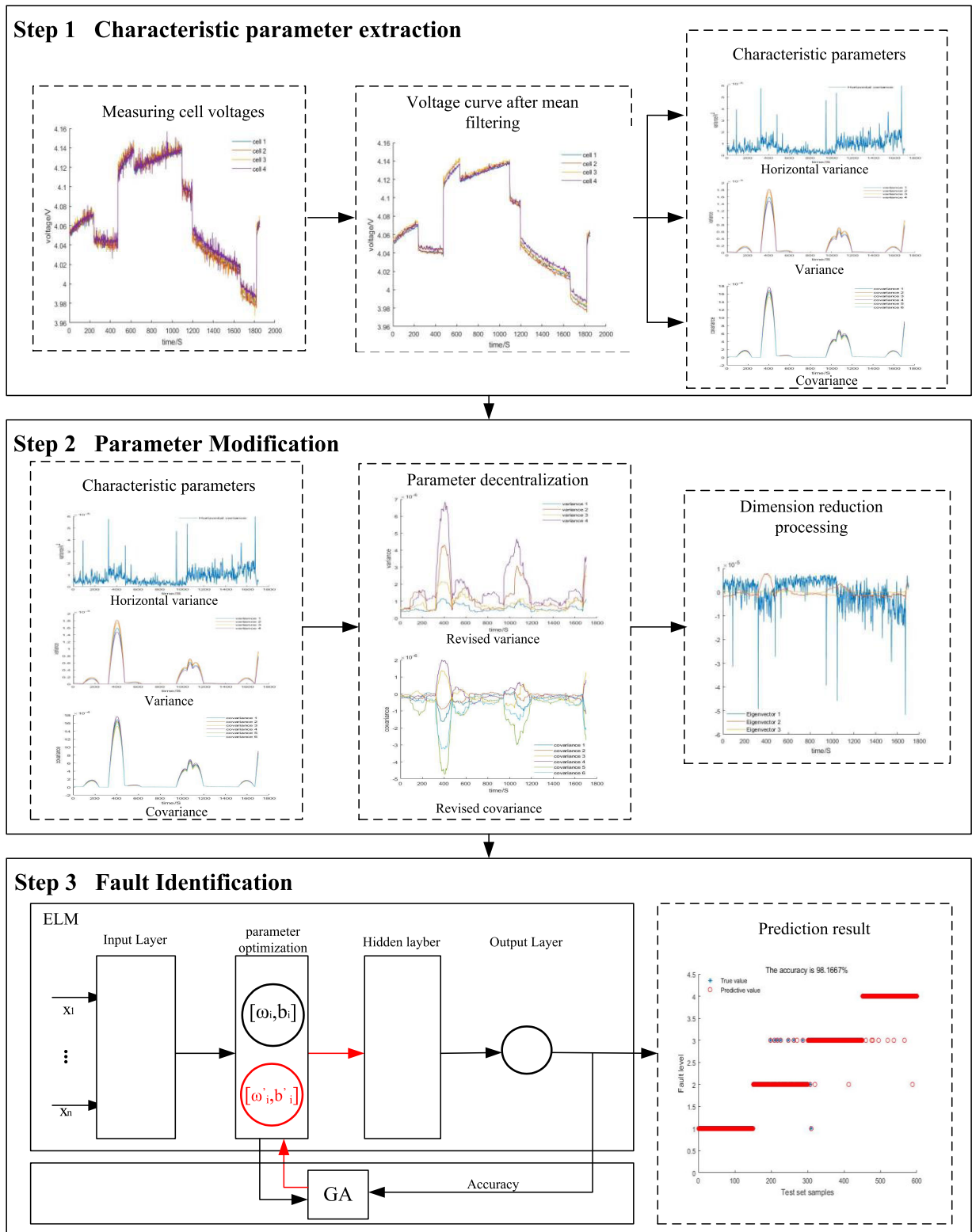


FIGURE 1. The overall framework of data processing.

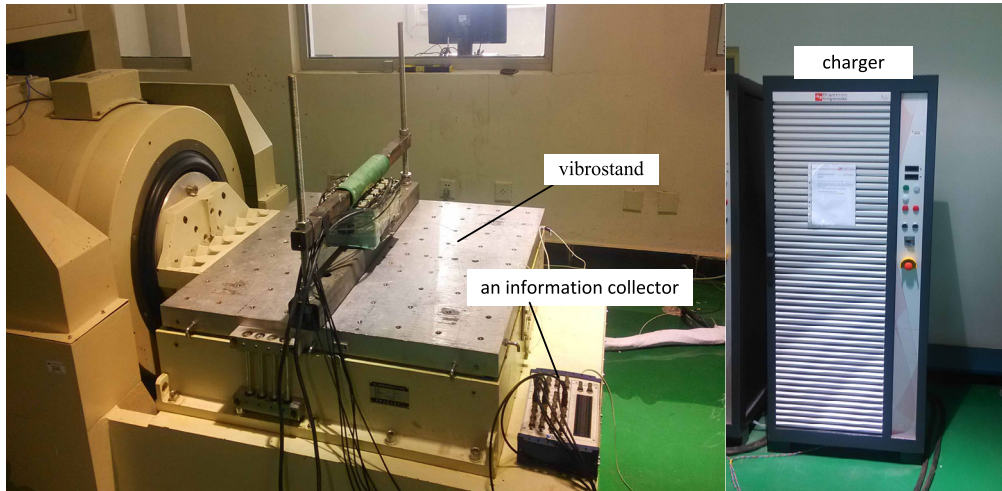


FIGURE 2. Picture of the battery test platform.

TABLE 1. Specification of the power battery for experiment.

PARAMETER	VALUE
Capacity/Ah	50
Charge cutoff voltage/V	4.2
Discharge cutoff voltage/V	3
Length/mm × Width/mm × Height/mm	170×43×100

expressed as:

$$Y = [y_1, y_2, \dots, y_Q]_{m \times Q}$$

$$y_j = \begin{bmatrix} y_{1j} \\ y_{2j} \\ \vdots \\ y_{mj} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^L \beta_{i1} g(\omega_i x_j + b_j) \\ \sum_{i=1}^L \beta_{i2} g(\omega_i x_j + b_j) \\ \vdots \\ \sum_{i=1}^L \beta_{im} g(\omega_i x_j + b_j) \end{bmatrix}_{m \times L} \quad (j = 1, 2, \dots, Q) \quad (2)$$

where:

$$X = \begin{bmatrix} x_1 & x_2 & \dots & x_Q \end{bmatrix}_{n \times Q}$$

$$= \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1Q} \\ x_{21} & x_{22} & \dots & x_{2Q} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nQ} \end{bmatrix}_{n \times Q} \quad (3)$$

$$w = [w_1 \ w_2 \ \dots \ w_L]_{L \times n}^T$$

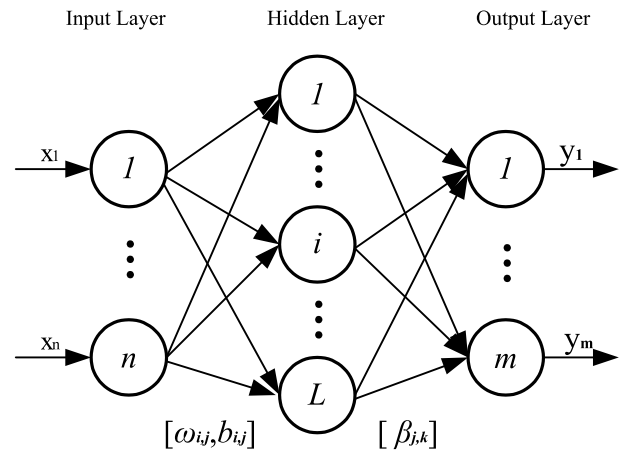


FIGURE 3. ELM network structure diagram.

$$= \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{L1} & w_{L2} & \dots & w_{Ln} \end{bmatrix}_{L \times n} \quad (4)$$

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_l \end{bmatrix}_{L \times 1} \quad (5)$$

$$\beta = [\beta_1 \ \beta_2 \ \dots \ \beta_L]_{L \times m}^T$$

$$= \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{L1} & \beta_{L2} & \dots & \beta_{Lm} \end{bmatrix}_{L \times m} \quad (6)$$

where $X = [x_1, x_2, \dots, x_Q]$ is the input matrix, and x_{ij} is the i th feature parameter of the j th sample, ($i = 1, 2, 3, \dots, n$, and $j = 1, 2, 3, \dots, Q$); n and Q represent the

number of feature parameters and samples, respectively; $w_i = [w_{i1}, w_{i2}, \dots, w_{iL}]^T$ is the connection weight between the i th neuron in the hidden layer and the input layer; L is the number of neurons in the hidden layer; b_i is the threshold of the i th neuron of the single hidden layer; $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]$ is the connection weight between the i th neuron in the hidden layer and the output layer, and m represents the number of nodes of the output layer; g is the activation function, and the sigmoid function is used in this manuscript, as follows:

$$g(x) = \frac{1}{1 + e^{-x}} \tag{7}$$

Equation (2) can be rewritten as follows:

$$H\beta = Y^T; \tag{8}$$

where the matrix H is the output weight matrix of the hidden layer, and its expression is as follows (9), as shown at the bottom of the page.

