This article has been accepted for publication in IEEE Transactions on Energy Conversion. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TEC.2026.3686467

Adaptive Pump Fault Detection for Vanadium Redox Flow Batteries Based on CNN-BiGRU and Transfer Learning

Binyu Xiong, *Senior Member, IEEE*, Shaojin Wang, Jinsong Wang, Peng Zhou, Quan Zhou, Changjun Xie, *Member, IEEE*, Yang Li, *Senior Member, IEEE*, and Peng Wang, *Fellow*, *IEEE*

***Abstract*—Vanadium redox flow batteries (VRBs) are vulnerable to pump malfunctions, which can cause severe problems like overcharging, overdischarging, and irreversible damage to the battery stack. Effective real-time fault detection methods for VRBs are essential but underdeveloped. This study proposes a data-driven approach for VRB pump fault detection using a convolutional neural network combined with bi-directional gated recurrent units to predict voltage based on measurable current, flow rate, and state of charge without relying on complex electrochemical models. To address the impact of battery aging and system variability, which are often overlooked in the literature, an offline transfer learning strategy with model tail fine-tuning is employed to adapt the parameters of the pre-trained model using a small amount of newly collected data. Pump faults are identified by comparing voltage prediction errors against a threshold derived from the 3σ criterion. Experimental results across six fault scenarios encompassing three distinct pump fault modes under two operational conditions demonstrate that the proposed method achieves timely and accurate fault detection, significantly improving the reliability of VRB systems.**

***Index Terms*—Vanadium redox flow battery; CNN-BiGRU; Fault diagnosis; transfer learning.**

I. Introduction

As an emerging energy storage technology, vanadium redox flow batteries (VRB) have shown great potential in renewable energy integration, grid power dispatch, and residential applications thanks to their long cycle life, high safety, and flexible power/capacity configuration [1]. However, VRBs are prone to various malfunctions during operation, such as proton exchange membrane damage, electrode corrosion, and pump faults [2]. Among them, pump faults are particularly critical, as they can reduce system efficiency and cause irreversible damage, leading to significant economic losses [3]. Therefore, developing an effective pump fault detection method is essential for ensuring the safe and reliable operation of VRBs.

In the broader context of battery systems, most fault analysis studies have focused on lithium-ion batteries, with relatively limited work addressing VRBs. Existing studies typically refer to fault diagnosis, involving not only fault detection but also fault identification and classification. Fault diagnosis methods for batteries can generally be categorized into model-based and data-driven approaches. *Model-based approaches* rely on physical mechanisms to estimate parameters such as voltage, state of charge (SOC), and temperature, which identify faults like short circuits, thermal runaway, and sensor failures through residual analysis. For example, Yang *et al.* [4] and Xu et al. [5] used extended Kalman filters (EKF) and the H_∞ observers to detect soft short circuits by comparing the estimated SOC against Coulomb counting results. Kong *et al.* [6] developed a pseudo-two-dimensional model for early micro-internal short circuits in lithium-ion batteries via impedance analysis. Roy *et al.* [7] proposed a real-time thermal runaway diagnosis method based on partial differential equation modeling, where battery faults were determined based on temperature residuals. Model-based approaches have also been investigated for sensor failure diagnosis, as seen in Lin *et al.* [8] applying a dual EKF to identify sensor and hysteresis faults in battery packs. Sepasiahooyi *et al*. [9] designed two model-based adaptive observers to detect new and aged lithium-ion battery SOC faults and voltage sensor faults. Although physically interpretable, model-based methods are often complex and computationally intensive, which poses challenges for real-time deployment, especially for highly nonlinear multiphysical systems such as VRBs.

In contrast, *data-driven approaches* eliminate the need for complex modeling and have gained significant attention in recent studies. These methods can be generally classified as those constructing fault classifiers based on fault data and those using neural network algorithms to predict battery behaviors for fault diagnosis. In classifier-based methods, a labeled fault dataset is first acquired and used to train machine learning models. For example, Yao *et al.* [10] applied discrete cosine filtering to denoise signals and used a support vector machine (SVM) to classify connection faults with an accuracy of over 97%. Jan *et al.* [11] similarly employed SVMs to classify artificially injected sensors. Ma *et al.* [12] introduced a graph-based autoencoder for sensor fault classification across different fault types. For VRBs, Qin *et al.* [13] developed an SVM-based method that uses the internal resistance change rates as a feature for pump fault detection, although its applicability under varying current profiles is limited due to the impact of current step changes on internal resistance.

These classifier-based methods, which typically rely on measured data for fault diagnosis, require constructing real fault datasets, a process that is often time-consuming and hazardous due to the difficulty of generating many battery failure types. Nevertheless, an alternative approach is to leverage predicted

This work is supported by the National Natural Science Foundation of China under grant no. 52177221. (*Corresponding author*: *Yang Li*).

Binyu Xiong, Shaojin Wang, Peng Zhou, Quan Zhou, and Changjun Xie are with School of Automation, Wuhan University of Technology, Wuhan 430070, China. (email: bxiong2@whut.edu.cn; shaojinwang@whut.edu.cn; 334938@whut.edu.cn; zhouquan910@whut.edu.cn; jackxie@whut.edu.cn).

Jinsong Wang is with HyperStrong Technology Co., LTD, Beijing, China. (email: wangjinsong@hyperstrong.com).

Yang Li is with the School of Electrical Engineering and Automation, Wuhan University, Wuhan 430070, China. (email: yangli@ieee.org).

Peng Wang is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: epwang@ntu.edu.sg).

data for fault diagnosis. For example, Hong *et al.* [14] used long short-term memory (LSTM) networks to predict battery voltage trajectories and assess operational safety via alarm thresholds with the consideration of external factors such as weather and driver behavior. Li *et al.* [15] further improved the prediction accuracy by integrating LSTM with an equivalent-circuit model. Ojo *et al.* [16], [17] employed LSTM to predict battery temperature and voltage for thermal fault and voltage anomaly detection. Li *et al.* [18] developed a CNN-LSTM neural network to predict battery temperature with the capability to diagnose thermal runaway up to 27 minutes in advance. Furthermore, to address the scarcity of labeled fault data, Luder *et al*. [19] proposed a big data generation platform to simulate and generate battery fault data, thereby significantly accelerating the development of the fault diagnosis industry.

Although significant progress has been made in battery fault diagnosis, several challenges remain unsolved for VRBs. First, existing VRB neural network models are applied directly after offline training, lacking an efficient mechanism for updating parameters with online adaptability. While Zhao *et al.* [20] proposed an incremental learning method, the high computational cost is a concern, as retraining the entire dataset is needed for real-world applications. Second, many studies rely on artificially induced faults, such as injecting faulty voltage signals and simulated thermal faults using heating pads [21], [22], [23], [24], which may not capture the complex dynamics of real-world VRB faults. Third, physically interpretable models, while valuable, often suffer from lower prediction accuracy and limited real-time applicability. Finally, Basic statistical models or simpler baseline approaches fail to cope with the time-varying and strongly coupled fault characteristics under complex operating conditions, resulting in numerous false alarms and missed detections.

