Data-Driven Fault Diagnosis of Lithium-Ion Battery Overdischarge in Electric Vehicles

Gan, Naifeng ; Sun, Zhenyu; Zhang, Zhaosheng; Xu, Shiqi ; Liu, Peng; Qin, Zian

DOI
10.1109/TPEL.2021.3121701

Publication date
2022

Document Version
Accepted author manuscript

Published in
IEEE Transactions on Power Electronics

Data-driven Fault Diagnosis of Lithium-ion Battery Over-discharge in Electric Vehicles

Naifeng Gan, Zhenyu Sun[1], Zhaosheng Zhang, Shiqi Xu, Peng Liu and Zian Qin, Senior *Member, IEEE*

***Abstract*—The over-discharge can significantly degrade a lithium-ion (Li-ion) battery's lifetime. Therefore, it is important to detect the over-discharge and prevent severe damage of the Li-ion battery. Depending on the battery technology, there is a minimum voltage (cut-off voltage) that the battery is allowed to be discharged in common practice. Once the battery voltage is below the cut-off voltage, it is considered as over-discharge. However, over-discharge will not lead to immediate failure of the battery, and if it is not detected, the battery voltage can increase above the cut-off voltage during charging process. How to detect an over-discharge has happened, while the current voltage is larger than cut-off voltage, thus becomes very challenging. In this paper, a machine learning (ML) based two-layer over-discharge fault diagnosis strategy for lithium-ion batteries in electric vehicles is proposed. The first layer is to detect the over-discharge by comparing the battery voltage with cut-off voltage, like what is utilized in common practice. If the battery voltage is larger than the cut-off voltage, the second layer, which is a detection approach based on eXtreme Gradient Boosting (XGBoost) algorithm, is triggered. The second layer is employed to detect the previous over-discharge. The proposed method is validated by real electric vehicle data.**

***Index Terms*—Electric vehicle (EVs), fault diagnosis, lithium-ion battery (LIB), over-discharge, eXtreme Gradient Boosting (XGBoost).**

I. Introduction

Electric vehicles (EVs) are highly valued and actively developed as practical approaches to tackle with energy crises and air pollution issues [1-3]. Lithium-ion batteries (LIBs), exhibiting outstanding energy density, low self-discharge rate, stable performance, and long lifespan, are becoming the mainstream energy storage system in EVs. However, as a relatively more active material compared with lead-acid, LIBs in EVs have encountered a lot of failures [4-6]. Meanwhile, the safety of EVs, especially the electric passenger cars, is of high importance. As a result, it is essential to detect the fault occurrence in the battery pack timely [7]. The main possible electrical faults of the battery system in EVs can be categorized into connection fault, sensor fault and battery abuse fault [8-10]. The battery abuse faults can be further divided into different types, i.e., short circuit [11], overcharge [12], and over-discharge [13]. Over-discharge is becoming an increasingly common issue in EV applications due to huge current strike, inappropriate design of management system (BMS), long-term storage and severe inhomogeneity among modules. The deep over-discharge can lead to many irreversible changes, i.e., Cu dissolution [14], impedance increase [7], and the solid electrolyte interphase (SEI) decomposition [15]. And then the thermal runaway of the battery is triggered. Although deep over-discharge is more harmful, it is easy to find by detecting the discharging voltage. However, the slight over-discharge is difficult to be detected timely by BMS because of the slight differences in the manufacturing and operation of normal cells [13]. And long-time slight over-discharge is easy to cause the deterioration of battery performance, such as permanent capacity fade. Therefore, it is necessary to detect an over-discharge of a battery cell happen in the past, while the battery cell voltage is higher than the cut-off voltage.

The over-discharge can be caused by factors including the failure of inconsistency among cells in the battery pack, self-discharge at low state-of-charge (SOC), and BMS failure, occurs commonly in the operation of EVs [13, 14]. Most existing studies have focused on investigating the failure mechanism and fault diagnosis in over-discharge conditions. For instance, Fear *et al.* [15] researched the copper dissolution phenomenon of LIBs under different depths of discharge (DODs). The copper dissolution of over-discharged cells can deposit on the electrode surfaces and bridge electrodes, which may lead to an internal short circuit (ISCr). Wu *et al.* [16] revealed the law of voltage–temperature change in LIBs with 110% DOD over-discharge. Based on the previous studies, it can be concluded that the voltage variation is the measurable and effective feature of over-discharge behavior.

The diagnostic approaches, which have been proposed in existing studies, are roughly classified into three categories, i.e., knowledge-, model-, and data-driven based methods [17, 18]. Knowledge-based methods utilize the expert system or fault

This work was supported by the National Key R&D Program of China (No. 2019YFE0107900).

1 The author (Zhenyu Sun) has equal contribuation with the first author.

(Corresponding authors: Peng Liu, roc726@163.com; Zian Qin, Z.Qin-2@tudelft.nl)

Naifeng Gan (e-mail: gannaifeng@163.com), Zhenyu Sun (bitzhenyu@163.com), Zhaosheng Zhang (e-mail: zhangzhaosheng@bit.edu.cn), Shiqi Xu (e-mail: 18910440672@163.com), Peng Liu (e-mail: roc726@126.com) is with the National Engineering Laboratory for Electric Vehicles, Beijing Institute of Technology, Beijing, China.

Zian Qin (Z.Qin-2@tudelft.nl) is with the Department of Electrical Sustainable Energy, Delft University of Technology, Delft 2628 CD, The Netherlands.

tree based on the knowledge or observation of battery systems to indicate the fault states of batteries [8]. For instance, Zhang *et al.* [19] proposed a fault diagnosis method based on a fault tree, which could learn the representative features automatically. However, due to the uncertainty and complexity of the battery system, there are still several problems in false knowledge representation and effective rules. In contrast, model-based fault diagnostics utilize the residuals obtained by comparing the model output with measurable signals to describe the evolutions of battery under different conditions [11, 20-22]. For example, Wang *et al.* [23] utilized a Kalman filter (KF) based state observer for joint estimation of both battery voltage and SOC using the battery model, which could quantitatively assess the positive and negative virtual insulation resistance for fault diagnosis. However, as the electrochemical system of batteries is nonlinear, KF may not achieve satisfactory performance because of its inherent properties. In this regard, Wei *et al.* [24] employed the strong tracking extended Kalman filter (ST-EKF) to establish a battery model and achieve online voltage estimation. The ST-EKF was able to improve the diagnostic accuracy against bad initial values. Besides, Sidhu *et al.* [25] proposed an adaptive nonlinear model-based fault diagnosis method that could detect multiple signature faults accurately. Nevertheless, due to the significant impact of model accuracy and data noise, the diagnosis accuracy of these methods can only be enhanced by using high precision modelling and inputting sufficient excitations.

Data-driven methods are widely used in fault diagnosis of LIBs without the need for accurate battery models [26-28]. Kim *et al.* [29] presented a feature extraction method based on discrete wavelet transform to detect the inconsistency fault of LIBs. Analogously, Kang *et al.* [30] proposed a multi-fault diagnostic method according to an interleaved voltage measurement topology for battery packs, which could diagnose several types of faults. There are plenty of differences between laboratory and real-world EVs, i.e., measurement accuracy, operation condition, and driving factors. Some fault diagnosis methods, which are based on the operation data of real-world EVs, have been proposed in recent years. Li *et al.* [31] proposed a sample entropy based LIBs fault diagnosis method verified by real-world data and this method can identify battery faults under various operating conditions. Analogously, Liu *et al.* [32] presented a mechanism of voltage fault diagnosis using entropy theory in the actual operation situation of EVs, which was verified to identify the abnormity of cell voltage and locate the fault cells. Besides, Zhao *et al.* [33] introduced a fault diagnosis method using 3σ multi-level screening strategy and a neural network algorithm to detect the abnormal voltage cells. However, these methods can only be used in specific vehicles by determining specific thresholds for fault diagnosis, which is not adaptive to various EVs models.

With the in-depth development of artificial intelligence, the machine learning methods, which learn the underlying battery fault mechanisms from massive battery training samples, are widely used in data-driven based fault diagnosis LIBs [34]. For instance, Hong *et al.* [35] developed a novel deep-learning-enabled method to perform accurate multi-forward-step voltage prediction for battery systems, which combined with alarm thresholds to implement the fault prognosis. Besides, Zhao *et al.* [36] developed a structured recurrent neural network (RNN) under various operating conditions trained with drive cycle data to predict the voltage and SOC accurately. However, these methods only utilize the operation data of batteries to train the deep-learning model, which seldom consider the electrochemical characteristics. To solve this problem, Li *et al.* [37] proposed a voltage abnormality fault diagnosis method combining the long short-term memory neural network (LSTM) and the equivalent circuit model (ECM), which could achieve accurate fault diagnosis for potential cell failure. Generally, these machine learning methods based on the operation data of EVs have high precision and robustness but with high computational cost.

The limitations of the fault diagnosis methods in literature can be summarized as following. Firstly, conventional fault diagnosis methods based on experimental data do not fully consider the complex operation conditions of practical vehicles and lack of verification with real vehicle operation data. Secondly, the fault diagnosis methods in literature show a trade-off between accuracy and computation time - they either have a low accuracy or long computation time. A reasonable balance between the two features is necessary. In this paper, a data driven based detection approach for over-discharged battery is proposed. Firstly, immense amount of data collected from the practical vehicular operation is utilized to train the approach offline. By this way, the approach is comparable with the other approach of highest accuracy that proposed in previous literatures. Secondly, by applying XGBoost algorithm, which is essentially an integrated machine learning algorithm based on decision tree and supports parallel computing, the computation time is greatly reduced. Eventually, the verification with real operation data of EVs shows that the proposed approach can effectively detect over-discharge of battery in the past, while the battery voltage is larger than the cut-off voltage.

This paper is organized as follows. Section II describes the data acquisition of experiment and real-world EVs, including data screening and pre-processing. Section III introduces the two-layer over-discharge fault diagnosis strategy briefly and presents the voltage prediction method in detail. Section IV provides the results of the diagnostic method based on experimental and real-world data and discusses the relationship between them discovered by the proposed method. Finally, Section V concludes the article. The scheme of the proposed over-discharge fault diagnosis method is shown in Fig. 1.