As rigorously proven in the theorems, the input weights and thresholds can be stochastically selected when the activation function is differentiable in natural fields. That is, w_i and b_i do not need to adjust the calculation. As a result, once the algorithm starts to learn, the relevant value of the H matrix will remain unchanged.

The number of the hidden layer neurons has a significant impact on the prediction result of ELM remarkably. According to the calculation principle of ELM, if the number L of nodes is not less than the number of samples, ELM can approximate these training samples with zero error. However, the number of hidden neurons is generally less than the number of training samples. The optimal solution β can be obtained, making the output error infinitely close to any small number.

The main task of using ELM for state identification is to find the best solution of formula (8) instead of the iterative method to reduce the error value, that is:

$$\beta = H^+Y = (H^T H)^{-1} H^T Y \tag{10}$$

where the H^+ represents the generalized Moore-Penrose inverse of the H matrix.

B. ELM OPTIMIZED BY GA

Compared with the traditional backpropagation (BP) neural network, ELM has fast learning speed, strong generalization ability, and it is not easy to fall into the local optimum. However, due to the random generation of input weights and thresholds in the hidden layer in the ELM initialization,

the neurons in the hidden layer have almost no regulation ability. Although the model’s accuracy can be improved by increasing the number of hidden layer neurons, its generalization ability is weakened. Furthermore, even if the number of hidden layer neurons is the same because the weight matrix of the input layer and the threshold of the hidden layer are randomly generated in the initial calculation, there is a certain difference in the input amount of each count, which will change the calculation results affect the reliability and verifiability of the results.

A genetic algorithm (GA) is a parallel random search optimization method that simulates the genetic mechanism and species evolution in nature, automatically obtaining the optimal search space under uncertain rules and adjusting the search direction adaptively. It first encodes the parameters to be optimized to form a tandem population. Then according to the fitness function, individuals are screened through selection, crossover, and mutation, and finally, the individuals with the best fitness are selected [47]. In this manuscript, GA is adopted to optimize some parameters such as w_i and b_i to build the best structural model and then establish a GA-ELM neural network to improve the robustness and accuracy of the synthesis algorithm, as shown in Fig. 4.

1) DATA ACQUISITION AND FEATURE VECTORS EXTRACTION

According to the experimental requirements, test and collect the relevant data signal, based on the extensive data processing method, conduct data cleaning and feature vector extraction.

2) DETERMINATION OF ALGORITHM TOPOLOGY

The number of the neuron’s input layer, an output layer, and a single hidden layer of the ELM is determined, and the iteration times, population number, and cut-off of GA are assigned.

3) INITIALIZATION OF POPULATION

The selected probability is determined by calculating the fitness function value of each generation population parameter, which provides the basis for the next generation population..

4) CALCULATION OF FITNESS FUNCTION VALUE

The ELM neural network’s input weights and hidden layer thresholds are generated randomly, and binary coding generates the initial population.

$$H = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & g(w_2 \cdot x_1 + b_2) & \dots & g(w_L \cdot x_1 + b_L) \\ g(w_1 \cdot x_2 + b_1) & g(w_2 \cdot x_2 + b_2) & \dots & g(w_L \cdot x_2 + b_L) \\ \vdots & \vdots & \ddots & \vdots \\ g(w_1 \cdot x_Q + b_1) & g(w_2 \cdot x_Q + b_2) & \dots & g(w_L \cdot x_Q + b_L) \end{bmatrix}_{Q \times L} \tag{9}$$

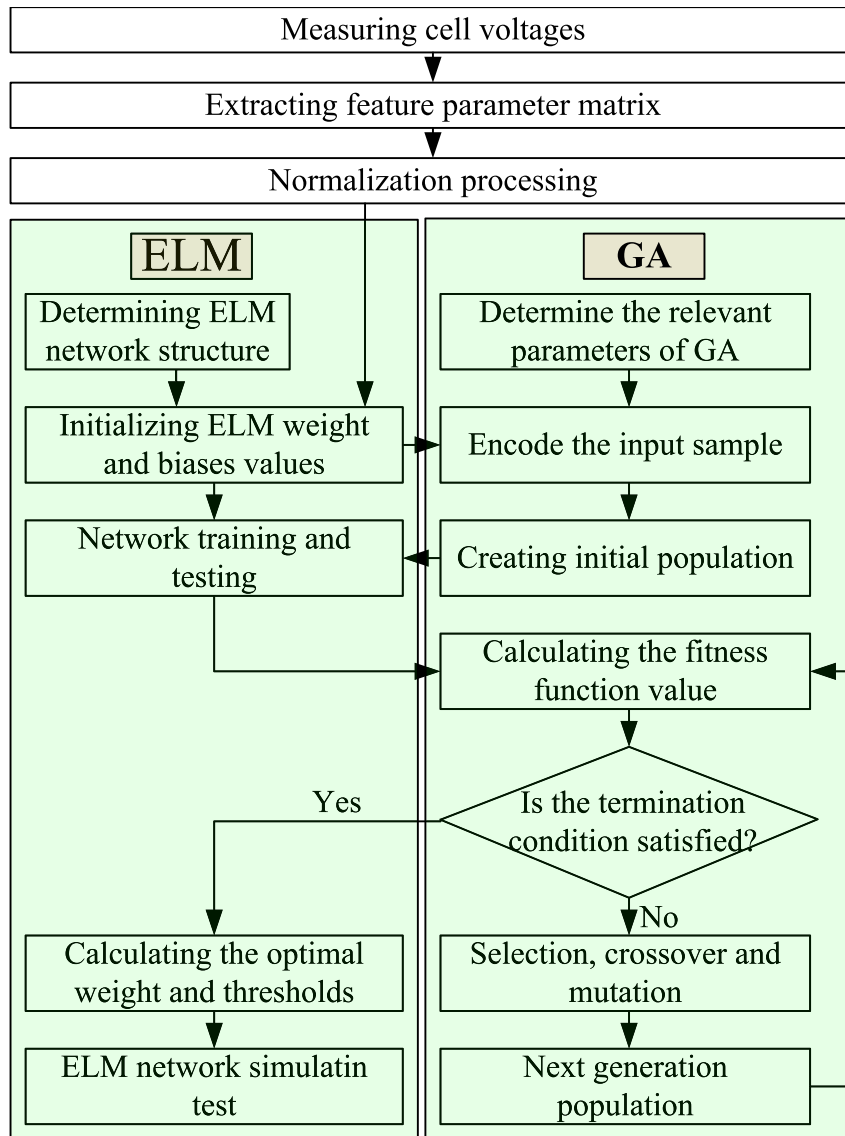


FIGURE 4. The network structure of GA-ELM.

5) FORMATION OF NEW POPULATION

According to the fitness value of individuals, the roulette method is adopted to select the chromosomes in each generation. Based on the genetic, crossover, mutation, and other operations, the amount of information used to optimize the selection population is carried out in various ways until the number of iterations is completed or the error meets the requirements and other constraints.

6) COMPLETION OF SIMULATION TEST

After iterative optimization, the optimized input weights and thresholds are obtained and then assigned to the ELM by decoding the final population. The complex matrix H of the hidden layer is calculated, and the weight matrixes between the output and hidden layers are solved according to the

formula (10). Finally, the trained network model is accepted for the simulation test.