In view of the above, this paper focuses on pump fault detection for VRBs rather than full-scale fault diagnosis. A voltage prediction model is developed based on a convolutional neural network with bidirectional gated recurrent units (CNN-BiGRU). First, the underlying mechanisms of VRB pump faults are mathematically analyzed, and the effects of different pump fault types on voltage behavior are systematically compared. In addition, the external characteristics of several common VRB faults are analyzed to clarify the physical mechanisms behind voltage anomalies. Next, an offline pre-training CNN-BiGRU model is established to predict the VRB voltage. To enhance adaptability under varying operating conditions, a transfer learning strategy, referred to as model tail fine-tuning, is adopted to update model parameters using a small amount of newly collected operational data. Pump faults are then detected by monitoring deviations between the predicted and measured voltages, with faults identified once the model error exceeds an optimal threshold.

The main contributions of this study include:

1) A highly accurate VRB voltage prediction model is developed using a CNN-BiGRU neural network, eliminating the need for complex electrochemical modeling.

2) A novel pump fault detection framework is proposed, which identifies faults by comparing the prediction error of the neural network model with a preset threshold.

3) A model fine-tuning approach is introduced, which enables the pre-trained model to adapt efficiently to new battery characteristics and operating conditions, thereby enhancing fault detection performance and transferability.

The rest of this paper is organized as follows: Section II presents the experimental setup and dataset, and describes the observed fault behaviors of VRBs. Section III analyzes these fault phenomena based on the fundamental operating principles of VRBs. Section IV details the proposed CNN-BiGRU-based VRB fault detection method with model tail fine-tuning. Section V evaluates the model's accuracy and the effectiveness of the fault detection approach. Concluding remarks are summarized in Section VI.

II. EXPERIMENTAL SETUP AND DATASET

A. Experimental Platform

To obtain operating data, a single-cell VRB experimental platform is established, as shown in Fig. 1. The single-cell VRB system (Wuhan Zhisheng New Energy Co., Ltd.) consists of a single cell, two pumps, piping systems, two tanks, and a battery testing system (BTS). The stack with a single cell comprises end plates, insulation sheets, current collectors, bipolar plates, gaskets, proton exchange membranes, and two electrodes, with detailed cell parameters provided in Table I. The BTS is responsible for delivering charging and discharging currents while simultaneously collecting experimental data. All sensor data, including stack voltage, current, and flow rate, are logged at a sampling frequency of 1 Hz.

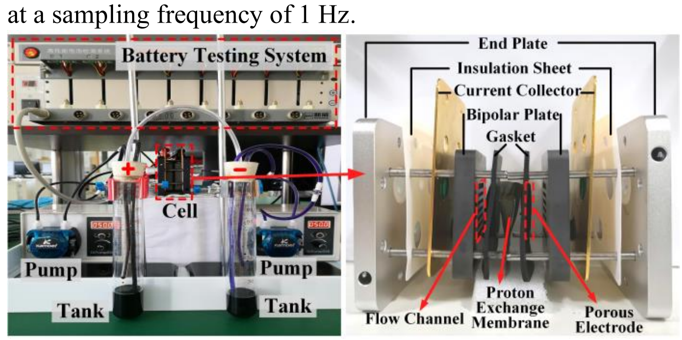


Fig. 1. VRB experimental platform.

TABLE I VRB SINGLE CELL PARAMETERS

Parameter	Value
Electrode size	3 cm×3 cm×0.435 cm
Membrane type	Nafion 212
Electrolyte volume	25 mL
Electrolyte concentration	1.7 mol/L
Nominal current density	200 mA/cm^2
Electrode cross-sectional area	1.305 cm^2 (3 cm×0.435 cm)
Active area	9 cm^2 (3 cm×3 cm)

B. Experimental Design

To comprehensively capture the operating characteristics of VRBs under different working conditions, five experimental scenarios are designed. Operation 1 is a single-cycle constant flow rate and constant current operation, which serves as the

This article has been accepted for publication in IEEE Transactions on Energy Conversion. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TEC.2026.3686467

baseline operating case. Operations 2 and 3 are carried out under a constant flow rate with variable current profiles. In Operation 2, the current increases linearly over time, whereas in Operation 3, the current decreases linearly. Operations 4 and 5 are designed to investigate the battery response under simultaneously varying flow rate and current conditions. In Operation 4, the current follows a stepwise variation, while in Operation 5, the current changes linearly.

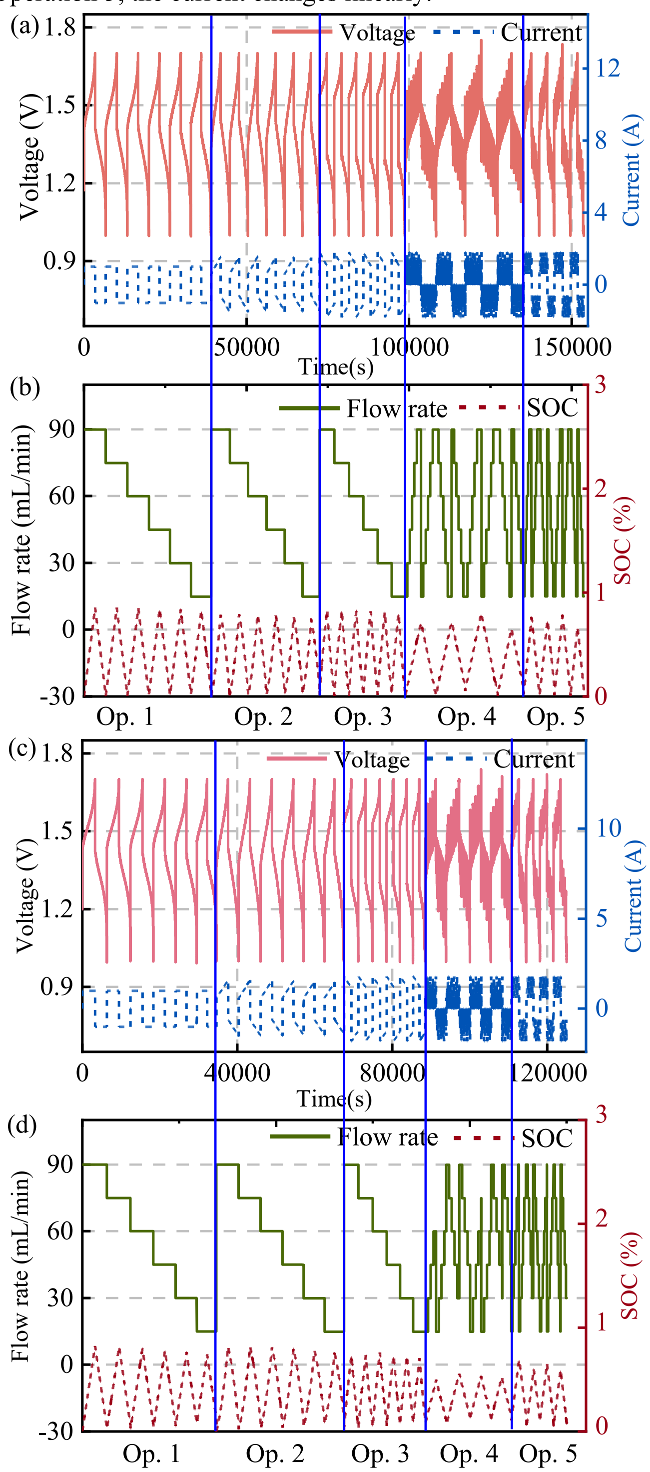


Fig. 2. VRB training set and sets. (a) Profiles of voltage and current for the training set; (b) Profiles of flow rate and SOC for the training set; (a) Profiles of voltage and current for the test set; (b) Profiles of flow rate and SOC for the test set.