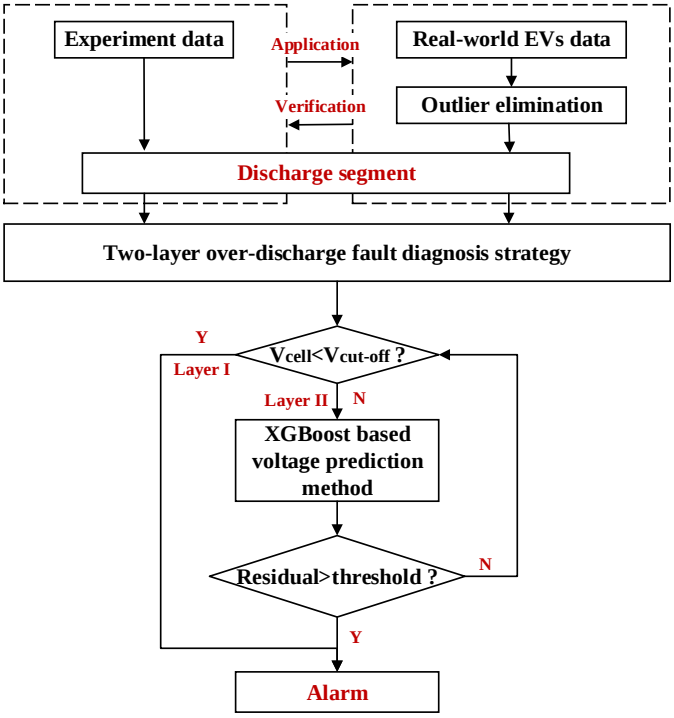


Fig. 1. The scheme of the proposed fault diagnosis method.

II. Data Acquisition

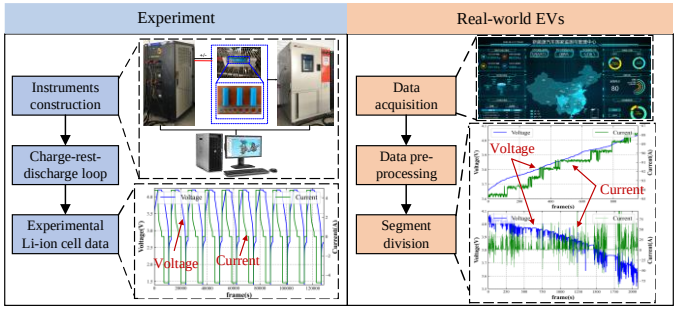


Fig. 2. The scheme of data acquisition for experiment and real-world EVs.

Fig. 2 illustrates the scheme of the data acquisition for the over-discharge fault diagnosis, including two parts of data utilized to train and verify the algorithm's effectiveness.

A. *Experimental data*

In reference [38], seven 21,700-size silicon-graphite/NCA lithium-ion battery cells with a rated discharge capacity of 4750mAh and voltage range of 2.5V-4.2V were selected for over-discharge testing. The instruments for testing are shown in Fig. 2. During the experiment, all cells were placed in a chamber to keep the almost constant temperature (25±0.5℃), and the frequency of data acquisition was 1Hz. Firstly, one of the selected cells (numbered as 6 in this paper) was discharged at 1C until the voltage no longer changes. The voltage profile during the process is shown in Fig. 3, the over-discharge could be roughly divided into 3 stages with 5 critical points. In stage I, the voltage dropped rapidly from the point A (2.5V, discharging cut-off voltage) to point D (0V). In stage II, the cell voltage reversed and continued to drop to the inflection point E (-0.55V). In stage III, the voltage curve rose to point F (-0.45V) and then dropped slowly, which indicated that irreversible damage happened in the cell. Finally, cell voltage rebound gradually and underwent a monotonic gradual increase to -0.33V approximately. In this paper, Stage I to Stage III are defined as slight over-discharge, deep over-discharge, and extreme over-discharge. The other 6 cells are numbered from 0 to 5 and they are discharged to the end points of A to F corresponding shown in Fig. 3.

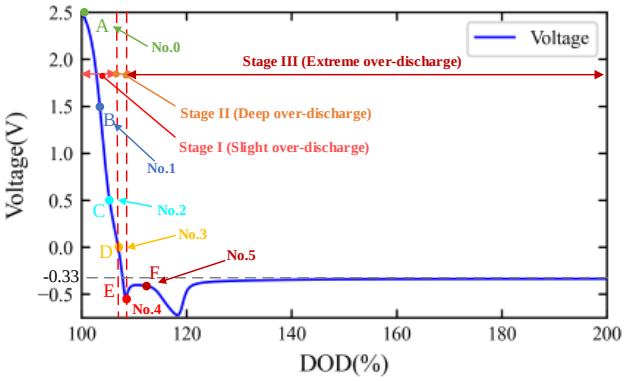


Fig. 3. The voltage curve of battery during unabridged over-discharge.

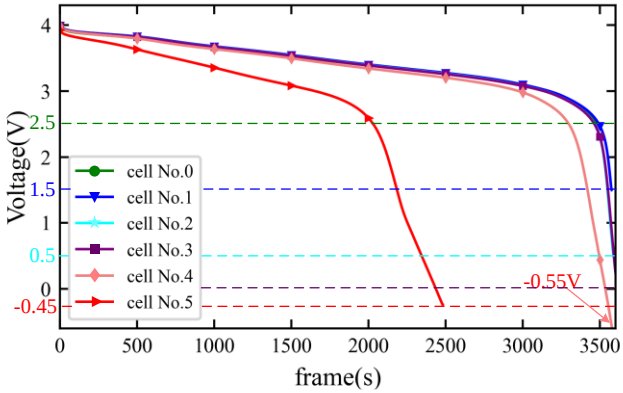


Fig. 4. The voltage curves of 6 cells during the last time over-discharge cycle.

Then, over-discharge cycling tests based on the aforementioned 6 end points are set as follows. Cells are charged by CC mode (1C) until reach the charge cut-off voltage, then charged by constant voltage (CV) mode of 4.2V until the current drops to 0.02C. After a 30 minutes rest, cells are discharged by CC mode (1C) to their respective DOD, resting for 30 minutes. The total cycle number is set to 30. If the cells could not be recharged, the cycle stopped. It is noteworthy that Cell No.4 and No.5 cannot be recharged after the cycle 15 and cycle 12. Thus, the last time over-discharge voltage curves of 6 cells are illustrated in Fig. 4. According to Fig. 4, compared with the fresh cell (No.0), both discharge time and discharge capacity in the last cycle (12#) of Cell No.5 reduced by almost 30%.

B. *Real-world EVs data*

The National Monitoring and Management Center for New Energy Vehicles (NMMC-NEV) is responsible for managing new energy vehicles, including pure electric vehicles, hybrid electric vehicles, plug-in hybrid electric vehicles, and fuel cell electric vehicles. There are various data in the NMMC-NEV including the location of vehicles and parameter values of power battery, i.e., voltage, current, mileage and temperature

[39].

In this paper, the operation data of 8 real-world EVs with two kinds of voltage acquisition errors (two vehicle models) are divided into charge and discharge segments. The information of vehicles employed in this paper is shown in Table I.

TABLE I
INFORMATION OF VEHICLE AND BATTERY PACK

Electric vehicle type	Battery type	Vehicle model (M)	Rated motor power	Cell nominal capacity	Voltage range
Battery electric transportati -on vehicle	Ternary lithium battery (NCA)	M1	30kW	2.5Ah	3.2V-4.2V
		M2	18kW	2.6Ah	2.5V-4.2V

It is noteworthy that the discharge characteristics of the battery in these vehicles are similar to the experimental cells as mentioned above. Based on the battery maintenance records, all vehicles were divided into two categories that shown in Table II. Due to the disorder of the practical vehicular operation data, data pre-processing is carried out to convert the original data into a well-organized form. The steps of data pre-processing include extreme voltage elimination, missing data filling and duplicate frames elimination. Considering this is not the focus of this paper, this part is introduced briefly. Then, the discharge segments are extracted after data pre-processing for subsequent study.

TABLE II
THE CATEGORIES OF 8 VEHICLES EMPLOYED IN THIS PAPER

Categories	Vehicle number	Vehicle model (M)	Voltage acquisition errors (mV)	Fault type
Normal vehicle	No.0	M1	20	\
	No.1	M1	20	
	No.5	M2	1	
	No.6	M2	1	
	No.7	M2	1	
Abnormal vehicle	No.2	M1	20	Over-discharge
	No.3	M1	20	Over-discharge
	No.4	M1	20	Inconsistency of battery cells

III. METHODOLOGY

A. *Two-layer over-discharge fault diagnosis strategy*

In EVs, once the voltage drops under the cut-off threshold, the under-voltage alarm in BMS will be triggered, which is different from the cells in the experimental test that can reach over 100% DoD. However, the slight over-discharge cannot be detected timely, and because of the inconsistency of the pack, some of the cells with low capacity may tend to be over-discharged slightly. Hence, it is necessary to establish a strategy to diagnose slight over-discharge in time. This paper proposes a two-layer over-discharge fault diagnosis strategy. In reference [14] and [15], the process of over-discharge mechanism is introduced. According to them, the stage of over-discharge can be represented by the voltage variation. Thus, when the voltage drops to the cut-off voltage, an over-charge will occur in the cell. For real world application, the fault diagnosis should be fast and real-time. Therefore, in order to reduce the calculation time, a single threshold is set to detect the over-discharge. The over-discharge warning will be triggered if the voltage drops below the cut-off voltage. When the cell voltage is higher than the cut-off voltage, the method of voltage prediction based on XGBoost will be employed to diagnose the over-discharge of the battery happened in the past. If no past over-discharge failure is detected, the cell voltage will continue to be monitored. The flowchart of strategy is illustrated in Fig. 5. It is noteworthy that the data screening and outlier elimination are necessary for pack voltage from the real-world EVs. There only a slight over-discharge in the real-world EVs and the cell voltage is barely lower than 0V. Since the fault diagnosis of Layer I only requires the pack's voltage, Layer II would be the main point for subsequent study.