IV. DATA ANALYSIS PROCESS

A. DENOISING PROCESSING

Based on the theoretical analysis of extensive data mining, it is necessary to clean up the real-time collected signal data and eliminate the bad points before the feature extraction of data to ensure the authenticity and reliability of the subsequent data processing results [48]. In the process of actual vehicle operation, the collected data often contain strong time-varying, nonlinear, and highly complex signal features, including standard operation signals, noise signals, and various fault signals, etc.

Under the ideal conditions, the voltage fluctuation should be consistent with the current variation trend; when the

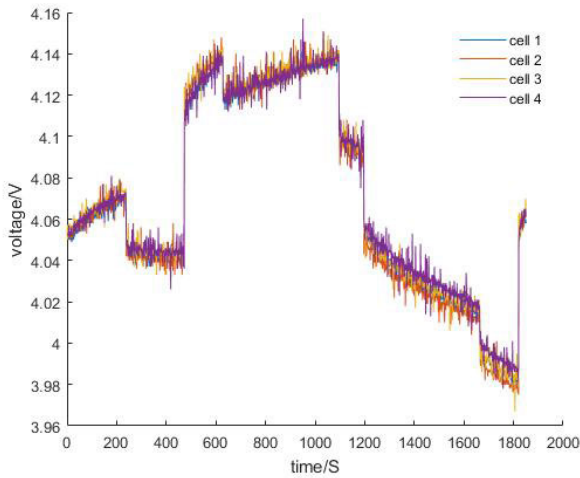


FIGURE 5. The original voltage data curve.

current is constant, the voltage curve should show a smooth upward or downward trend with time. However, in the actual measurement process, as shown in Fig. 5, in the actual vehicle application of the battery system, the noise signal can not be avoided entirely. When the voltage signal is affected by the current fluctuation, there will be many small fluctuations on the collected signal curve, which is similar to the information at the initial stage of the fault, enhancing the challenge of accurate fault diagnosis of the system.

Mean filtering is called linear filtering, and its core is to remove the mutation value of the numerical value by calculating the average value of one point and adjacent points. The equation is as follows:

$$f'(n) = \frac{1}{M} \sum_{i=n-l}^{n+l} f(i) \quad (11)$$

where $f'(n)$ is the measured value after mean filtering; $f(n)$ represents the actual measured value; n indicates the number of measurements, $n = 1, 2, 3 \dots, N$, and N is the total amount of measured data. $f(i)$ is the filtering center; M is the window length of the filter, which satisfies the following conditions: $M = 2l + 1$.

Since the influence of initial fault on voltage fluctuation is similar to that of noise interference filtering, if the filtering times are too many, the initial mark is not easy to detect. Even the voltage changes with the currency fluctuation will be seriously distorted. Therefore, to ensure the effectiveness of voltage fluctuation and keep the battery authenticity in the initial fault state to the greatest extent, this manuscript only carries out simple mean filtering, as shown in Fig. 6(a).

Although the filtered voltage curve still contains noise, the amplitudes of the curve fluctuation within each frequency band are improved to some extent.

B. FEATURE PARAMETERS EXTRACTION

Generally, data mining extracts a series of feature parameters from the time domain or frequency domain of the collected signal data, called feature functions. The feature extraction is

the critical content of machine learning, which can transform the original data into a more representative representation of the potential problems and features for the prediction model.

1) COVARIANCE MATRIX

The covariance matrix is commonly used to judge the total error of two variables in probability theory and statistics. Generally, if the changing trend of the two feature quantities is consistent, the covariance c_{ij} is positive, and $c_{ij} \in (0, 1]$; if the changing direction of the two distinct quantities is opposite, the covariance c_{ij} is negative, and $c_{ij} \in [-1, 0)$; if the two specific quantities are independent of each other, $c_{ij} = 0$.

If a data set consists of N groups of data values (x_1, x_2) , where $n = 1, 2, \dots, N$, then the covariance formula can be as follows:

$$c_{(x_1, x_2)} = \frac{1}{n-1} \sum_{i=1}^n \left(x_{i,1} - \frac{1}{n} \sum_{i=1}^n x_{n,1} \right) \left(x_{i,2} - \frac{1}{n} \sum_{i=1}^n x_{n,2} \right) \quad (12)$$

where the $c_{(x_1, x_2)}$ represents the covariance value between x_1 and x_2 , a value without unit.

In actual vehicle operation, the state data of the battery system is generated and collected continuously in real-time, resulting in a vast amount of signal data, which aggravates the dependence on hardware. The real-time prediction of the covariance algorithm is based on the principle of the covariance calculation, and m is set as the window length value of processing data, which is convenient for simplifying the calculation. As shown in Fig. 6(b), the red shaded part is the data processing window. The covariance values can be described as:

$$c_{(v_1, v_2)} = \frac{1}{m-1} \sum_{i=1}^m \left(v_{i,1} - \frac{1}{m} \sum_{i=1}^m v_{m,1} \right) \times \left(v_{i,2} - \frac{1}{m} \sum_{i=1}^m v_{m,2} \right) \quad (13)$$

where the V_i represents the voltage value of the i th cell in the acquisition signal.

The covariance matrix for the system in this manuscript can be expressed as (14), shown at the bottom of the next page.

where m is the window length for calculating the covariance value, k is the total number of the cells. $C_{(i,j)}$ is the covariance vector value between cell i and cell j . N is the total number of collected signal data, and $N-m+1$ expresses the dimension value of the covariance matrix.

2) HORIZONTAL VARIANCE MATRIX

The above two feature parameters mainly show the fluctuation curve of cell parameters with time. To obtain the feature parameters that can accurately and comprehensively represent the current system's working state, the battery system's variance when $t = t_i$ is used as the feature parameter to

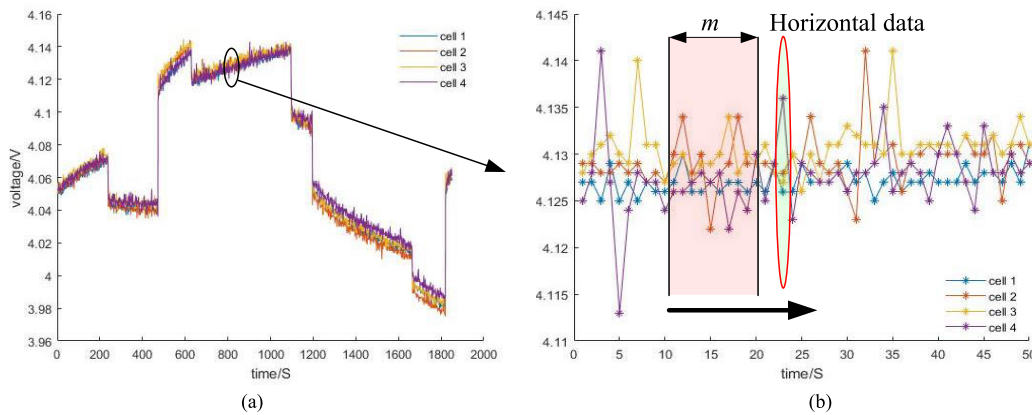


FIGURE 6. The filtered voltage data curve. (a) global voltage curve, (b) local voltage curve.

display the current distribution of the system data, as shown in Fig. 6. (b). The equation can be expressed as follows:

$$ss_j^2 = \frac{1}{k} \sum_{i=1}^k (x_{i,j} - \bar{x})^2 \quad (15)$$

where x_{ij} indicates the voltage signal value of the i -th cell at the j -th time, $i = 1, 2, 3, \dots, k$; $j = 1, 2, 3, \dots, N$, k is the number of single cells, and N is the total number of sampling data.

The horizontal variance matrix can be defined:

$$SS^2 = [ss_1^2 \quad ss_2^2 \quad \dots \quad ss_N^2] \quad (16)$$

3) VARIANCE MATRIX

In statistics, the variance is the average of the square value of the difference between each sample value and the average of all sample values, which mainly represents the deviation degree of samples. The data processing window value m is set according to the covariance processing method. The signal of the red shadow part in Fig. 6(b) is the variance data of every single cell calculated. The formula is as follows:

$$s^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \bar{x})^2 \quad (17)$$

The variance matrix value can be shown as:

$$S^2 = [s_1^2 \quad s_2^2 \quad \dots \quad s_k^2] \quad (18)$$

where the S_2 is a high latitude matrix with the dimension $(N-M+1) \times k$, and s_i^2 is the i th cell's variance value.