Since the pump flow rate must be adjusted manually, the changes under the conditions described above are all step changes. Two batteries from the same batch of materials are selected for the experiments: data from Cell A is used for training, and data from Cell B serves as the test set. Detailed information about the training and test sets is shown in Fig. 2, with all data sampled at a 1-second interval.

The BTS cannot directly provide the SOC of the battery, so an estimation of the SOC is needed. It is crucial to ensure that the inputs to the model do not contain any incorrect information. While advanced SOC estimation methods, such as EKF and neural networks, offer more accurate estimates using measured battery voltage, they are more complex. Given that the SOC estimation is not the focus of the present study, the much simpler Coulomb counting method is used, i.e.,

$$\mathrm{SOC}(t) = \mathrm{SOC}_0 + \Delta\mathrm{Cap}(t) / \mathrm{Cap}_N, \tag{1}$$

$$\Delta\mathrm{Cap}(t) = \int_0^t I(\tau) d\tau, \tag{2}$$

where t is time, SOC_0 is the initial SOC, which can be obtained from the battery OCV-SOC curve, Cap_N is the nominal battery capacity, determined through charging and discharging experiments with small currents, and I denotes the applied current. Nominal charging and discharging capacities are 1107.175 mAh and 1070.925 mAh, respectively. We use their average value, 1089.05 mAh, as the nominal capacity ΔCap under alternating positive and negative current conditions.

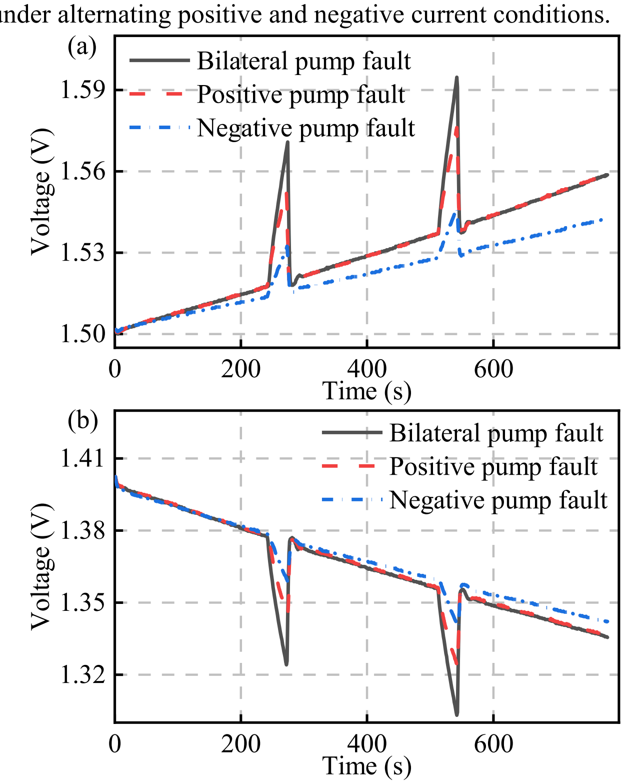


Fig. 3. Impact of different types of pump faults on voltage at 1000 mA charging and discharging current. (a) Charging; (b) Discharging.

To show the effect of pump faults on battery voltage, positive pump fault, negative pump fault, and bilateral pump fault were triggered during battery operation at 1000 mA charging and discharging currents, with the results shown in Fig. 3. It can be seen that the rate of voltage change increases when a pump fault occurs. A *positive pump fault*, in particular, causes a larger voltage deviation than a *negative pump fault*. In contrast, a *bilateral pump fault*, i.e., faults that simultaneously occur in both pumps, leads to the most significant voltage fluctuation. Once the fault is cleared, the electrolyte quickly flows from the tanks into the stack, allowing the voltage to return to its normal state rapidly. These observed phenomena of voltage variations will be further analyzed and explained in Section III.

III. ANALYSIS OF THE VRB PUMP FAULT MECHANISM

To investigate the mechanism of pump faults, it is essential to first explain the working principle of the VRB. Unlike lithium-ion batteries, VRB Numerical modelling and in-depth analysis of a multi-stack vanadium flow battery module incorporating transport delays utilize an external electrolyte that is actively pumped to circulate between the tanks and the stack. The governing equations that describe the variation of concentrations of vanadium ions in the stack can be derived based on mass conservation and charge conservation by ignoring side reactions and ion migration, i.e.,

$$V_t \frac{dc_i^t}{dt} = Q(c_i^c - c_i^t). \tag{3}$$

In a stack consisting of N_c cells, we have

$$\begin{aligned} V_c \frac{dc_i^c}{dt} &= Q(c_i^t - c_i^c) + \frac{N_c I}{zF} \\ &+ \frac{S}{d}(a_2 k_2 c_2^c + a_3 k_3 c_3^c + a_3 k_3 c_3^c + a_4 k_4 c_4^c), \end{aligned} \tag{4}$$

where V_t and V_c represent the volumes of the single tank and the half-cell, respectively, c_i^t and c_i^c denote the concentrations of vanadium ions with valence i (i = 2, 3, 4, and 5) in the tank and the cell, respectively, Q is the flow rate, N_c denotes the number of cell in the stack, I represents the applied current, defined as positive for charging and negative for discharging, S is the electrode area, d denotes the membrane thickness, a_i and k_i are the term coefficients and the diffusion coefficients of vanadium ions with valence i (i = 2, 3, 4, and 5), respectively. The values of a_i are shown in Table II.

On the right-hand side of (4), the first term is the effect of electrolyte flow rate, the second term is the role of the redox reaction (applied current), and the last term is the impact of ion crossover and self-discharge reactions (side reactions). While side reactions such as self-discharge do affect long-term capacity loss and open-circuit voltage decay, their impact on short-term voltage dynamics during operation (the focus of our fault detection model) is relatively minor. Therefore, the side reactions are assumed to be negligible in this work [25].

TABLE II VALUES OF TERM COEFFICIENTS

i	a_2	a_3	a_4	a_5
2	-1	0	-1	-2
3	0	-1	2	3
4	3	2	-1	0
5	-2	-1	0	-1

The terminal voltage of the VRB is composed of open-circuit voltage (OCV), ohmic losses, concentration overpotentials, and activation overpotentials, as calculated by

$$E_{\text{cell}} = E_{\text{OCV}} + IR_{\text{cell}} + \eta_{\text{con}} + \eta_{\text{act}}, \tag{5}$$

where η_{act} can be neglected due to the porous electrodes with a large effective surface area in VRBs, and R_{cell} denotes the cell ohmic resistance. E_{OCV} and η_{act} can be respectively expressed by

$$E_{\text{OCV}} = E_{\text{eq}} + \frac{RT}{zF}\ln\left(\frac{c_2^c c_5^c}{c_3^c c_4^c}\right), \tag{6}$$

and

$$\begin{aligned} \eta_{\text{con}} &= \eta_{\text{con}}^+ + \eta_{\text{con}}^- \\ &= \frac{RT}{zF}\ln\left(1 - \frac{i}{zFk_m c_r^+}\right) + \frac{RT}{zF}\ln\left(1 - \frac{i}{zFk_m c_r^-}\right), \end{aligned} \tag{7}$$

where E_{eq} is the formal potential, i denotes the current density, k_m is the local mass transfer coefficient, c_r^+ and c_r^+ are the molar concentrations of positive and negative reactants, respectively. They are calculated by [25]

$$k_m = 1.6 \times 10^{-3} (Q/S_{EC})^{0.4}, \tag{8}$$

$$c_r^+ = \begin{cases} c_4^c, & \text{charging} \\ c_5^c, & \text{discharging} \end{cases}, \tag{9}$$

$$c_r^- = \begin{cases} c_3^c, & \text{charging} \\ c_2^c, & \text{discharging} \end{cases}, \tag{10}$$

where S_{EC} is the electrode cross-sectional area.