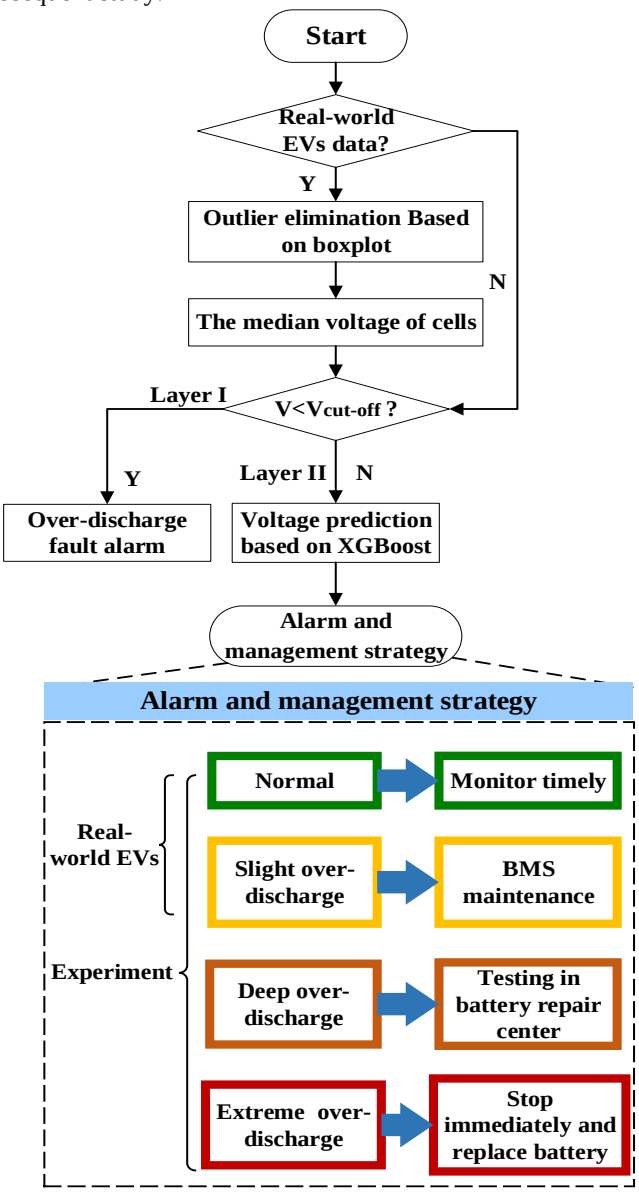


Fig. 5. The flowchart of two-layer over-discharge fault diagnosis strategy.

B. *XGBoost prediction algorithm*

The XGBoost with the full name of eXtreme Gradient Boosting, is a well-developed gradient boosting framework that

proposed by Friedman [40]. In recent years, XGBoost has been widely used in the fields of financial market prediction, forecasting medical disease and power load prediction [41-43]. Gradient boosting algorithm is a machine learning approach that applied in building predictive tree-based models. This is an integrated algorithm that adds new function to fit the residual of the last model. This algorithm is called as boosting [44]. The XGBoost is well suit to multivariate multistep-ahead time series prediction. In essence, voltage prediction is a regression process. The decision tree is the foundation of XGBoost, but it is prone to be overfitting. When overfitting occurs, the generalization performance of model will decrease. Compared with GBDT (Gradient Boosting Decision Tree), the target function of XGBoost has a regularizer. Thus, the iterative effect and operability of each iteration round will be improved. In addition, XGBoost supports parallel computing, which reducing computing time greatly. Considering the real-time monitoring in real-world application, XGBoost is employed to predict the voltage of discharge segments of experimental testing and real-world EVs, which is illustrated in Fig. 6.

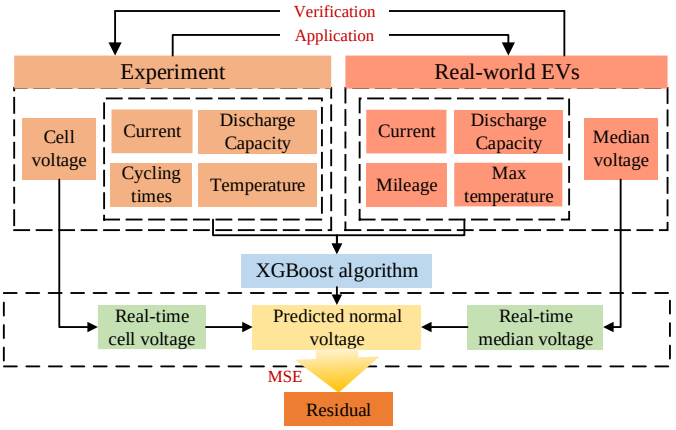


Fig. 6. The voltage prediction process with features and result based on XGBoost algorithm.

Classification and regression trees (CART) are the typical binary decision trees in XGBoost, which can be used in the classification and regression process [40]. If the predicted result is discrete, CART will generate the classification decision tree. On the contrary, once the result is continuous, it will generate the regression decision tree, the output of which is the mean value of all predicted samples belong to a certain leaf node [40]. The *Gain_σ* is chosen to evaluate the behavior of split attribute. Once the *Gain_σ* obtained by certain attributes is smaller, there is little difference between two sub-samples obtained by dichotomy. Meanwhile, the certain attribute can be defined as splitting attribute. The continuous predicted results, total variance and *Gain_σ*, which are obtained from sample set *S*, can be calculated by

$$\sigma(S) = \sqrt{\sum (y_k - \mu)^2} \tag{1}$$

$$Gain\_\sigma_A(S) = \sigma(S_1) + \sigma(S_2) \tag{2}$$

Where μ and y_k denote the mean value of the predicted result in sample set *S* and the predicted result of sample *k*, respectively. The $Gain\_\sigma_A(S)$ is the *Gain_σ* that gains from the attribute *A*, which divides *S* into two parts.

The two parts calculated by attribute A are continued to be split by the optimized method mentioned above. Finally, as for sample set *S*, the minimum value of optimal dichotomous scheme is chosen. It is worth noting that, XGBoost performs parallel computing in split attribute optimization with high training efficiency, which is the main difference from GDBT [45].

The target optimization function of the XGBoost model is given by

$$obj^{(t)} = \sum_{i=1}^{N} l(y_i, y_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) + constant \tag{3}$$

Where l is the loss function, $f_t(x_i)$ is the value of i-th leaf node in t-th regression tree and $\Omega(f_t)$ is the regularizer. Herein, mean square error (MSE) is employed as the loss function to ensure the global optimum.

The discharge voltage prediction based on the XGBoost regression algorithm can be expressed as

$$Y_i = \sum_{t=1}^{T} f_t(X_i) \tag{4}$$

Where $f_t(X_i)$ is the t-th regression tree, T is the total number of regression trees and Y_i is the i-th predicted voltage. Moreover, to predict the voltage timely with newly fed operation data, the sliding windows is employed. Finally, regression trees are optimized step by step and the final predicted voltage is the sum of results from all trees.

In reference [35] and [37], Pearson correlation coefficient (PCC) is utilized to calculate the correlation between ambient temperature and voltage. The correlation between them is only 0.16, which indicates that those two parameters have no correlation. The PCCs between the voltage and different input features are shown in Fig. 7. When it comes to vehicle states, current, SOC and temperature can directly affect the battery voltage. Besides, as the accumulated mileage and cycling times increase, the battery's capacity keeps decaying, which may enlarge the inconsistency of battery packs and shorten the cycle life. So, both of them have apparent impacts on the voltage and can be used as the ageing characteristics of the battery in EVs and experiment, respectively [37].

	Cell voltage
Ambient temperature	0.16
Current	0.42
Mediem voltage	0.67
SOC	0.96
Speed of vehicle	0.19
Cell voltage	1
Max temperature	0.22
Discharge capacity	0.35

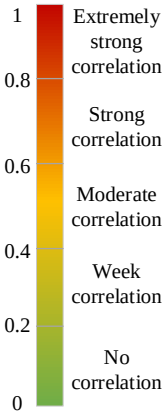


Fig. 7. PCC values between voltage and different input features.

The features of the experiment and real-world EVs in this paper are shown in Table III. Online voltage prediction should be updated timely when a new time step of data are provided. Thus, a sliding time window is used to extract the training interval and to feed the features of calculation intervals into the trained XGBoost for voltage prediction. The training set, testing set and verification set with the sliding window are illustrated

TABLE III
THE FEATURES EMPLOYED IN XGBOOST ALGORITHM

Sources of data	Features	Description
Experiment	Discharge capacity	The cumulative capacity of cells during discharge record by Arbin test equipment
	Current	The discharge current (1C in this paper)
	Temperature	Temperature measured by thermocouple on the cell surface
	Cycling times	Numbers of charge-discharge cycles
	Voltage	The discharge voltage of cell
Real-world EVs	Discharge capacity	Calculated by ampere-hour integral based on the discharge current
	Current	The discharge current of battery pack
	Max temperature	The highest measured temperature of each probe of battery pack
	Mileage	The cumulative mileage of EV
	Voltage	The median value among all cell voltages after eliminating outliers

in Fig. 8. As shown in Fig. 8, the sample set of time series $V_{prediction}=\{(X_k,Y_k)\}$ is constructed by using the sliding window and the size of which is w. $X_i=\{S_k,S_{k+1},\ldots,S_{k+w-1}\}$ is the input data of XGBoost. S_k includes all parameters mentioned in Table III, $Y_k=\{S_{k+w}\}$ is the output data after prediction, and k is the sequence number of the sample set. When the data length does not reach the size of sliding window w, it will be saved and accumulated. Only if the data length is larger than w, the data of previous w frame will be assimilated and utilized to predict the next frame. Thus, the voltage prediction model is given by

$$Y_{prediction}=\left\{(X_i,Y_i,Y_{i+1},\ldots,Y_N)\right\} \quad (5)$$

The size of the training data has a certain influence on the prediction accuracy of XGBoost. An extensive training data will contain too many features, which may increase training complexity and training time for XGBoost. However, the XGBoost will not learn the features and result in unsatisfactory prediction results if the amount of training data is too small. Therefore, the size of the sliding time window in this paper is determined by a comparison experiment.

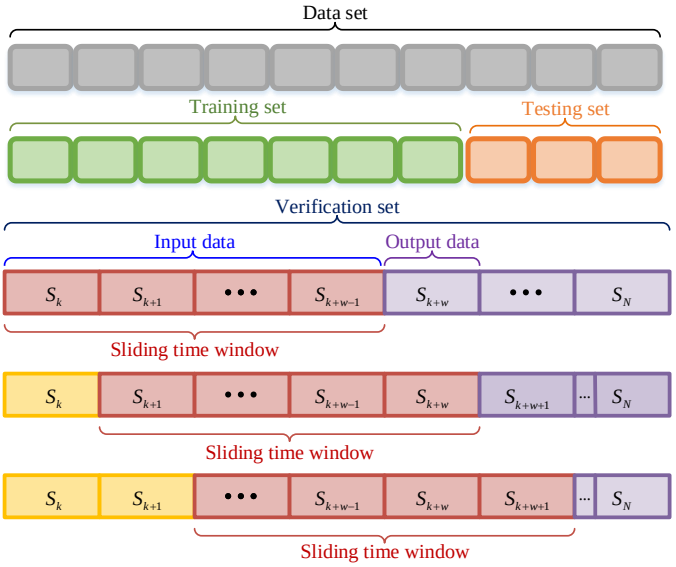


Fig. 8. The training set, testing set and verification set as well as the schematic diagram of sliding window in this paper.