C. MODIFIED FEATURE PARAMETERS MATRIX

The actual voltage covariance curves for each cell are shown in Fig. 7, which indicates some similarities between each cell's signals but still quite different. Mainly in two aspects: firstly, the overall covariance curve of the movement is consistent with the fluctuation of the voltage signal; secondly, the covariance signal has a slight diversification due to the difference in the working state of each cell. Therefore, the fluctuation value of covariance data can be summarized into two parts, as shown in Fig. 7. C_1 represents the influence of real-time current on voltage data, which leads to the fluctuation of covariance data. C_2 represents the relative difference and primarily reflects the cell performance variable, mainly by the connection impedance, internal resistance, health state, and aging condition, and is the core object in this manuscript.

Because the value of C_1 is much larger than the value of C_2 , the complex issues that can represent the working state of the system contained in the collected signal can not be highlighted, which increases the risk of accurate state identification. The filtered voltage data is decentralized to reduce the current influence on the acquisition signal, and the equation is as follows:

$$v'_{ij} = v_{ij} - \frac{1}{k} \sum_{i=1}^k v_{ij} \quad (19)$$

where v_{ij} is the collected signal data of the j th cell at the i th time, and $i = 1, 2, 3, \dots, N$; $j = 1, 2, 3, \dots, k$.

The voltage change curve after decentralization is obtained based on the above equation, as shown in Fig. 8.

$$C = [C_{1,2} \quad C_{1,3} \quad \dots \quad C_{1,k} \quad C_{2,3} \quad C_{2,4} \quad \dots \quad C_{2,k} \quad \dots \quad C_{k-1,k}]$$

$$= \begin{bmatrix} c_{1,2}^1 & c_{1,3}^1 & \dots & c_{1,k}^1 & c_{2,3}^1 & c_{2,4}^1 & \dots & c_{2,k}^1 & \dots & c_{k-1,k}^1 \\ c_{1,2}^2 & c_{1,3}^2 & \dots & c_{1,k}^2 & c_{2,3}^2 & c_{2,4}^2 & \dots & c_{2,k}^2 & \dots & c_{k-1,k}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{1,2}^{N-m+1} & c_{1,3}^{N-m+1} & \dots & c_{1,k}^{N-m+1} & c_{2,3}^{N-m+1} & c_{2,4}^{N-m+1} & \dots & c_{2,k}^{N-m+1} & \dots & c_{k-1,k}^{N-m+1} \end{bmatrix} \quad (14)$$

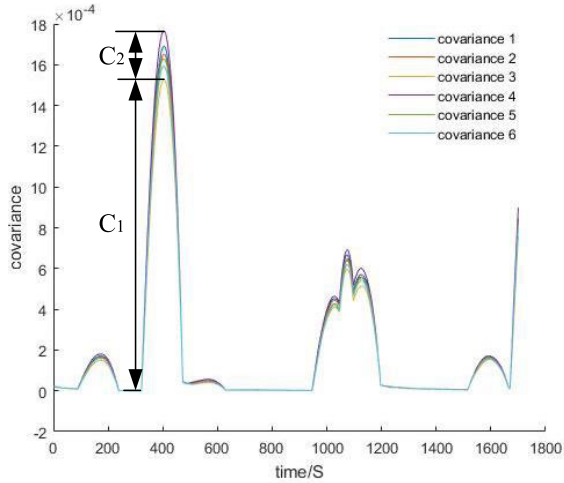


FIGURE 7. The voltage covariance data curve.

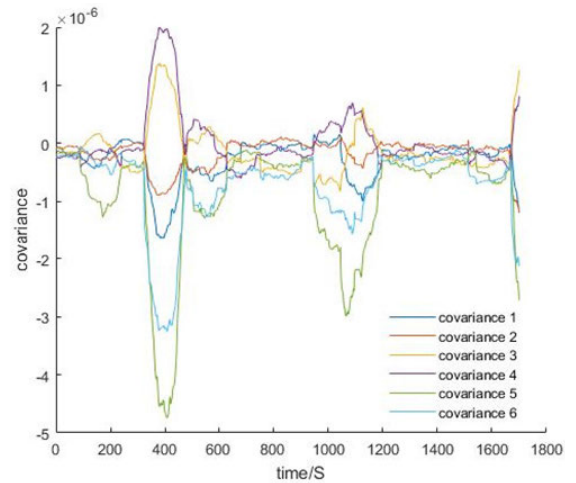


FIGURE 9. The covariance curve of the decentralized voltage data.

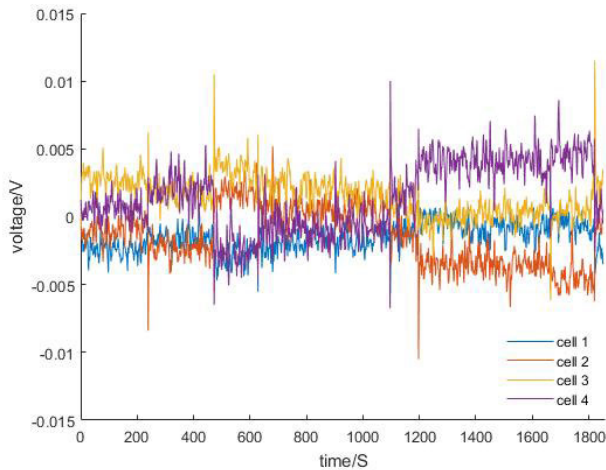


FIGURE 8. The decentralized voltage curve.

However, compared with the voltage variation trend in Fig. 6(a), the current influence on the decentralized voltage variation trend has been dramatically weakened, resulting in the relative variation of each cell being highlighted.

According to equation (14), a novel covariance curve is acquired, as shown in Fig. 9. Compared with Fig. 7, the results show that the modified data has better sensitivity for varying voltage signals. Especially when the current is set to a fixed value, the differences for the covariance curves of each cell voltage are highlighted, which will be conducive to the accurate identification of the system.

Based on the above description of the data signal processing method, the new voltage variance matrix is calculated according to the formula (17), as shown in Fig. 10(b). Meanwhile, to further verify the method's effectiveness, the variance matrices of the centralized and non-decentralized voltage data are calculated, respectively, providing parameters for comparing the subsequent results. The specific content will be discussed in the fifth section.

From Fig. 10, it is seen that the curve in Fig. 10(a) is the same as the standard covariance curve, which is mainly caused by the change of the current real-time data, and the signal difference between the cells is relatively weak in the constant current state. However, the variance matrix curve of the decentralized voltage data decreases in the overall order of magnitude, as shown in Fig. 10(b). As a result, the diversity of the cell signal changes is further increased, which is easy to enhance the system state identification later.

According to the horizontal covariance matrix's calculation formula, it is found that the data of decentralization and non-decentralization remain unchanged, as shown in Fig. 11.

D. PRINCIPAL COMPONENT ANALYSIS (PCA)

In the actual operation of electric vehicles, the battery pack usually consists of hundreds of cells in a series-parallel mode, resulting in many feature parameters such as covariance and variance among the cells, which leads to excessive data redundancy and increases the risk of system misjudgment.

Principal component analysis (PCA) is a standard technique to reduce the parameter dimension, effectively they are reducing parameter redundancy and improving the fault diagnosis efficiency. The data correlation analysis can transform the original data into effective parameters, independent of each other and contain the primary information. PCA first analyzes the features of the standardized correlation matrix and calculates the eigenvalues and corresponding eigenvectors. Then according to the cumulative variance contribution rate..