Plugging (6) and (7) into (5) yields

$$\begin{aligned} E_{\text{cell}} &= E_{\text{eq}} + \frac{RT}{zF}\ln\left(\frac{c_4^c c_5^c}{c_2^c c_3^c}\right) + IR_{\text{cell}} \\ &+ \frac{RT}{zF}\ln\left(1 - \frac{i}{zFk_m c_r^+}\right) + \frac{RT}{zF}\ln\left(1 - \frac{i}{zFk_m c_r^-}\right). \end{aligned} \tag{11}$$

For a stack consisting of N_c individual cells, the stack voltage is the sum of the voltage of each cell, i.e., $E_{\text{stack}} = N_c \cdot E_{\text{cell}}$.

When a pump fault occurs, the electrolyte in the tanks fails to flow into the stack, resulting in $Q = 0$. Consequently, (4) can be simplified to

$$V_c \frac{dc_i^c}{dt} = \frac{N_c I}{zF}. \tag{12}$$

In this scenario, the electrolyte concentration in the stack cannot be diluted by the electrolyte from the tank. In addition, since the stack's volume is much smaller than that of the tank, the electrolyte concentration in the stack will change drastically during charging and discharging. According to the Nernst equation (6), these changes in electrolyte concentrations will lead to rapid fluctuations in battery voltage. Specifically, when the current is positive, the voltage increases, while it decreases when the current is negative. Since the positive standard potential is much larger than the negative standard potential, the impact of a positive pump fault on voltage is more pronounced than that of a negative pump fault. Furthermore, a bilateral pump fault, although rare, can cause the most significant voltage deviation, which can be understood as the combined effect of two individual single-side pump faults mentioned above. The analysis provides a clear explanation for the phenomenon of voltage variations observed in Fig. 3.

In practical VRB systems, voltage anomalies may arise from various fault mechanisms, not only pump faults. Pump faults mainly affect the electrolyte circulation and mass transport, leading to increased concentration polarization and a gradual voltage decrease or increase during load operation. In contrast, internal short circuits reduce the OCV and accelerate voltage decay during rest due to internal self-discharge. Poor electrical contact primarily increases the ohmic resistance of the system, resulting in instantaneous voltage drops when current is applied. These distinct external characteristics provide important clues for distinguishing different fault types.

IV. Method for VRB Pump Fault Detection

In this section, a method for pump fault detection is proposed based on the VRB fault mechanism, which provides timely and accurate alarms upon the occurrence of a pump fault without the need to identify the specific type of fault. The framework of the proposed method is illustrated in Fig. 4.

First, battery data are collected under three operating conditions as introduced in Section II: 1) constant flow rate and constant current; 2) constant flow rate and variable current; 3) variable flow rate and variable current. These data are then processed to train a proposed CNN-BiGRU, resulting in a pre-trained model. Model tail fine-tuning is subsequently applied to adapt the model to the current operating state of the battery. Finally, the pump faults are detected in real time by comparing the prediction error of the model with a preset threshold.

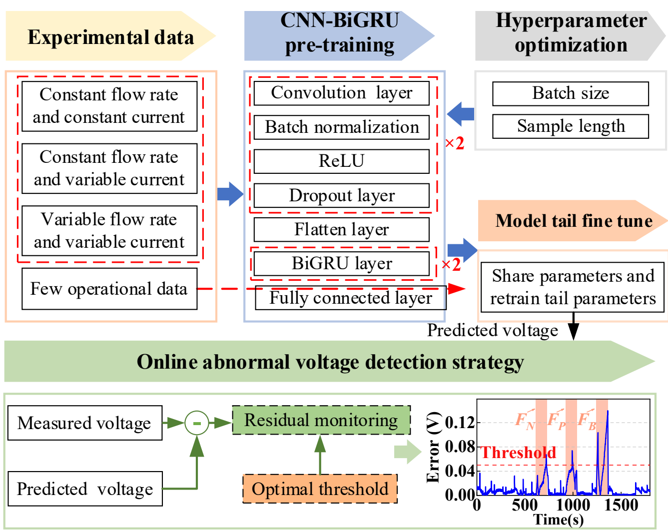


Fig. 4. Framework of the proposed pump fault detection method.

A. CNN-BiGRU Neural Network

1) One-Dimensional Convolution Neural Network

CNNs have been extensively used for prediction and classification tasks due to their powerful feature extraction capabilities and fast computational speed. CNNs use sliding windows to compute feature values within each window, enabling local feature extraction [26], [27]. The output $x_{j,\text{out}}^{l}$ of the l-th CNN layer is calculated by

$$x_{j,\text{out}}^{l}=\sum_{i\in m_j} x_i^{l-1}\cdot w_{ij}^{l}+b_j^{l}, \tag{13}$$

where x_i^{l-1} represents the input matrix, w_{ij}^{l} denotes the convolution kernel, b_j^{l} is the bias matrix, and m_j refers to the convolution computation objective.

CNNs can be classified into 1D-CNNs and 2D-CNNs, with 1D-CNNs typically used for sequential data and 2D-CNNs for spatial data. Since the input in the present work is 2D, a 1D-CNN is used. The structure of the 1D-CNN used in the proposed model is shown in Fig. 5(a). As illustrated, we employ two identical CNN layers, each consisting of a convolutional layer, a batch normalization (BN) layer, a ReLU activation function, and a dropout layer. The BN layer normalizes a small batch of samples, accelerating model training and enhancing stability. The ReLU activation function introduces nonlinearity into the model, which can improve the ability to detect complex patterns. The dropout layer randomly deactivates certain neurons during training to prevent overfitting. The output from the CNN layers is passed through a fully connected dense layer to adjust the dimensions, making them compatible with the BiGRU layer that is introduced next.

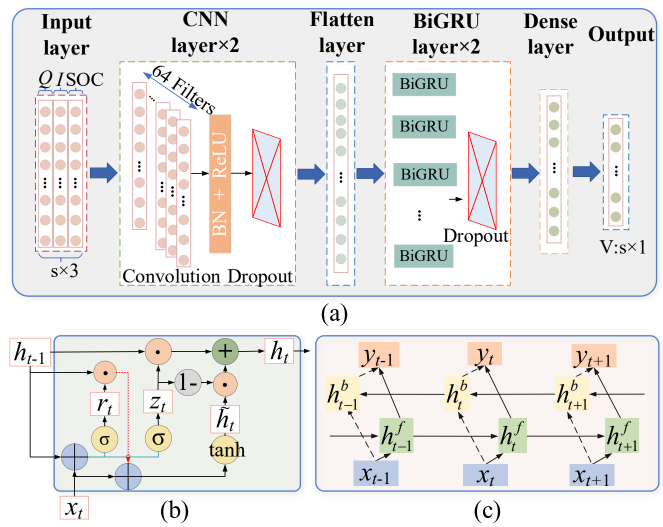


Fig. 5. Model structures of (a) CNN-BiGRU, (b) GRU, and (c) BiGRU.