In order to minimize the error between the predicted value and real value, parameter optimization is employed to get the prediction model accurately. On the one hand, the hyperparameters of XGBoost are set as the initial values. And on the other hand, the loss of results obtained from the testing data is compared. Therefore, the general steps to determine the hyperparameter of XGBoost are as follows and the flowchart of parameter tuning is illustrated in Fig. 9:

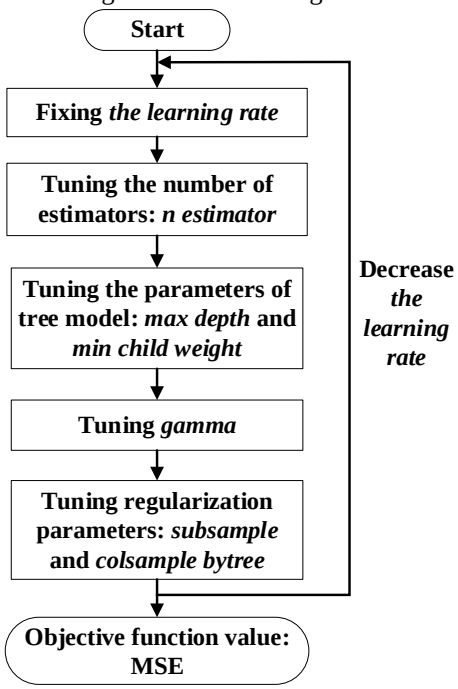


Fig. 9. The flowchart of hyper-parameters tuning.

(1) Firstly, *the number of estimators* is tuned to optimize the accuracy of XGBoost with other parameters fixing. The accuracy (also known as generalization error) of the model is determined by variance, deviation and noise, which is given by

$$E=bias^2+var+\varepsilon^2 \quad (6)$$

Where E is the generalization error, *bias* is the deviation, *var* is the variance and ε is the noise. By adjusting the *bias* and *var*, *the number of estimators* when the generalization error of minimum value is obtained.

(2) The next parameters to be tuned are the *max depth* and *min child weight* of trees which can affect the conservativeness or complexity of the algorithm.

(3) *Gamma* is tuned to make the model more conservative.

(4) Different combinations of *subsample* and *colsample bytree* are tuned to prevent overfitting.

(5) The last parameter to be tuned is the *learning rate*. Generally, the *learning rate* should be lower to prevent overfitting.

MSE is adopted in this paper to evaluate the prediction result on the testing data during each cross-validation. In order to make the model more conservative, 10-fold cross-validation is employed to obtain the optimized tuning result. The data utilized for training XGBoost model are divided into training data and testing data in the proportion 7:3, as illustrated in Fig. 8.

C. Prediction accuracy and residual

After the parameter optimization, the discharge data of normal cell are utilized to train the prediction model. MSE represents the prediction accuracy and the residual between predicted value and real value is given by

$$MSE = \frac{1}{M}\sum_{m=1}^{M}(y_m - y_m)^2 \tag{7}$$

$$\Delta V = y_m - y_m \tag{8}$$

IV. Results and Discussions

A. Over-discharge fault diagnosis based on experimental data

The experimental discharge data of Cell No.0 are employed to establish the voltage prediction model based on XGBoost. According to the maximum residuals calculated by experimental data, the thresholds for over-discharge fault diagnosis are carried out.

The size of training data determined by cycling times and has a certain influence on the prediction accuracy, is discussed. Firstly, the training data with different cycling times are selected to build the prediction model by the untuned XGBoost. Then, the MSE of different cycling times is compared. In order to shorten the training time, the first 6th discharge segments data of Cell No.0 are used for testing. The comparative result about prediction accuracy and operation time are shown in Fig. 10. It illustrates that the MSE decreases continually and the operation time tends to be longer with the larger amount of training data, simultaneously. The prediction accuracy improves significantly when the cycling times is over *four-cycle*, whereas the operation time of *five-cycle* increases hardly. Hence, in this paper, the first five discharge segments and last discharge segment data are adopted as training data and testing data, respectively.

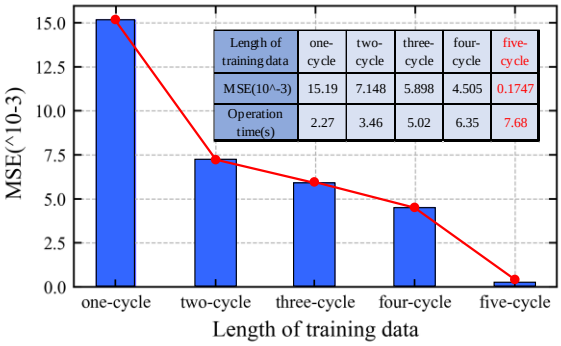


Fig. 10. The prediction accuracy and operation time under different length of training data.

In order to improve the accuracy of the model, the hyperparameters of XGBoost require to be optimized. Herein, tree booster is employed in voltage prediction. It is necessary to select the parameters, which exert the dominant influence in prediction results, to optimize preferentially. The general steps as mentioned in chapter III are utilized to determine the optimization order of hyperparameters. The first priority for optimization is *the number of estimators*. Besides, with the increase of *max depth*, the model is likely to have less bias but tend to be overfit. Nevertheless, the model will be more conservative with the increase of *min child weight.* Therefore, *max depth* and *min child weight*, which can help prediction model to keep balance, are chosen as the parameters in second priority to optimize. The ranges of both parameters are 3-9 and 1-6, respectively. Fig. 11 illustrates the variation of MSE with the change of three parameters noted above. Finally, the best combination of other parameters is obtained by the traversal search method, whose MSE is the minimum one of the MSE under different parameter combinations. In this paper, *the number of estimators*, *max depth* and *min child weight* are 50, 5,4 respectively.

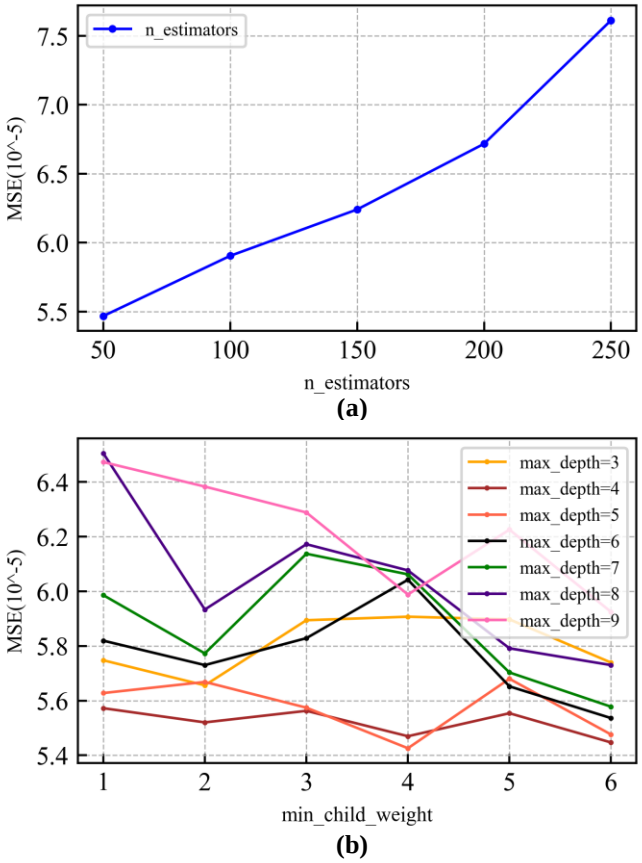


Fig. 11. The MSE under the variations of (a) *n_estimators*, (b) *max_depth* and *min_child_weight*.

After parameters optimization, the best combination of parameters is selected to establish the voltage prediction model. In this paper, the first five-cycle discharge segments of Cell No.0 are employed as training data to train the voltage prediction model of normal discharge. Meanwhile, the accuracy of the model is tested by the discharge data of Cell No.0 during the final cycle. The prediction result of Cell No.0 (normal discharged cell) during the last discharge segment is shown in Fig. 12(a). The MSE is 3.46×10^{-5} and the maximum of residual is 0.018V. In order to verify the superiority of the method, the MSEs and operation time of proposed method are compared with the LSTM method, the LSTM-ECM method, the Linear-Regression-based method and the SVM-based method, which are illustrated in Table IV. According to Table IV, we can draw 2 conclusions as follows. Firstly, the proposed method has the minimum value of MSE compared with the other four methods.

It verifies that the proposed method has the ideal accuracy of voltage prediction. Secondly, compared to other methods, the operation time of the proposed method is relatively short, which can ensure the timeliness of the prediction result.

TABLE IV
THE MSE AND OPERATION TIME OF PROPOSED METHOD AND THE OTHER FOUR METHODS

Voltage prediction method	Prediction accuracy (MSE)	Operation time (s)
LSTM [35]	7.04×10^{-3}	65.27
LSTM-ECM [37]	6.60×10^{-5}	69.42
Linear-Regression-based method [46]	7.84×10^{-4}	2.31
SVM-based [46]	3.14×10^{-3}	48.13
Proposed method	3.46×10^{-5}	7.68

We input one of the discharge segments of over-discharged cells to the mentioned model to predict the voltage of the over-discharge segment. Then, we obtain the residuals of real value and predicted value. The over-discharge fault diagnosis method thresholds are calculated by the prediction result of those six cells which are illustrated in Fig. 12(a)-(f). In order to prevent the false positives, the maximum residual of Cell No.1, the DOD of which is 103.3%, is taken into consideration to set the threshold of Normal Stage.