For a given feature parameter matrix A:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix} \quad (20)$$

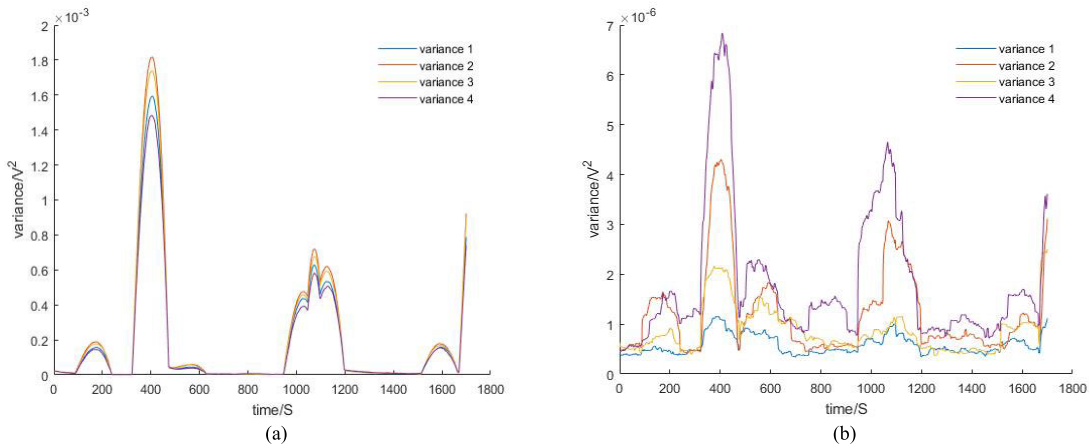


FIGURE 10. The voltage variance data curve (a) non-decentralized variance curve, (b) decentralized variance curve.

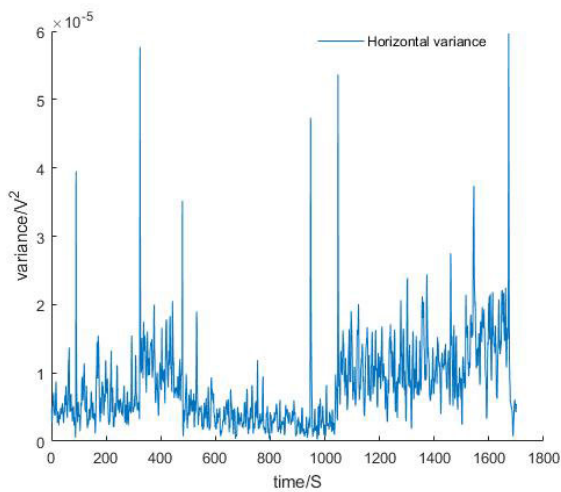


FIGURE 11. The horizontal variance-covariance data curve.

where n and p indicate the number of the specimen and feature parameters, respectively, a_{ij} is the j th feature parameter of the i th model.

Then the sample data is standardized, and the new feature parameter matrix A^* is:

$$A^* = \begin{bmatrix} a_{11}^* & a_{12}^* & \cdots & a_{1p}^* \\ a_{21}^* & a_{22}^* & \cdots & a_{2p}^* \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}^* & a_{n2}^* & \cdots & a_{np}^* \end{bmatrix} \quad (21)$$

$$a_{ij}^* = (a_{ij} - a_j') / s_j \quad (22)$$

where a_j' and s_j represent the mean and variance of the j th feature parameter, respectively.

Incidence matrix R of feature parameters:

$$R = \frac{A^{*T} A^*}{n - 1} \quad (23)$$

The eigenvalue decomposition is performed on the matrix R according to Eq. (24).

$$R\mu_i = \lambda_i\mu_i \quad (24)$$

where λ_i is the i th eigenvalue of matrix R , and μ_i is the corresponding eigenvector ($i = 1, 2, 3, \dots, p$). The eigenvalues satisfy $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, and $\mu_1 \geq \mu_2 \geq \dots \geq \mu_p$ are the corresponding eigenvectors.

According to Eq. (25) and (26), each feature parameter's variance contribution rate η_i and cumulative variance contribution rate η_{sum} are calculated. When $\eta_{sum} \geq 95\%$, it shows that the first m principal components contain most of the sample information.

$$\eta_i = \lambda_i / \sum_{i=1}^p \lambda_i \quad (25)$$

$$\eta_{sum}(m) = \sum_{i=1}^m \eta_i \quad (26)$$

The eigenvector corresponding to the principal component can be expressed as:

$$U_{p \times m} = [u_1 \quad u_2 \quad \cdots \quad u_m] \quad (27)$$

Finally, a new data set $Z_{n \times m} = A_{n \times p}^* U_{p \times m} = (z_1 \ z_2 \ \cdots \ z_p)$ is obtained. It can be found from the derivation that the newly acquired principal component variable z is a linear combination of the original variable a , indicating that each principal component variable z_i will contain all the valuable information of the actual variables a .

According to the calculation process of the above algorithm, the eigenvalues of the decentralized and non-decentralized matrices of the collected signal are obtained, respectively, as shown in Table. 2. (D: decentralized; N: non-decentralized)

From Table.2, it can be seen that based on the above equation, the proportion of eigenvalues of each order matrix

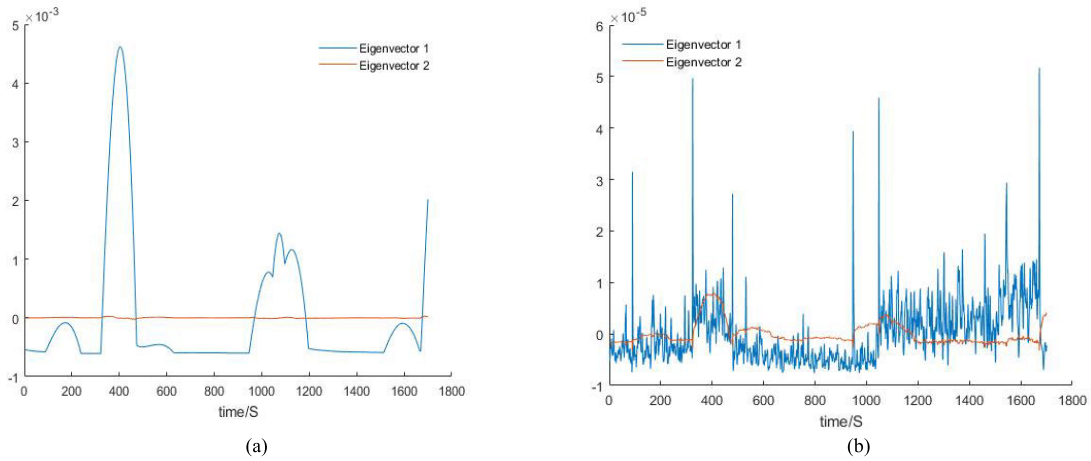


FIGURE 12. The eigenvector curve (a) after non-decentralization, (b) after decentralization.

TABLE 2. The first three eigenvalues proportion (%).

Serial number	1		2		3	
	N	D	N	D	N	D
η_i	99.99	87.07	0.0047	11.86	0.0027	0.55
η_{sum}	99.99	87.07	99.9947	98.93	99.9974	99.84

is obtained. The proportion of the first three order eigenvalues of non-decentralizing and decentralizing are as follows: 99.99%, 0.0047%, 0.0027% and 87.07%, 11.86%, 0.55%.

The first-order eigenvalues of the non-decentralizing matrix account for 99.99%, while the first-order eigenvalues of the decentralizing matrix only account for 87.07%. According to the $\eta_{sum} \geq 95\%$, the first two columns of eigenvectors are selected as the final simplified features parameter matrix to facilitate the comparison between the decentralized and non-decentralized processing. The eigenvector is shown in Fig. 12. Since the second eigenvalue of the non-decentralizing matrix accounts for 0.0047%, the corresponding eigenvector is only a horizontal line close to 0, as shown in Fig. 12. (a).

V. VERIFICATION

A. DETERMINATION OF THE MOVING WINDOW LENGTH

To realize the online state identification, the feature values of the collected signals need to be calculated in real-time. However, during the test process of the battery system, the voltage signal will be collected and sent by the sensor according to the set sampling frequency, resulting in an increasing amount of signal data. Therefore, the information can be updated by setting the moving window length to ensure sensitivity to the fault. Meanwhile, the excessive amount of data can be avoided, and the requirements for hardware systems can be reduced, which guarantees the engineering application of the algorithm in this manuscript. It should be noted that the

TABLE 3. The meaning of the state category.