2) Bi-directional Gated Recurrent Unit

In a recurrent neural network (RNN), the same weighting parameters are applied to the inputs at each time step, allowing the model to learn temporal relationships in the data. Common types of recurrent neural networks include RNN, LSTM, and GRU. The GRU is preferred due to its excellent performance and fewer parameters. Its standard (forward) structure is shown in Fig. 5(b), and the governing equations are

$$z_t=\sigma\left(W_z\cdot[h_{t-1},x_t]+b_z\right), \tag{14}$$

$$r_t=\sigma\left(W_r\cdot[h_{t-1},x_t]+b_r\right), \tag{15}$$

$$\tilde{h}_t=\tanh(W_h\cdot[r_t\odot h_{t-1},x_t]+b_h), \tag{16}$$

$$h_t=\left(1-z_t\right)\odot h_{t-1}+z_t\odot\tilde{h}_t, \tag{17}$$

where σ is the activation function, the operator $\odot$ denotes the Hadamard product, z_t is the update gate, r_t is the reset gate, $\tilde{h}_t$ is the candidate output, h_t is the output, W_z, W_r, and W_h are the weight matrices, b_z, b_r, and b_h are the bias terms, and the subscript t represents the time step.

In the BiGRU, both past and future time step information is used when processing time series data [28], [29]. As shown in Fig. 5(c), this is achieved through two separate loop structures: one for propagating information from past to future (forward GRU) and the other for propagating information from future to past (backward GRU). The equations for the backward GRU are similar to (14)−(17), except that the subscripts t−1 are replaced with t+1. The output of the BiGRU y_t at time step t is obtained by combining the outputs of the two GRUs, i.e.,

$$y_t = [h_t^f, h_t^b], \tag{18}$$

where h_t^f and h_t^b are the outputs of the forward and the backward GRU, respectively.

B. Holistic Hyperparameters Optimization

Given the large number of hyperparameters involved in the proposed model, identifying a single optimal configuration is challenging. Thus, a comprehensive hyperparameter optimization is conducted using the Optuna framework, and the optimization workflow is illustrated in Fig. 6. Table III summarizes the hyperparameters of the proposed CNN-BiGRU model and their respective search ranges used during the optimization process.

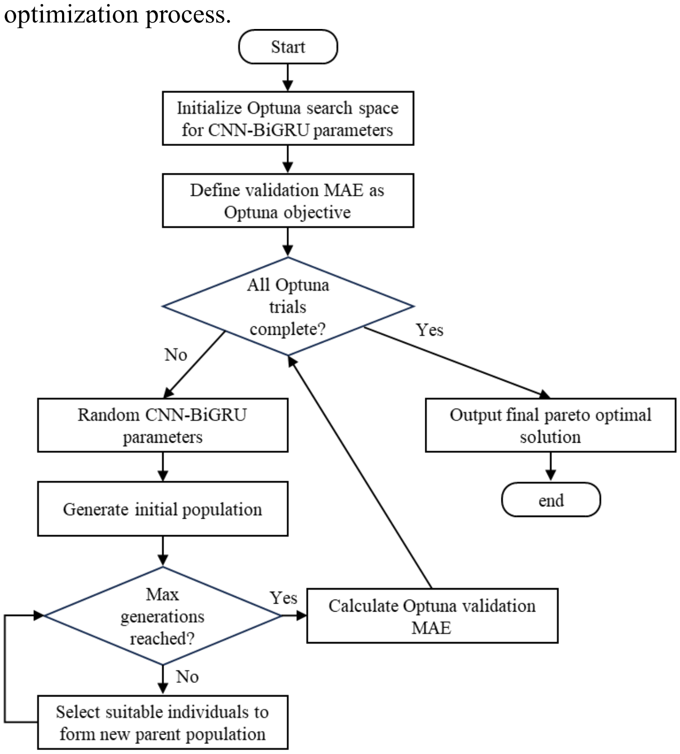


Fig. 6. Flowchart of hyperparameter optimization based on Optuna.

Fig. 7(a) illustrates a Pareto analysis to quantify the trade-off between model complexity and prediction accuracy. The results indicate that the MAE decreases as the number of model parameters increases. However, the marginal improvement becomes noticeably smaller once the parameter size exceeds approximately 0.5 M (million), indicating diminishing returns in prediction accuracy. Therefore, a balanced configuration with about 0.46 M parameters is selected as the Pareto optimal solution, as it significantly reduces the prediction error while avoiding excessive model complexity and computational overhead. Table IV lists the hyperparameter sets corresponding to the selected optimal, compact, and high-accuracy configurations.

Furthermore, the statistical importance of the hyperparameters affecting model accuracy is analyzed. A sensitivity analysis is performed by varying the main hyperparameters within predefined ranges, as illustrated in Fig. 7(b). This analysis helps identify the most influential design parameters and provides additional justification for the selected model configuration.

TABLE III HYPERPARAMETERS SEARCH SPACE OF THE CNN-BIGRU MODEL

Category	Hyperparameter	Search Range
Model topology	CNN channels	32 – 128
	Kernel size	3 – 5
	GRU hidden units	64 – 256
Optimization	Learning rate	10^{-4} – 10^{-3}
Activation	Activation function	ReLU / LeakyReLU / ELU
Optimizer	Optimizer	Adam / RMSprop / SGD
Training strategy	Batch size	8 – 32
	Epochs	80 – 150
Regularization	Dropout rate	0.1 – 0.3

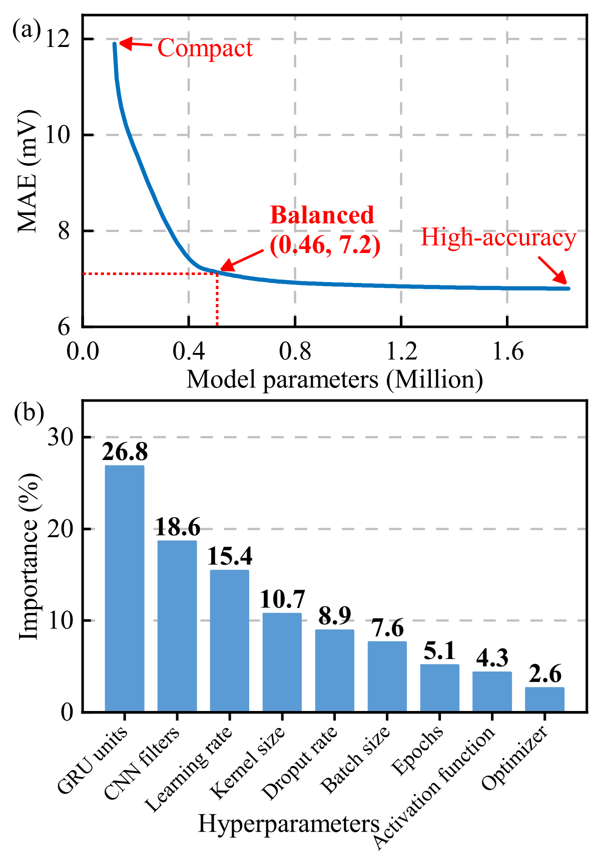


Fig. 7. Selection criteria and importance analysis of hyperparameters of the proposed CNN-BiGRU model. (a) Pareto trade-off between model size and validation MAE; (b) Hyperparameter importance analysis.