As illustrated in Fig. 12(a), the ΔV from equation (7), representing the deviation of the predicted value from the normal, always fluctuates symmetrically around 0V. Simultaneously, when the discharge voltage is about to reach the cut-off voltage (2.5V), the ΔV approaches to 0V. Nevertheless, as shown in Fig. 12(b), the ΔV of Cell No.1 decreases to 0V, then increases until reaches a peak voltage of 0.02V approximately, following a decreasing process and then

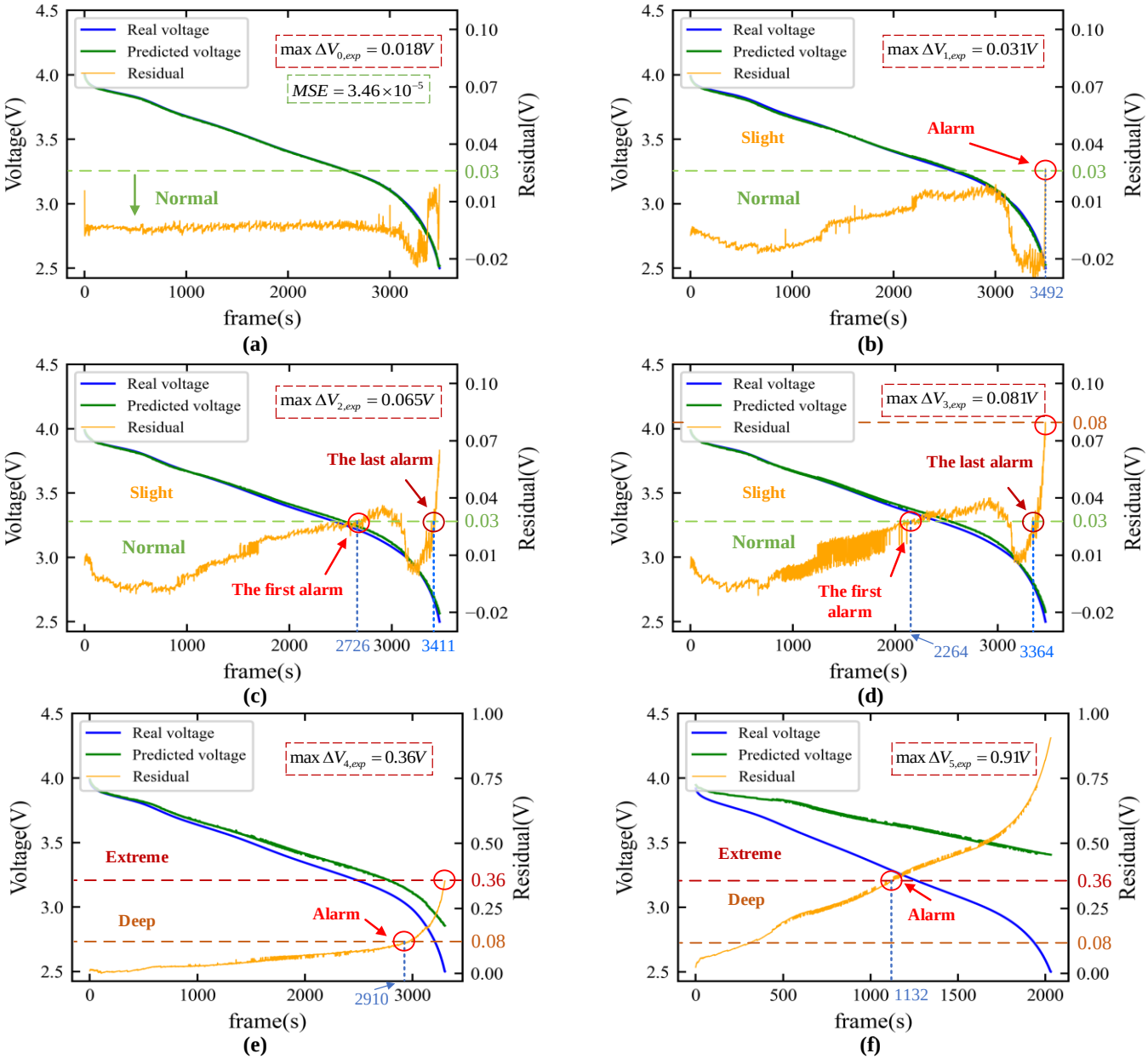
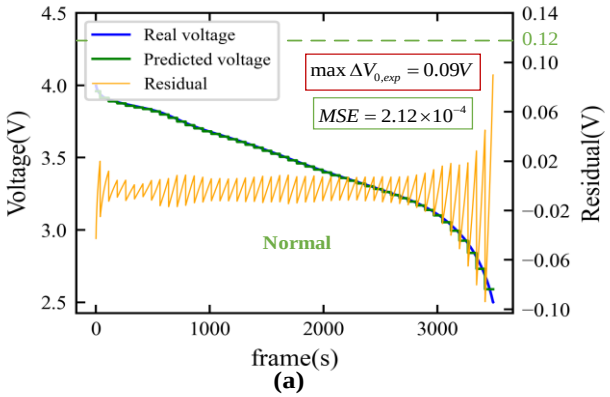


Fig. 12. Discharge voltage prediction result (a) Cell No.0. (b)-(f) Cell No.1-5.

reversing to a negative value. Finally, the ΔV curve increases shapely since the discharge voltage approaches to 2.5V and exceeds the first threshold, which indicates that the battery has been over-discharged slightly in the past. Besides, the ΔV curves of Cell No.2 and No.3 in Fig. 12(c) and Fig. 12(d) have the similar transform tendency, which exceeds the first threshold and keeps increasing until reaching a peak voltage. Then, the ΔV curves decrease slightly, following a monotonic gradual increase without fluctuation. Thus, the over-discharge alarm could be triggered at least 3 times and when the last alarm is triggered, the discharge voltage is away from the normal. Because the point D in Fig. 3 (Cell No.3) is the demarcation point of slight over-discharge and deep over-discharge, the $\max\Delta V_{3,exp}$ is selected to be the second threshold. To figure out the influence of DOD on over-discharge fault diagnosis method, the voltage and ΔV curves of cell No.4 and No.5 are presented in Fig. 12(e) and Fig. 12(f), respectively. It's notable that the ΔV curves increase monotonically and exceed the second threshold. As mentioned above, the point E in Fig. 3 is the critical point between deep over-discharge and extreme over-discharge. Therefore, the $\max\Delta V_{4,exp}$ is chosen to be the third threshold of the over-discharge process. Once the ΔV is over the $\max\Delta V_{4,exp}$, the alarm of the third threshold will be trigged during its normal discharge process. As for Cell No.5, the DOD of which is 108.3%, the alarm of the third threshold is trigged, which indicates that the extreme over-discharge had happened in the past. In a word, based on the over-discharge fault diagnosis method in this paper, the different degrees of over-discharge happened in the past could be diagnosed, which will avoid cycle aging and thermal runaway of battery.

However, the data acquisition errors of the voltage sensor in the experimental environment are smaller than those in real-world EVs, resulting in the difference in prediction accuracy and residual value. It's notable that the accuracy of diagnosis result is strongly related to the residual value. The maximum voltage theoretic error of battery testing instrument is about 0.35mV, whereas for vehicle No.0 to No.4, the voltage error is 20mV. In light of that, the voltage accuracy of experimental data is adopted to the same of EVs and the threshold of fault diagnosis method is recalculated for real-world application. As shown in Fig. 13, the MSE of the adjusted method is 2.12×10^{-4}. Moreover, the threshold between normal discharge and slight over-discharge is 0.12V, which is calculated by the maximum ΔV of Cell No.0 and No.1.



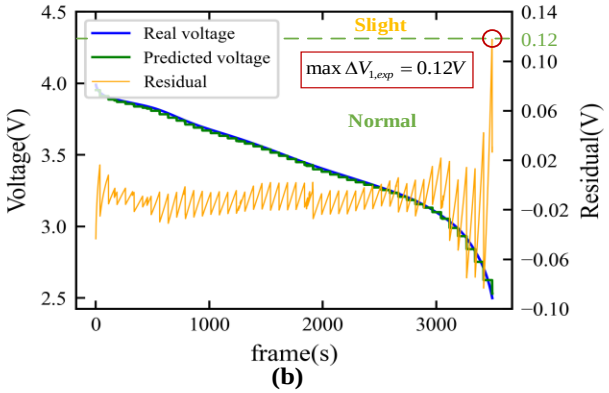


Fig. 13. Discharge voltage prediction result when the voltage error is 20mV (a) Cell No.0. (b) Cell No.1.

B. Application and verification in real-world EVs

It is hard to analyze the over-discharge fault that hidden beneath the cell voltage due to the strong nonlinearity, instability and inconsistency of lithium-ion batteries in real-world EVs [47]. In this paper, the discharge segments of vehicles No.0 to No.4 are employed to verify the validity of the diagnosis method that is established by experimental data.

The discharge data of normal vehicle No.0 are used to train the XGBoost model for real-world EVs. Since the input features of real-world EVs and experiment are slightly different, the features' importance after normalization of the two data sources is shown in Fig. 14.

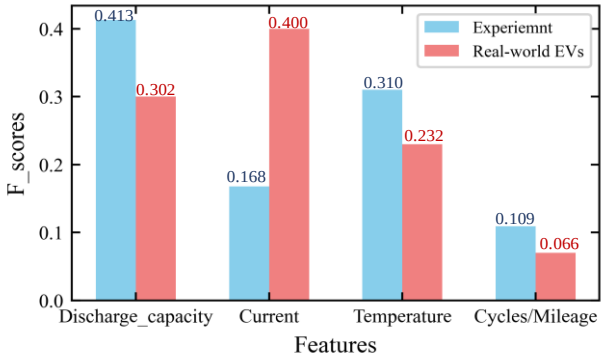


Fig. 14. Features' importance of experimental data and real-world EVs data.

As Fig. 14 illustrates, the discharge capacity based on experimental data has the highest F score. On the contrary, the F score of current is higher than discharge capacity from real-world EVs data. It can be explained as the current of real-world EVs fluctuates violently during the operation, whereas it is constant in the experimental tests. Therefore, the current of EVs has a great influence on voltage prediction model. At the same time, it is proved that the voltage prediction method is able to adapt the real-world EVs operation.

After size selection of training data and parameter optimization, the MSE is 8.21×10^{-5} and the maximum residual is 0.09V. It is almost consistent with the result which is calculated by the adjusted experimental data.

As mentioned above, 5 vehicles retrieved from NMMC-NEV are employed to verify the reliability and robustness of diagnosis method in this paper and the results of vehicle No. 0 to No. 4 are shown in Fig. 15 to Fig. 17.