Fault level number	Number meaning
I	Good condition
II	Primary fault
III	Intermediate fault
IV	Severe fault

window length m is closely related to the sensitivity and reliability of the fault identification. According to (12)-(18), for the same group of voltage signal data, the shorter the window length is, the more pronounced the fluctuation of the state parameter curve is, indicating that the higher the diagnostic precision of real-time fault identification is. However, in the process of real-time data acquisition, the noise signal will always exist in the collected signal to a varying degree. If the window length is too small, the feature parameters will be relatively increased by the noise, which will affect the robustness to a certain extent. So this manuscript sets the optimal window length to 150 to obtain strong robustness to noise signal and low computational complexity.

B. DIVISION OF DATA SETS

With the battery experiment, the fault degree is gradually severe, from the initial hard to detect to the final mars burst, and the temperature has a sharp increase. At the same time, the fluctuation degree of the voltage signal also raises rapidly. Therefore, the voltage values of different cycle periods are selected in this manuscript, as shown in Fig. 13.

As shown in Fig. 13, the voltage curve fluctuation in the graph increases sharply with the progress of the experiment. According to the fluctuation degree of the collected signal, the working state of the battery module presents four situations, namely: good condition (I), the primary fault condition (II), intermediate fault condition (III), and severe fault condition (IV), as shown in Table. 3

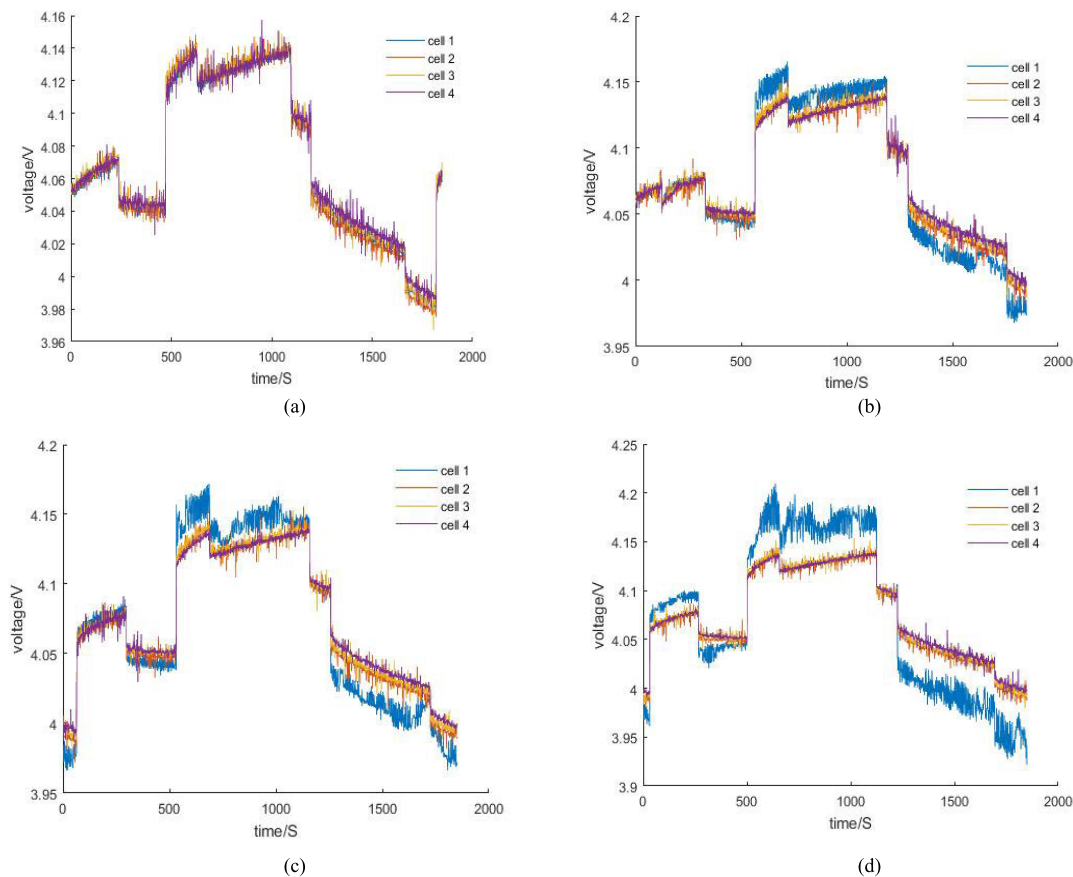


FIGURE 13. The voltage fluctuation curve of different cycles (a) n = 5, (b) n = 8, (c) n = 10, (d) n = 12.

C. RESULTS COMPARATIVE ANALYSIS

This manuscript selects 6800 sets of data, including 1360 data sets as the training set, mainly used to acquire the connection weights between the hidden layer and the output layer and the threshold of the hidden layer in the ELM model, and the rest of the data as the test set to verify the algorithm’s precision. Meanwhile, to further improve the engineering practicability and robustness of the algorithm, the test set and training set data are randomly generated.

In the paper, the PCA matrix of the modified feature matrix (PMFM) and the PCA matrix of the fundamental feature matrix (PUFYM) are both adopted as the input of the GA-ELM model, respectively.

As shown in Fig. 14, by comparing the position and change trend of the two curves, it shows that the accuracy of ELM with PMFM increases from 94.25% to 97.83%, and the absolute error is only 2.17%. The result is relatively acceptable when combined with the principle of no more than 5% deviation in a limited data set. While the accuracy of ELM with PUFYM increases from 92.05% to 94.5%, and the absolute error is 5.5% (>5%), which is lower than that of ELM with PUFYM. Especially compared with the first generation, the accuracy of ELM with PMFM is 2.2% higher than that of ELM with PUFYM without the GA influence, which indicates

TABLE 4. The prediction accuracy of GA-ELM.

Serial number	Prediction accuracy without GA %		Final prediction accuracy %		Accuracy appreciation %	
	PUFYM	PMFM	PUFYM	PMFM	PUFYM	PMFM
1	92.25	94.75	94.77	97.83	2.52	3.08
2	91.73	93.75	94.83	97.07	3.1	3.32
3	92.17	95.01	95.19	97.97	3.02	2.96

that the modified feature parameter matrix has more vital state representation ability and is conducive to the accurate identification of system state. Secondly, the output accuracy of GA for ELM in two forms is improved by 3.558% and 2.45%, respectively, which shows that ELM has strong generalization ability and high precision. The identification accuracy is completed in 30 iterations.

In the manuscript, PMFM and PUFYM are adopted repeatedly to work out the identification accuracy, and the results are shown in Table. 4

Table. 4 shows that before the GA algorithm is used, the three diagnostic accuracy rates of ELM with PUFYM are 92.25%, 91.73%, and 92.17%, respectively, with an average of 92.05%. after GA optimization, an average accuracy increased by 2.88%. However, without GA optimization,

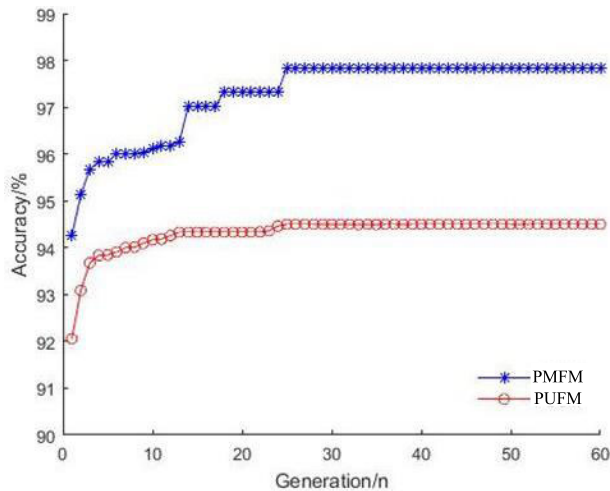


FIGURE 14. Evolutionary, iterative graph of genetic algorithms.

the diagnosis accuracy of ELM with PMFM is 94.75%, 93.75%, and 95.01%, respectively, and the average value is 94.5%. After GA optimization, the accuracy is raised at 97.83%, 97.07%, and 97.97%, with an average increase of 3.12%.

Through many experimental tests, the data fluctuation in Table. 4 shows the same regular as the previous analysis, which further verifies the effectiveness and authenticity of the algorithm proposed in this manuscript. However, it is easy to find that the accuracy of each identification is different by observing the data of the prediction results. Mainly because the division of the training set and the test set is random to verify further the robustness and reliability of the comprehensive algorithm proposed in this paper, which results in the input of the GA-ELM model is not the same, which lead to some errors in the prediction results.