TABLE IV TYPICAL HYPERPARAMETER CONFIGURATIONS AND CORRESPONDING MODEL PERFORMANCE

Hyperparameters	Compact	**Balanced (selected)**	High-accuracy
CNN filters	32	64	128
Kernel size	3	3	5
GRU units	64	128	256
Learning rate	10^{-3}	10^{-3}	5×10^{-4}
Activation	ReLU	ReLU	ELU
Optimizer	Adam	Adam	Adam
Parameters (M)	0.12	0.46	1.83
MAE (mV)	11.92	7.20	6.84
RMSE (mV)	15.53	9.62	9.11

C. Model Training

1) Training Algorithm

During training, the neural network model is continuously evaluated based on the difference between the predicted output and the true value. This difference is quantitatively measured using a loss function, specifically the mean squared error. The model is trained using a gradient descent algorithm that minimizes the loss function until the model's output converges to the preset target. In this case, the Adam optimizer is employed, which adjusts the learning rate of each parameter by calculating first-order and second-order moment estimates of the gradient, thereby facilitating more efficient training of the neural network. The corresponding equations are as follows.

First, the 1st-order moment M_t and the 2nd-order moment G_t of the gradient are estimated as

$$M_t = \beta_1 \cdot M_{t-1} + (1-\beta_1) \cdot g_t, \tag{19}$$

$$G_t = \beta_2 \cdot G_{t-1} + (1-\beta_2) \cdot g_t^2, \tag{20}$$

and then corrected by

$$\hat{M}_t = M_t / (1-\beta_1^t), \tag{21}$$

$$\hat{G}_t = G_t / (1-\beta_2^t), \tag{22}$$

where β_1 and β_2 denote the attenuation rates and are set to 0.9 and 0.999, respectively.

Finally, the model parameters are updated by

$$\theta_t = \theta_{t-1} - \alpha \cdot \hat{M}_t / (\sqrt{\hat{G}_t} + \varepsilon), \tag{23}$$

where $\varepsilon = 10^{-9}$ and α is the learning rate of the model.

2) Model Tail Fine-Tuning

The predictive accuracy of the deployed model can be significantly affected in a real-world VRB system due to the following reasons. First, VRB aging, such as capacity degradation, can lead to increased errors in SOC estimation. Second, the model may be applied to a VRB with different internal resistances and SOC-OCV characteristics. Both can cause a significant increase in model errors over time. To address these challenges, the model's parameters must be adjusted to enhance its applicability, a process known as transfer learning. One effective transfer learning strategy is model fine-tuning, where a pre-trained model, developed from source data, is further trained on a task-specific dataset (the target dataset) to optimize its performance [30]. Pre-trained models are typically trained on large-scale datasets to capture a wide range of features to enhance the model's performance on a specific task using a smaller, task-specific dataset. This approach is widely used in training large language models as it requires less time, data, and computational resources compared to training a model from scratch. The specific implementation methods are: 1) When the target dataset is small and closely related to the source dataset, only the parameters of the final few layers of the neural network need to be modified. 2) When the target dataset is small and has low relevance to the source dataset, the first few layers of the pre-trained model are frozen, and only the parameters of the remaining layers are trained. 3) When the target dataset is large and highly correlated with the source dataset, the weights of the pre-trained model are used to initialize the model, which is then fine-tuned through retraining.

Given the high similarity in voltage characteristics among different VRB batteries, the first type of implementation described above, namely *model tail fine-tuning*, is employed. As shown in Fig. 5, the proposed CNN-BiGRU framework consists of five network layers, arranged from output to input as follows: a fully connected layer, two BiGRU layers, and two CNN layers. Taking Cell B under Operation 4 as an example, the data from the first cycle are used as the training set for model fine-tuning, while the data from the remaining three cycles serve as the testing set. Fig. 8 illustrates the impact of fine-tuning different network layers on model prediction error. The number of fine-tuning layers is counted starting from the output of the model.

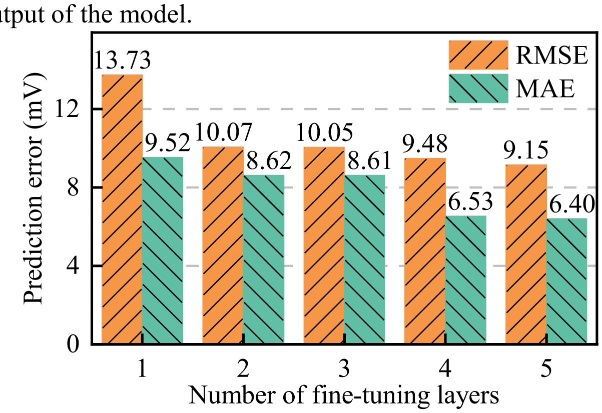


Fig. 8. Impact of fine-tuning different network layers on model performance. 1: A fully connected layer; 2: A fully connected layer and the second BiGRU layer; 3: A fully connected layer and two BiGRU layers; 4: A fully connected layer, two BiGRU layers, and the second CNN layer; 5: A fully connected layer, two BiGRU layers, and two CNN layers.

To achieve a balance between model accuracy and computational burden, the number of fine-tuning layers is set to two. Specifically, two CNN layers and the first BiGRU layer are frozen (i.e., their weights are fixed) during the transfer learning process. These layers are responsible for extracting the universal electrochemical features, such as the intrinsic VO^{2+}/VO_2^+ and V^{2+}/V^{3+} redox kinetics. The final BiGRU layer and the subsequent fully connected output layers are designated as the "model tail". By retraining the tail, the model can adaptively learn corrections caused by different batches or

This article has been accepted for publication in IEEE Transactions on Energy Conversion. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TEC.2026.3686467

manufacturers, as well as variations in the state of the battery. Compared with other transfer learning strategies, this approach mitigates the risk of catastrophic forgetting and demonstrates greater robustness when training on relatively small experimental datasets, thereby improving stability across different manufacturers and operating conditions. Furthermore, adopting a "tail fine-tuning" strategy can significantly reduce the number of parameters that need to be optimized.

3) Threshold Selection

Threshold selection is critical in fault diagnosis. A smaller threshold can lead to a higher false alarm rate, while a larger threshold may delay the detection of the fault. The 3σ criterion is used here to determine the threshold. According to this criterion, for data that follow or approximately follow a normal distribution, approximately 99.73% of the values lie within the interval (μ-3σ, μ+3σ), with any values outside this range considered outliers. Consequently, the threshold is set as μ+3σ for the data corresponding to non-fault conditions. To minimize the effects of capacity degradation and ensure that the data better conform to a normal distribution, data corresponding to low to medium SOC should be used to determine μ+3σ. Moreover, since the performance of VRBs may vary between manufacturers, e.g., some batteries may experience more rapid capacity degradation, it is important to incorporate a margin into the threshold to prevent frequent false alarms. Therefore, the threshold in this study is calculated as

$$\text{threshold} = \mu + 3\sigma + \text{margin}. \tag{24}$$

As discussed in the previous section, the model predicts the voltage at the s-th time step using the data from the previous 1 to s time steps. As a result, if the error caused by the fault becomes detectable at the 2nd time step, the model only performs predictions at the s-th time step. This will introduce a delay of (s-2) time steps in fault detection. Although pump faults in VRBs are typically not catastrophic, it is still important to allow sufficient time for maintenance personnel to respond when a fault occurs. To solve this problem, a sliding window approach can be used in practical detection [31]. Specifically, during each prediction, only the voltage data from the last k time steps are retained, and after the k time steps, the model makes the next prediction. For k = 1, this approach allows for voltage predictions at every time step. In addition, the selection of k should take into account the computational resources of the platform. It is important to note that this approach does not affect the model's input or output dimensions and, thus, does not compromise the model's accuracy.