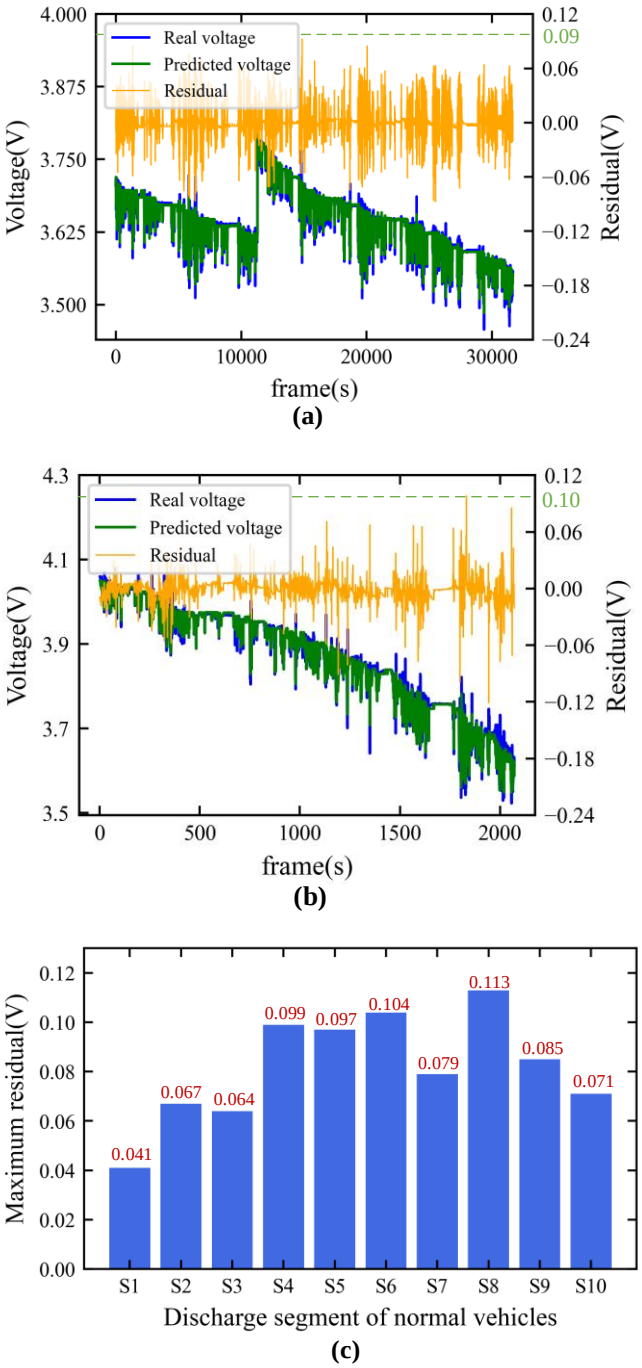
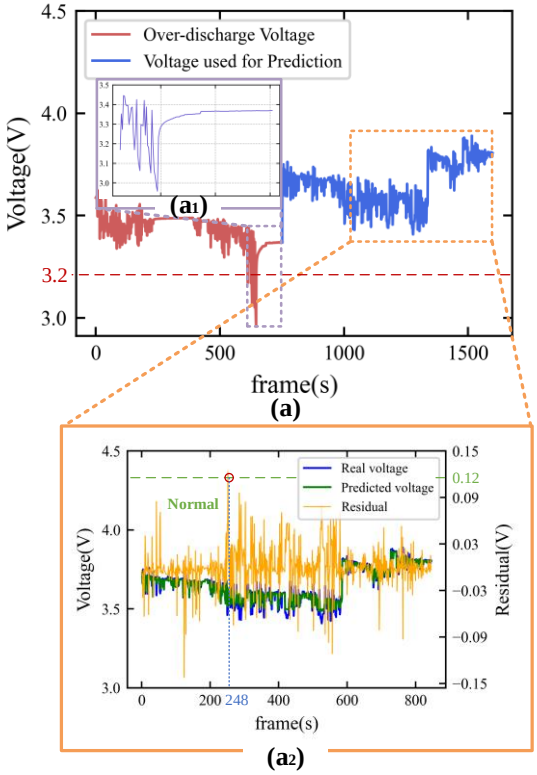


Fig. 15. The prediction results of normal vehicle (a) No.0 and (b) No.1; (c) The maximum residual of 10 discharge segments obtained from two normal vehicles.

The testing data of those vehicles are selected from one single discharge segment. Combined with the aforementioned threshold, the normal vehicles prediction and diagnosis results are depicted in Fig. 15(a) and Fig. 15(b). In both figures, the ΔV curves are almost fluctuated symmetrically around 0V. Besides, in Fig. 15(c), there are ten discharge segments obtained by vehicle No.0 and No.1 and the maximum ΔV of which are below 0.12V. This serves as an example to attest that the proposed fault diagnosis method will not trigger alarm falsely, which can verify the reliability of the method.

Furthermore, based on the over-discharge fault vehicle No.2 and No.3, the fault alarm will be triggered and the over-discharge failure happened in the past can be detected as shown in Fig. 16(a) and Fig. 16(b). The discharge segment voltage curve of vehicle No.2 illustrated in Fig. 16(a) are divided into over-discharge part and predicted part. The predicted result of discharge segment for prediction is illustrated in Fig. 16(a_2) and the predicted residual of vehicle No.2 has exceeded 0.12V (slight over-discharge line). Besides, the alarm time of vehicle No.2 is 14:24:40 2019-01-11 (248 frames) and the time when the over-discharge occurred is 09:31:40 2019-01-11 as shown in Fig. 16(a_1). In addition, as for vehicle No.3, the residuals have exceeded slight over-discharge alarm line more than 50 times shown in Fig. 16(b_2), which indicates that the slight over-discharge has happened before. In Fig. 16(b) and 16(b_1), we can find that the slight over-discharge has happened in 10:38:36 2019-01-07. The predicted results of both faulty vehicles verify the robustness of the proposed method. As illustrated in Fig. 16(a) and Fig. 16(b), the unabridged voltage curves of those two vehicles indicate that the slight over-discharge have happened and this method can detect the over-discharge fault by the subsequent segment.



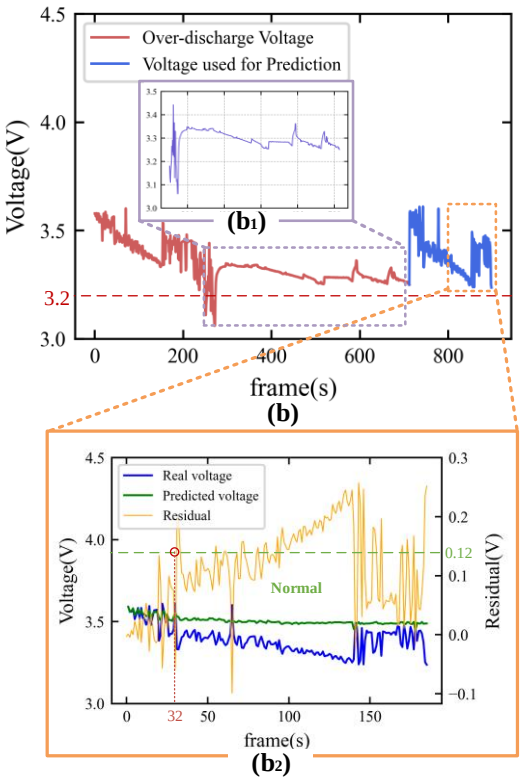


Fig. 16. The unabridged voltage curves of selected discharge segment collected by vehicle (a) No.2 and (b) No.3. The over-discharge part in selected segment of vehicle (a1) No.2 and (b1) No.3. The prediction results of over-discharge vehicle (a2) No.2 and (b2) No.3.

To verify the necessity of outlier elimination based on boxplot, vehicle No.4 with severe inconsistency of cells voltage but no over-discharge failure is employed, the prediction result of two discharge segments are illustrated in Fig. 17. If there is no outlier elimination, the voltage of certain cells may deviate from the normal values. It can lead to the inaccuracy of median voltage calculated by all cells, which is why maximum ΔV exceeds the threshold, as shown in Fig. 17(a) and Fig. 17(c). However, after outlier elimination based on boxplot, the ΔV curves keep below the threshold from beginning to end without false positives. It proves that outlier elimination of real-world EVs data is important for fault diagnosis.

C. Discussion the influence of data acquisition error

As mentioned above, the voltage acquisition error of experimental data is smaller than that of real-world EVs. Therefore, it is necessary to adjust the accuracy of the obtained voltage data in experimental tests to fit the practical vehicular operation. Generally, according to the statistics from NMMC-NEV, there are two types of voltage acquisition error in real-world EVs, which are 20mV and 1mV, respectively. The practical vehicular diagnosis threshold, the error of which is 20mV, has been discussed. Therefore, the real-world EVs diagnosis threshold of 1mV will be discussed in this part. Since the voltage accuracy of the experimental data has been adjusted to 1mV, the threshold which can be used to distinguish the normal discharge and over-discharge state is obtained. The value of the adapted threshold is 0.065V, which is almost 2.17 times as the original value. As shown in Fig. 18(a) and Fig.