In conclusion, PMFM is more suitable as the state feature parameter. The accuracy of the state diagnosis of the ELM algorithm is improved by the GA optimization algorithm, which makes the proposed comprehensive diagnosis method more efficient and accurate.

VI. CONCLUSION

This paper studies the response of the voltage signal of each cell to the change of fault degree when the battery system is charging or discharging based on data-driven theory. There are some bad points and noise interference signals in the collected data. The data cleaning is done by simple mean filtering to improve the signal's practicability. Because of the battery system's strong time-varying, nonlinear and multi-parameter coupling, it is pretty challenging to identify the fault's location and time with a single parameter. Therefore, variance, covariance, and horizontal variance are selected as feature parameters to determine the system's current state. However, the traditional feature parameters fluctuate violently with the current values, which submerges the diversity of individual signals and increases the fault diagnosis's risk.

Furthermore, the modified feature parameters are introduced as the state input parameters of the ELM algorithm for the first time. Because the connection weights and thresholds of the ELM model are generated randomly, it is easy to cause some hidden layer nodes to fail, reducing the generalization ability and accuracy. Therefore, fault identification's comprehensive diagnosis method is proposed based on ELM optimized by GA algorithm. Meanwhile, the voltage faults are divided into four levels according to the intensity of volatility in the voltage data. Finally, the effectiveness and robustness of the proposed method are further proved through the comparison of various ways and the verification of measured data. In conclusion, the comprehensive diagnosis method of the fault identification proposed in this paper can accurately and efficiently identify the relevant information about the fault and reflect the intensity of the current system work, providing an innovative theoretical mechanism and technical support for the future battery system intelligent management.

Due to the rapid development of computer hardware technology, deep learning has been widely used. Combined with the characteristics of the power battery itself, such as solid time-varying, nonlinear and multi-parameter, intelligent diagnosis with high identification accuracy, good reliability, and strong robustness will be implemented in the subsequent work.

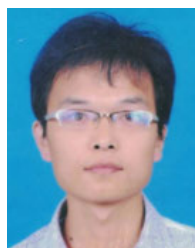
DECLARATION OF COMPETING INTEREST

The authors declared that they have no conflicts of interest. The authors alone are responsible for the content and writing of this article.

REFERENCES

- [1] Z. P. Cano, D. Banham, S. Ye, A. Hintennach, J. Lu, M. Fowler, and Z. Chen, "Batteries and fuel cells for emerging electric vehicle markets," *Nature Energy*, vol. 3, no. 4, pp. 279–289, Apr. 2018.
- [2] Y. Wang, D. Meng, R. Li, Y. Zhou, and X. Zhang, "Multi-fault diagnosis of interacting multiple model batteries based on low inertia noise reduction," *IEEE Access*, vol. 9, pp. 18465–18480, 2021.
- [3] 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.
- [4] T. Hu, K. Li, H. Ma, H. Sun, and K. Liu, "Quantile forecast of renewable energy generation based on indicator gradient descent and deep residual BiLSTM," *Control Eng. Pract.*, vol. 114, Sep. 2021, Art. no. 104863.
- [5] 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, Jan. 2019.
- [6] L. Yao, S. Xu, A. Tang, F. Zhou, J. Hou, Y. Xiao, and Z. Fu, "A review of lithium-ion battery state of health estimation and prediction methods," *World Electr. Vehicle J.*, vol. 12, no. 3, p. 113, Aug. 2021.
- [7] L. Liu, X. Feng, M. Zhang, L. Lu, X. Han, X. He, and M. Ouyang, "Comparative study on substitute triggering approaches for internal short circuit in lithium-ion batteries," *Appl. Energy*, vol. 259, pp. 114–143, Mar. 2020.
- [8] M. A. Hannan, M. M. Hoque, A. Hussain, Y. Yusof, and P. J. Ker, "State-of-the-art and energy management system of lithium-Ion batteries in electric vehicle applications: Issues and recommendations," *IEEE Access*, vol. 6, pp. 19362–19378, 2018.
- [9] M. Seo, M. Park, Y. Song, and S. W. Kim, "Online detection of soft internal short circuit in lithium-ion batteries at various standard charging ranges," *IEEE Access*, vol. 8, pp. 70947–70959, 2020.