V. RESULTS AND DISCUSSION

The model's performance is evaluated using the mean absolute error (MAE) and root mean square error (RMSE), i.e.,

$$\text{MAE} = \frac{1}{m}\sum_{i=1}^{m}\left|(V_i - \hat{V}_i)\right|, \tag{25}$$

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(V_i - \hat{V}_i)^2}, \tag{26}$$

where m is the number of samples, and V_i and $\hat{V}_i$ represent the measured and predicted voltages, respectively. As model performance improves, the values of these metrics decrease.

A. Hyperparameter Tuning

Since the performance of the pre-trained model directly affects the accuracy of the fine-tuned model, optimizing hyperparameters during the pre-training phase is crucial. The hyperparameters to be optimized are batch size (BS) and sample length, s. BS refers to the number of samples processed in one training iteration, influencing the efficiency of model training. Based on experience, the range of BS values is set to {4, 8, 16, 32, 64, 128}, and the model's performance under different BSs is illustrated in Fig. 9(a). The bar graphs represent the performance metrics of the model on the test set, while the curve shows the time required to train the model for a single epoch. As shown in Fig. 9(a), a larger BS leads to lower model performance but reduces training time due to the fact that a larger BS can traverse through all the data more quickly. Conversely, a smaller BS results in more frequent parameter updates during training, which improves model generalization. Considering the training time and model accuracy, BS is selected as 16.

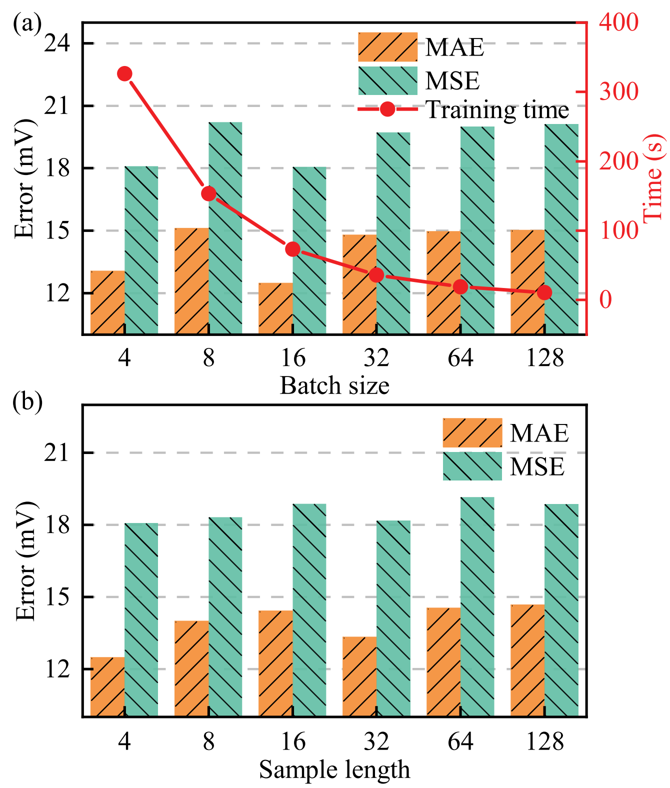


Fig. 9. Results of hyperparameter optimization. Prediction errors with (a) different batch sizes and (b) different sample lengths.

In time series forecasting, the sample length s of a sample influences forecasting accuracy. The interval of s is set to {5, 10, 15, 20, 25, 30}, and the forecasting results of the model under different s are shown in Fig. 9(b). It can be seen that the model achieves the best performance metric at s = 5. This suggests that the battery voltage is most influenced by recent data, while data from earlier time steps may negatively impact the prediction accuracy.

B. Model Validation

The battery voltage profile experiences a sharp drop at very low SOC levels, making operations in these regions generally

This article has been accepted for publication in IEEE Transactions on Energy Conversion. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TEC.2026.3686467

inadvisable for safety reasons. Therefore, in this section, only data corresponding to a depth of discharge of approximately 85% are considered for model validation.

1) Results of Model Tail Fine-Tuning

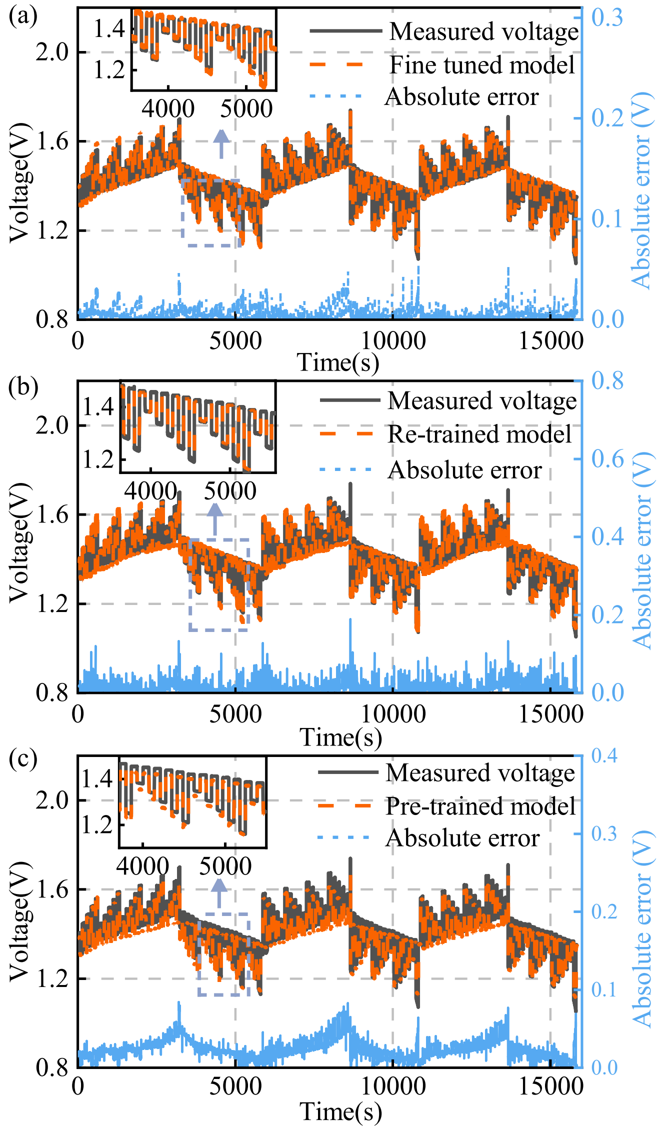


Fig. 10. Prediction results and errors of the models obtained from different training methods under step flow rate and step current operating condition (Operation 4). (a) Fine-tuned model; (b) Re-trained model; (c) Pre-trained model.

In order to examine the effectiveness of the model tail fine-tuning method, data from the first charging and discharging cycle in Operation 4 are used to fine-tune the pre-trained model. The prediction results of different training methods for the remaining data in Operation 4 are compared in Fig. 10. Specifically, the results and errors of the fine-tuned model are illustrated in Fig. 10 (a). The retrained model using only the data from the first charging and discharging cycle in Operation 4 is shown in Fig. 10 (b). The results and errors of the pre-trained model are shown in Fig. 10(c). Model performance metrics in Operation 4 and other conditions with different training methods are given in Table V.