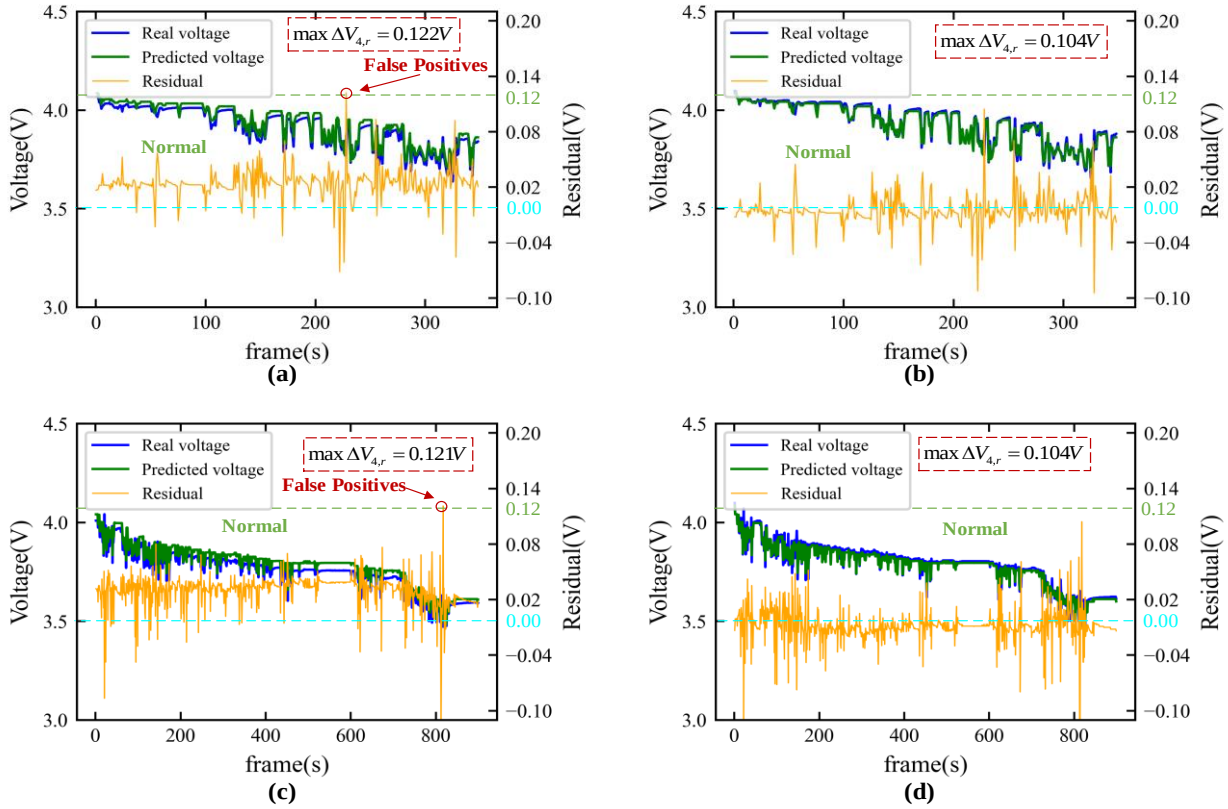


Fig. 17. The prediction result of vehicle No.4 from selected discharge segment (a) 1# and (c) 2# without outlier elimination; The prediction result from selected discharge segments (b) 1# and (d) 2# with outlier elimination utilizing boxplot.

18(b), the adapted threshold is captured by the prediction results of Cell No.0 and Cell No.1.

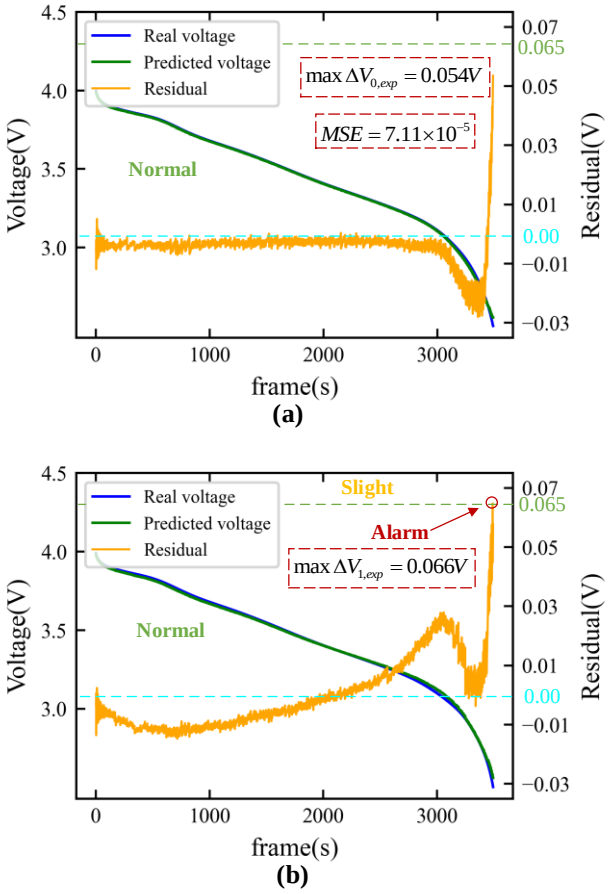
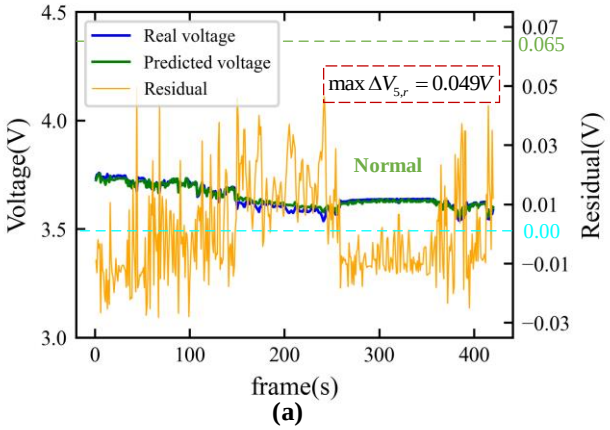


Fig. 18. The prediction results when the voltage data accuracy of 1mV for (a) Cell No.0 and (b) Cell No.1.

After that, the normal vehicles No.5 to No.7 with the error of 1mV are employed to verify the reliability of the new threshold. 10 discharge segments from these 3 vehicles are captured and the maximum ΔV are calculated, the results of which are illustrated in Fig. 19(a) and Fig. 19(b).



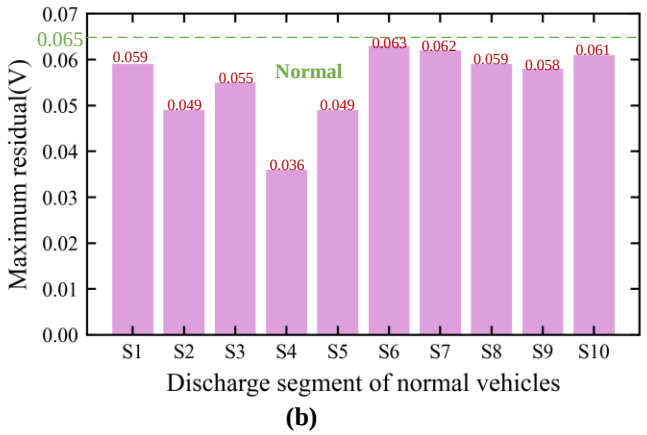


Fig. 19. The prediction result with voltage acquisition error of 1mV for (a) the one discharge segment from vehicle No.5; (b) 10 discharge segments from vehicle No.5 to No.7.

We can find that, all results of discharge segments are under the new threshold, according to the proposed diagnosis method. This serves as an example to illustrate that the proposed fault diagnosis method can well-adapted to the real-world EVs data with different voltage acquisition errors. It also confirms that the voltage acquisition error has a significant influence on the setting of threshold. In order to apply the threshold to practical vehicular diagnosis, an adaptive threshold method is proposed in this paper.

TABLE V
THE RELATIONSHIP BETWEEN VOLTAGE ACQUISITION ERROR AND CORRECTION FACTOR

Voltage acquisition error (V)	Threshold (V)	Correction factor ω
0.001	0.065	2.17
0.02	0.12	4

The correction factor ω can be used to calculate the threshold of fault diagnosis method, which can detect the over-discharge fault of real-world EVs with different voltage errors. The ω is given by

$$\omega = \frac{R_{E_i,r}}{R_{E_0,exp}} \tag{9}$$

Where $R_{Ei,r}$ represents the threshold between normal discharge and over-discharge in real-world EVs when the voltage acquisition error is E_i; $R_{E0,exp}$ is the threshold calculated by experimental data with the initial voltage accuracy. The value of voltage error, threshold and correction factor are elucidated in Table V.

And the relationship between correction factor and voltage error are fit by linear function, which is given by

$$\omega = 96.5E + 2.07 \tag{10}$$

Where E is the voltage acquisition error of real-world EVs. For different types of Lithium-ion battery, the threshold of fault diagnosis method can be obtained by slight over-discharge experiment directly. According to (10), the correction factor with different errors can be calculated to get the over-discharge fault diagnosis threshold of real-world EVs.

V. CONCLUSION

This paper presents a two-layer over-discharge fault

diagnosis method for LIBs in EVs based on XGBoost algorithm. The data from lab experiments with high accuracy and from real-world EVs are combined to establish the fault diagnosis model. The features that have influence on voltage during the discharge process are selected as the characteristic parameters in XGBoost. The residual of real-time voltage and predicted voltage is employed to be the basis to set the diagnostic threshold. The prediction accuracy represented by MSE, which is obtained from normal discharge, is 3.46×10^{-5}. Three thresholds are determined by the experimental over-discharge data to divide the different degrees of over-discharge. After adjusting the lab experiment voltage accuracy, the proposed approach greatly fits to the real EVs' over-discharge diagnosis. A large quantities of operation data collected from normal and faulty EVs are utilized to validate the feasibility of the proposed approach. The result shows that the proposed approach can effectively detect an over-discharge happening in the past. Therefore, the proposed method is a very useful tool in health diagnosis of batteries during EVs maintenance, service and operation.

REFERENCES

[1] J. Du and D. Ouyang, "Progress of Chinese electric vehicles industrialization in 2015: A review," *Applied Energy*, vol. 188, pp. 529-546, 2017.

[2] J. Wen, D. Zhao, and C. Zhang, "An overview of electricity powered vehicles: Lithium-ion battery energy storage density and energy conversion efficiency," *Renewable Energy*, vol. 162, pp. 1629-1648, 2020.

[3] Z. Jianan, Z. Lei, S. Fengchun, and W. Zhenpo, "An Overview on Thermal Safety Issues of Lithium-ion Batteries for Electric Vehicle Application," *IEEE Access*, vol. 6, pp. 1331-46, 2018.

[4] J.-N. Shen, J.-J. Shen, Y.-J. He, and Z.-F. Ma, "Accurate State of Charge Estimation with Model Mismatch for Li-Ion Batteries: A Joint Moving Horizon Estimation Approach," *IEEE Transactions on Power Electronics*, vol. 34, no. 5, pp. 4329-4342, 2019.

[5] M. A. Hannan, M. S. H. Lipu, A. Hussain, and A. Mohamed, "A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: challenges and recommendations," *Renewable & Sustainable Energy Reviews*, vol. 78, pp. 834-54, Oct. 2017.

[6] W. Qingsong, P. Ping, Z. Xuejuan, C. Guanquan, S. Jinhua, and C. Chunhua, "Thermal runaway caused fire and explosion of lithium ion battery," *Journal of Power Sources*, vol. 208, pp. 210-24, Jun. 2012.