- [10] K. Liu, X. Hu, H. Zhou, L. Tong, W. D. Widanage, and J. Marco, "Feature analyses and modeling of lithium-ion battery manufacturing based on random forest classification," *IEEE/ASME Trans. Mechatronics*, vol. 26, no. 6, pp. 2944–2955, Dec. 2021.
- [11] 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," *J. Power Sources*, vol. 417, pp. 132–144, Mar. 2019.
- [12] T. Han, Y.-F. Li, and M. Qian, "A hybrid generalization network for intelligent fault diagnosis of rotating machinery under unseen working conditions," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.
- [13] Z. Chen, G. He, J. Li, Y. Liao, K. Gryllias, and W. Li, "Domain adversarial transfer network for cross-domain fault diagnosis of rotary machinery," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 8702–8712, Nov. 2020.
- [14] J. Hong, Z. Wang, F. Ma, J. Yang, X. Xu, C. Qu, J. Zhang, T. Shan, Y. Hou, and Y. Zhou, "Thermal runaway prognosis of battery systems using the modified multiscale entropy in real-world electric vehicles," *IEEE Trans. Transport. Electrification*, vol. 7, no. 4, pp. 2269–2278, Dec. 2021.
- [15] Y. Wang, J. Tian, Z. Sun, L. Wang, R. Xu, M. Li, and Z. Chen, "A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems," *Renew. Sustain. Energy Rev.*, vol. 131, Oct. 2020, Art. no. 110015.
- [16] J. Wei, G. Dong, and Z. Chen, "Lyapunov-based state of charge diagnosis and health prognosis for lithium-ion batteries," *J. Power Sources*, vol. 397, pp. 352–360, Sep. 2018.
- [17] K. Liu, X. Hu, Z. Yang, Y. Xie, and S. Feng, "Lithium-ion battery charging management considering economic costs of electrical energy loss and battery degradation," *Energy Convers. Manage.*, vol. 195, pp. 167–179, Sep. 2019.
- [18] L. Yao, Z. Fang, Y. Xiao, J. Hou, and Z. Fu, "An intelligent fault diagnosis method for lithium battery systems based on grid search support vector machine," *Energy*, vol. 214, Jan. 2021, Art. no. 118866.
- [19] Z. Lu, X. L. Yu, L. C. Wei, F. Cao, L. Y. Zhang, X. Z. Meng, and L. W. Jin, "A comprehensive experimental study on temperature-dependent performance of lithium-ion battery," *Appl. Thermal Eng.*, vol. 158, Jul. 2019, Art. no. 113800.
- [20] B. Xia, Y. Shang, T. Nguyen, and C. Mi, "A correlation based fault detection method for short circuits in battery packs," *J. Power Sources*, vol. 337, pp. 1–10, Jan. 2017.
- [21] Y. Pan, X. Feng, M. Zhang, X. Han, L. Lu, and M. Ouyang, "Internal short circuit detection for lithium-ion battery pack with parallel-series hybrid connections," *J. Cleaner Prod.*, vol. 255, May 2020, Art. no. 120277.
- [22] H. Maleki and J. N. Howard, "Effects of overdischarge on performance and thermal stability of a Li-ion cell," *J. Power Sources*, vol. 160, pp. 1395–1402, Oct. 2006.
- [23] M. Chen, F. Bai, W. Song, J. Lv, S. Lin, Z. Feng, Y. Li, and Y. Ding, "A multilayer electro-thermal model of pouch battery during normal discharge and internal short circuit process," *Appl. Thermal Eng.*, vol. 120, pp. 506–516, Jun. 2017.
- [24] X. Feng, C. Xu, X. He, L. Wang, G. Zhang, and M. Ouyang, "Mechanisms for the evolution of cell variations within a $\text{LiNi}_x\text{Co}_y\text{Mn}_z\text{O}_2/\text{graphite}$ lithium-ion battery pack caused by temperature non-uniformity," *J. Cleaner Prod.*, vol. 205, pp. 447–462, Dec. 2018.
- [25] R. Zhao, J. Liu, and J. Gu, "Simulation and experimental study on lithium ion battery short circuit," *Appl. Energy*, vol. 173, pp. 29–39, Jul. 2016.
- [26] W. Fang, P. Ramadass, and Z. Zhang, "Study of internal short in a Li-ion cell-II. Numerical investigation using a 3D electrochemical-thermal model," *J. Power Sources*, vol. 248, pp. 1090–1098, 2014.
- [27] Q. Yu, C. Wan, J. Li, R. Xiong, and Z. Chen, "A model-based sensor fault diagnosis scheme for batteries in electric vehicles," *Energies*, vol. 14, no. 4, p. 829, Feb. 2021.
- [28] X. Feng, Y. Pan, X. He, L. Wang, and M. Ouyang, "Detecting the internal short circuit in large-format lithium-ion battery using model-based fault-diagnosis algorithm," *J. Energy Storage*, vol. 18, pp. 26–39, Aug. 2018.
- [29] X. Feng, X. He, L. Lu, and M. Ouyang, "Analysis on the fault features for internal short circuit detection using an electrochemical-thermal coupled model," *J. Electrochem. Soc.*, vol. 165, no. 2, pp. A155–A167, Jan. 2018.
- [30] Y. Kang, B. Duan, Z. Zhou, Y. Shang, and C. Zhang, "Online multi-fault detection and diagnosis for battery packs in electric vehicles," *Appl. Energy*, vol. 259, Feb. 2020, Art. no. 114170.
- [31] X. Li and Z. Wang, "A novel fault diagnosis method for lithium-ion battery packs of electric vehicles," *Measurement*, vol. 116, pp. 402–411, Feb. 2018.
- [32] K. Liu, Y. Shang, Q. Ouyang, and W. D. Widanage, "A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery," *IEEE Trans. Ind. Electron.*, vol. 68, no. 4, pp. 3170–3180, Apr. 2021.
- [33] 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," *Appl. Energy*, vol. 196, pp. 289–302, Jun. 2017.
- [34] 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 Trans. Power Electron.*, vol. 34, no. 10, pp. 9709–9718, Oct. 2019.
- [35] A. Tourani, P. White, and P. Ivey, "Analysis of electric and thermal behaviour of lithium-ion cells in realistic driving cycles," *J. Power Sources*, vol. 268, pp. 301–314, Dec. 2014.
- [36] 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," *J. Cleaner Prod.*, vol. 187, pp. 950–959, Jun. 2018.
- [37] K. Liu, X. Hu, J. Meng, J. M. Guerrero, and R. Teodorescu, "RUBoost-based ensemble machine learning for electrode quality classification in Li-ion battery manufacturing," *IEEE/ASME Trans. Mechatronics*, early access, Oct. 18, 2021, doi: [10.1109/TMECH.2021.3115997](https://doi.org/10.1109/TMECH.2021.3115997).
- [38] Q.-Q. Yu, R. Xiong, L.-Y. Wang, and C. Lin, "A comparative study on open circuit voltage models for lithium-ion batteries," *Chin. J. Mech. Eng.*, vol. 31, no. 1, pp. 76–84, Aug. 2018.
- [39] S. Akash, C. Sumana, and S. W. Sheldon, "Machine learning-based data-driven fault detection/diagnosis of lithium-ion battery: A critical review," *Electronics*, vol. 10, p. 1139, May 2021.
- [40] J. Hong, Z. Wang, W. Chen, L. Wang, P. Lin, and C. Qu, "Online accurate state of health estimation for battery systems on real-world electric vehicles with variable driving conditions considered," *J. Cleaner Prod.*, vol. 294, Apr. 2021, Art. no. 125814.
- [41] C. Sbarufatti, M. Corbetta, M. Giglio, and F. Cadini, "Adaptive prognosis of lithium-ion batteries based on the combination of particle filters and radial basis function neural networks," *J. Power Sources*, vol. 344, pp. 128–140, Mar. 2017.
- [42] 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," *Appl. Energy*, vol. 207, pp. 354–362, Dec. 2017.
- [43] H. Wang, S. Li, L. Song, and L. Cui, "A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals," *Comput. Ind.*, vol. 105, pp. 182–199, Feb. 2019.
- [44] 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," *Appl. Energy*, vol. 251, Oct. 2019, Art. no. 113381.
- [45] G. Zheng, W. Hua, Z. Qiu, and Z. Gong, "Detecting water depth from remotely sensed imagery based on ELM and GA-ELM," *J. Indian Soc. Remote Sens.*, vol. 49, no. 4, pp. 947–957, Nov. 2020.
- [46] Z. M. Yaseen, S. O. Sulaiman, R. C. Deo, and K.-W. Chau, "An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction," *J. Hydrol.*, vol. 569, pp. 387–408, Feb. 2019.
- [47] G. Renner and A. Ekárt, "Genetic algorithms in computer aided design," *Comput. Aided Des.*, vol. 35, no. 8, pp. 709–726, Jul. 2003.
- [48] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, Apr. 2015.



LEI YAO received the Bachelor of Engineering degree in mechanical design, manufacturing and automation from the Henan University of Science and Technology, in June 2008, the Master of Engineering degree in mechanical design and theory from Chongqing University, in June 2011, and the Doctor of Engineering degree in vehicle engineering from the Beijing Institute of Technology, in January 2016. Since March 2016, he has been a Lecturer with the Department of Vehicle Engineering, School of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry. His research interests include charging strategies for new energy power batteries and fault diagnosis for group systems.



SHIMING XU was born in Zhengzhou, Henan, China, in 1998. He is currently pursuing the Graduate degree with the School of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry. His research interests include charging strategies for new energy power batteries and fault diagnosis for group systems.



YANQIU XIAO received the Bachelor of Engineering degree from the School of Mechanical Engineering, Henan University of Science and Technology, in June 2004, and the Doctor of Engineering degree from the School of Mechanical and Vehicle, Beijing Institute of Technology, in January 2009. He was a Visiting Scholar with the Industrial Engineering Laboratory, School of Engineering, Nanyang Technological University, Singapore, from September 2008 to December

2008. He was a Visiting Scholar with the Advanced Manufacturing Laboratory, School of Economics and Engineering Management, University of Luneburg, Germany, from December 2013 to December 2014. Since January 2009, he has been the Vice President with the School of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry. His research interests include digital design and manufacturing and intelligent electromechanical systems.



JUNJIAN HOU received the Bachelor of Engineering degree in mechanical manufacturing and automation (automobile) major from the Henan University of Science and Technology, in June 2004, the Master of Engineering degree in vehicle engineering from the Beijing University of Aeronautics and Astronautics, in April 2007, and the Doctor of Engineering degree in mechanical design and theory from Shanghai Jiao Tong University, in September 2011. He was a Postdoctoral

Fellow with the Technology Research Institute, Zhengzhou Yutong Bus Company Ltd., from September 2011 to April 2014. Since April 2014, he has been the Director of the Department of Vehicle Engineering, School of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry. His research interests include automotive system dynamics, vibration and noise control, sound source identification, and intelligent acoustic fault diagnosis.



XIAOYUN GONG received the Ph.D. degree from Zhengzhou University, in 2013. She is currently a Professor majoring in mechanical engineering with the Zhengzhou University of Light Industry, Zhengzhou, China. Her current research interests include vibration signal processing and rotating machinery fault diagnosis.



ZHIJUN FU (Member, IEEE) received the Doctor of Engineering degree in vehicle engineering from the University of Science and Technology Beijing, in June 2003. He is currently a Joint Training Doctoral Student with Canada Concordia University. Since January 2019, he has been an Associate Professor with the Vehicle Engineering Department, Zhengzhou University of Light Industry. His research interests include intelligent autonomous vehicle decision-making and dynamic control,

driver behavior data mining and reinforcement learning, distributed estimation/control and intelligent vehicle applications, optimal control, and energy saving assistance for vehicle powertrain systems.



AIHUA TANG (Member, IEEE) was born in 1980. He received the Ph.D. degree from the Beijing Institute of Technology. He is currently an Associate Professor with the School of Vehicle Engineering, Chongqing University of Technology.

...