[7] J. Zhang and J. Lee, "A review on prognostics and health monitoring of Li-ion battery," *Journal of Power Sources*, vol. 196, no. 15, pp. 6007-6014, 2011.

[8] B. K. Mandal, A. K. Padhi, Z. Shi, S. Chakraborty, and R. Filler, "Thermal runaway inhibitors for lithium battery electrolytes," *Journal of Power Sources*, vol. 161, no. 2, pp. 1341-1345, 2006.

[9] M. Chen, Y. He, C. De Zhou, Y. Richard, and J. Wang, "Experimental Study on the Combustion Characteristics of Primary Lithium Batteries Fire," *Fire Technology*, vol. 52, no. 2, pp. 365-385, 2016.

[10] S. Dey, Z. A. Biron, S. Tatipamula, N. Das, S. Mohon, B. Ayalew and P. Pisu, "Model-based real-time thermal fault diagnosis of Lithium-ion batteries," *Control Engineering Practice*, vol. 56, pp. 37-48, Nov. 2016.

[11] 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," *Journal of Energy Storage*, vol. 18, pp. 26-39, 2018.

[12] R. Dongsheng, F. Xuning, L. Languang, H. Xiangming, and O. Minggao, "Overcharge behaviors and failure mechanism of lithium-ion batteries under different test conditions," *Applied Energy*, vol. 250, pp. 323-32, 2019.

[13] O. Dongxu, W. Jingwen, C. Mingyi, and W. Jian, "Impacts of Current Rates on the Degradation Behaviors of Lithium-Ion Batteries under Over-Discharge Conditions," *Journal of the Electrochemical Society*, vol. 166, no. 14, pp. 3432-40, 2019.

[14] R. Guo, L. Lu, M. Ouyang, and X. Feng, "Mechanism of the entire overdischarge process and overdischarge-induced internal short circuit in lithium-ion batteries," *Scientific Reports*, vol. 6, p. 30248, 2016.

[15] C. Fear, D. Juarez-Robles, J. A. Jeevarajan, and P. P. Mukherjee, "Elucidating Copper Dissolution Phenomenon in Li-ion Cells Under Overdischarge Extremes," *Journal of the Electrochemical Society*, vol. 165, no. 9, pp. 1639-47, 2018.

[16] W. Chao, Z. Chunbo, S. Jinlei and J. Jianhu, "Fault mechanism study on Li-ion battery at over-discharge and its diagnosis approach," *IET Electrical Systems in Transportation*, vol.7. no.1,pp. 48-54,2017.

[17] L. Languang, H. Xuebing, L. Jianqiu, H. Jianfeng, and O. Minggao, "A review on the key issues for lithium-ion battery management in electric vehicles," *Journal of Power Sources*, vol. 226, pp. 272-88, Mar. 2013.

[18] 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, 2015.

[19] Z. Zehan, L. Shuanghong, X. Yawen, and Y. Yupu, "Intelligent simultaneous fault diagnosis for solid oxide fuel cell system based on deep learning," *Applied Energy*, vol. 233-234, pp. 930-42, Jan. 2019.

[20] H. Xiaosong, L. Wenxue, L. Xianke, and X. Yi, "A Comparative Study of Control-oriented Thermal Models for Cylindrical Li-ion Batteries," *IEEE Transactions on Transportation Electrification*, vol. 5, no. 4, pp. 1237-53, Dec. 2019.

[21] C. Zeyu, X. Rui, T. Jinpeng, S. Xiong, and L. Jiahuan, "Model-based fault diagnosis approach on external short circuit of lithium-ion battery used in electric vehicles," *Applied Energy*, vol. 184, pp. 365-74, Dec. 2016.

[22] S. Jeongeun and D. Yuncheng, "Model-Based Stochastic Fault Detection and Diagnosis of Lithium-Ion Batteries," *Processes*, vol. 7, no. 1, p. 38 (19 pp.), Jan. 2019.

[23] W. Yujie, T. Jiaqiang, C. Zonghai, and L. Xingtao, "Model based insulation fault diagnosis for lithium-ion battery pack in electric vehicles," *Measurement*, vol. 131, pp. 443-51, Jan. 2019.

[24] W. Jingwen, D. Guangzhong, and C. Zonghai, "Model-based fault diagnosis of Lithium-ion battery using strong tracking Extended Kalman Filter," *Energy Procedia*, vol. 158, pp. 2500-5, Feb. 2019.

[25] 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. 1002-11, Feb. 2015.

[26] G. Xianzhi, X. Rui, and C. C. Mi, "A Data-Driven Bias-Correction-Method-Based Lithium-Ion Battery Modeling Approach for Electric Vehicle Applications," *IEEE Transactions on Industry Applications*, vol. 52, no. 2, pp. 1759-65, Mar. 2016.

[27] X. Bing, S. Yunlong, N. Truong, and C. Mi, "A correlation based fault detection method for short circuits in battery packs," *Journal of Power Sources*, vol. 337, pp. 1-10, Jan. 2017.

[28] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2560-2574, Aug. 2007.

[29] K. Jonghoon, "Discrete wavelet transform-based feature extraction of experimental voltage signal for Li-Ion cell consistency," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1150-61, Mar. 2016.

[30] K. Yongzhe, D. Bin, Z. Zhongkai, S. Yunlong, and Z. Chenghui, "A multi-fault diagnostic method based on an interleaved voltage measurement topology for series connected battery packs," *Journal of Power Sources*, vol. 417, pp. 132-44, Mar. 2019.

[31] L. Xiaoyu, D. Kangwei, W. Zhenpo, and H. Weiji, "Lithium-ion batteries fault diagnostic for electric vehicles using sample entropy analysis method," *Journal of Energy Storage*, vol. 27, p. 101121 (11 pp.), Feb. 2020.

[32] L. Peng, S. Zhenyu, W. Zhenpo, and Z. Jin, "Entropy-Based Voltage Fault Diagnosis of Battery Systems for Electric Vehicles," *Energies*, vol. 11, no. 1, p. 136 (15 pp.), Jan. 2018.

[33] Z. Yang, L. Peng, W. Zhenpo, Z. Lei, and H. Jichao, "Fault and defect diagnosis of battery for electric vehicles based on big data analysis methods," *Applied Energy*, vol. 207, pp. 354-62, Dec. 2017.

[34] 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.

[35] H. Jichao, W. Zhenpo, and Y. Yongtao, "Fault prognosis of battery system based on accurate voltage abnormity prognosis using long short-term memory neural networks," *Applied Energy*, vol. 251, pp. 296-309, Oct. 2019.

[36] Z. Ruxiu, P. J. Kollmeyer, R. D. Lorenz, and T. M. Jahns, "A compact unified methodology via a recurrent neural network for accurate modeling of lithium-ion battery voltage and state-of-charge," in *2017 IEEE Energy Conversion Congress and Exposition (ECCE), 1-5 Oct. 2017*, Piscataway, NJ, USA, 2017: IEEE, in 2017 IEEE Energy Conversion Congress and Exposition (ECCE), pp. 5234-41.

[37] L. Da, Z. Zhaosheng, L. Peng, W. Zhenpo, and Z. Lei, "Battery Fault Diagnosis for Electric Vehicles Based on Voltage Abnormality by Combining the Long Short-Term Memory Neural Network and the Equivalent Circuit Model," *IEEE Transactions on Power Electronics*, vol. 36, no. 2, pp. 1303-15, Feb. 2021.

[38] Z. P. Wang, S. Q. Xu, X. Q. Zhu, H. Wang, L. W. Huang, J. Yuan and W. Q. Yang, "Effects of short-term over-discharge cycling on the performance of commercial 21,700 lithium-ion cells and the identification of degradation modes," *Journal of Energy Storage*, vol. 35, 2021.

[39] Y. Hao, W. Zhenpo, L. Peng, Z. Zhaosheng, and L. Yang, Voltage Fault Diagnosis of Power Batteries based on Boxplots and Gini Impurity for Electric Vehicles. Presented at *2019 Electric Vehicles International Conference (EV), 3-4 Oct. 2019*, Piscataway, NJ, USA, 2019: IEEE, in 2019 Electric Vehicles International Conference (EV), p. 5 pp.

[40] T. Chen and C. Guestrin, XGBoost: A scalable tree boosting system. Presented at *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2016, August 13, 2016 - August 17, 2016*, San Francisco, CA, United states, 2016, vol. 13-17-August-2016: Association for Computing Machinery, in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794.

[41] Q. Tang, G. Xia, X. Zhang, and F. Long, "A Customer Churn Prediction Model Based on XGBoost and MLP," in 2020 International Conference on Computer Engineering and Application, ICCEA 2020, March 27, 2020 - March 29, 2020, Guangzhou, China, 2020: Institute of Electrical and Electronics Engineers Inc., in Proceedings - 2020 International Conference on Computer Engineering and Application, ICCEA 2020, pp. 608-612.

[42] A. Singh, A. Dhillon, N. Kumar, M. S. Hossain, G. Muhammad, and M. Kumar, "eDiaPredict: An Ensemble-based Framework for Diabetes Prediction," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 17, no. 2s, 2021.

[43] S. Ji, X. Wang, W. Zhao, and D. Guo, "An application of a three-stage XGboost-based model to sales forecasting of a cross-border e-commerce enterprise," *Mathematical Problems in Engineering*, 2019.

[44] A. Galicia, R. Talavera-Llames, A. Troncoso, I. Koprinska, and F. Martinez-alvarez, "Multi-step forecasting for big data time series based on ensemble learning," *Knowledge-Based Systems*, vol. 163, pp. 830-841, 2019.

[45] Y. Senyan, W. Jianping, D. Yiman, H. Yingqi, and C. Xu, "Ensemble Learning for Short-Term Traffic Prediction Based on Gradient Boosting Machine," *Journal of Sensors*, vol. 2017, p. 7074143 (15 pp.), 2017.

[46] S. Song, C. Fei, and H. Xia, "Lithium-Ion Battery SOH Estimation Based on XGBoost Algorithm with Accuracy Correction," *Energies*, vol. 13, no. 4, 2020.

[47] F. Fei, H. Xiaosong, H. Lin, H. Fengling, L. Yang, and Z. Lei, "Propagation mechanisms and diagnosis of parameter inconsistency within Li-Ion battery packs," *Renewable & Sustainable Energy Reviews*, vol. 112, pp. 102-13, 2019.