The predicted voltage of the pre-trained model deviates significantly from the measured voltage during the middle and late stages of each charging and discharging cycle, primarily due to battery capacity decay, resulting in an inaccurate SOC estimation. The retrained model exhibits a slightly lower overall error but significant local errors. This training method partially learned the current state of the battery, yet insufficient data prevented the model from fully capturing the battery voltage characteristics. Among the three models, the model fine-tuned from the pre-trained model achieved the highest accuracy, with an average error as low as 6.838 mV. The maximum model error is only 40 mV, excluding instances of higher SOC at the end of charging. In addition, as can be seen from Table V, the fine-tuned model errors are significantly lower than the pre-trained model and the newly trained model in all operations except Operation 3, highlighting the superiority of this training approach. Performance metrics for the re-trained model are omitted in Operations 1-3 since the voltage characteristics can only be learned under a single flow rate using data from a single cycle in these operations.

TABLE V COMPARISON OF PERFORMANCE METRICS WITH DIFFERENT TRAINING METHODS

Op.	Model	MAE	RMSE
1	Pre-trained model	1.191×10^{-2}	1.309×10^{-2}
	Fine-tuned model	4.042×10^{-3}	5.773×10^{-3}
2	Pre-trained model	9.137×10^{-3}	1.082×10^{-2}
	Fine-tuned model	8.573×10^{-3}	1.005×10^{-2}
3	Pre-trained model	9.866×10^{-3}	1.270×10^{-2}
	Fine-tuned model	1.099×10^{-2}	1.370×10^{-2}
4	Pre-trained model	23.99×10^{-2}	27.36×10^{-2}
	Fine-tuned model	6.630×10^{-3}	9.835×10^{-3}
	Retrained model	11.87×10^{-2}	16.48×10^{-2}
5	Pre-trained model	14.86×10^{-2}	1.734×10^{-2}
	Fine-tuned model	9.499×10^{-3}	1.320×10^{-2}
	Retrained model	1.069×10^{-2}	1.3207×10^{-2}

2) Model Validation Under Various Operating Conditions

Similarly, the first charging and discharging cycle data for each condition in the test set are used for fine-tuning the model. Prediction results and model errors for different operating conditions are shown in Fig. 11. Performance metrics for each operation are presented in Table V. Results for the model under constant flow and constant current conditions (Operation 1) are illustrated in Fig. 11 (a). It can be seen that the predicted voltage of the model at each flow rate closely matches the measured voltage, with a low MAE of 4.042 mV. This demonstrates the model's ability to capture the effect of flow rate on voltage. In addition, the error of the model is almost limited to 20 mV, excluding the stage of high SOC at the end of charging.

As illustrated in Fig. 11 (a), most studies in the literature have only evaluated model accuracy under constant flow and constant current conditions (Operation 1), which is insufficient for practical systems. In this section, we aim to assess the model's performance under complex operating conditions to demonstrate its robustness. The results and errors of the model under constant flow rate variable current (Operations 2 and 3) are shown in Figs. 11(b) and (c). Although current variations increase prediction complexity, the model maintains high accuracy. The model prediction results and error under the variable flow rate and variable current (Operation 5) are shown in Fig. 11(d). As indicated in Table V, the model achieves an MAE of 9.499 mV, with a maximum error of around 50 mV. Despite these challenging conditions, the low model error demonstrates its excellent performance.

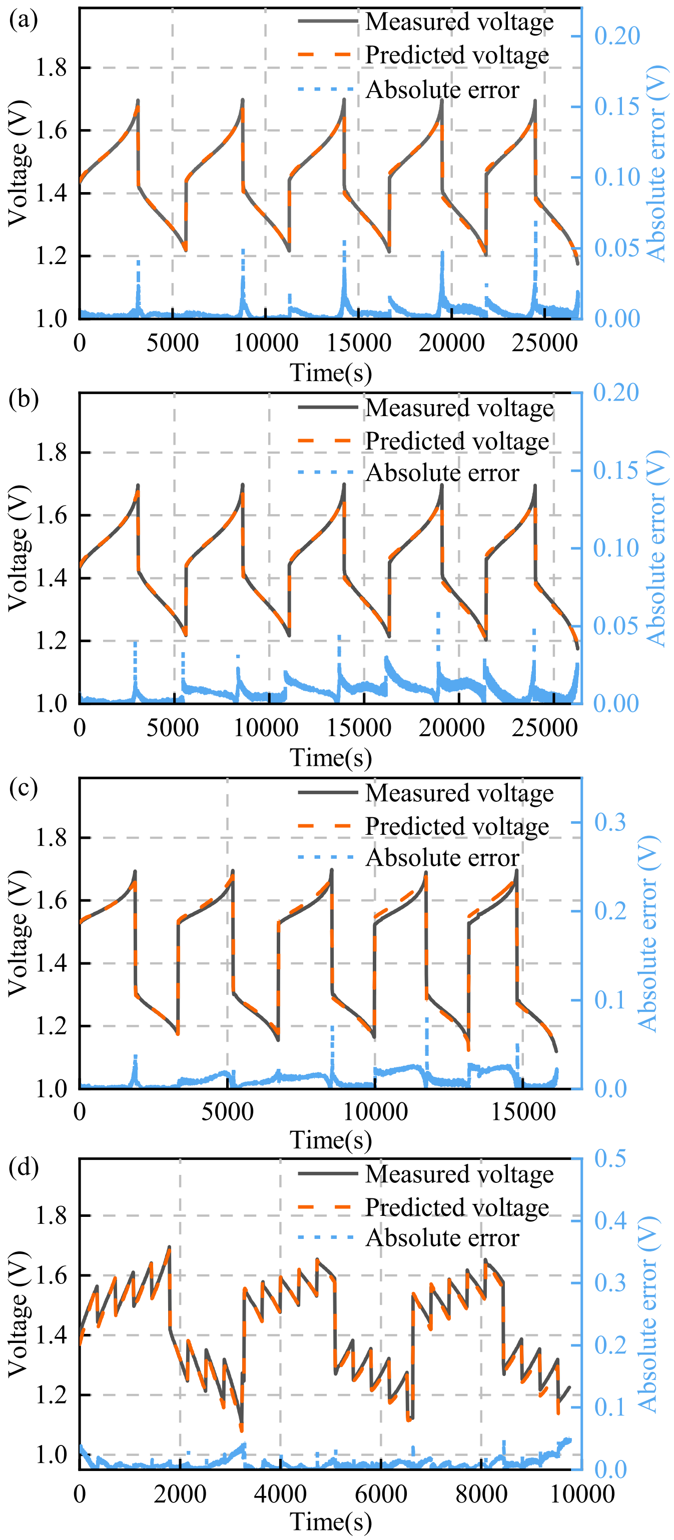


Fig. 11. Prediction results and errors of the model for different operating conditions. (a) Constant flow rate and constant current operating (Operation 1); (b) Constant flow rate, linearly rising current condition (Operation 2); (c) Constant flow rate, linearly decreasing current condition (Operation 3); (d) Step flow rate, linearly rising current condition (Operation 5).

3) Model Validation Using Different Battery Types

In this section, data from a different type of battery, denoted by Cell C, is used to assess the model's applicability. A set of constant-current charging and discharging data and a set of variable-current charging and discharging data of Cell C are employed for model fine-tuning. The former is used to learn the OCV-SOC characteristics, while the latter helps to understand the internal resistance characteristics. In addition, the flow rate is consistently set to 60 mL/min. Prediction results and errors of Cell C are illustrated in Fig. 12, where results and errors for the constant current conditions are shown in Fig. 12(a), and results and errors for the variable current conditions are shown in Fig. 12(b). It can be seen that the model errors are very low in both scenarios, generally below 20 mV. The maximum error of 46 mV under constant current conditions occurs at the charging and discharging transitions, while the maximum error under the variable-current condition is 35 mV.

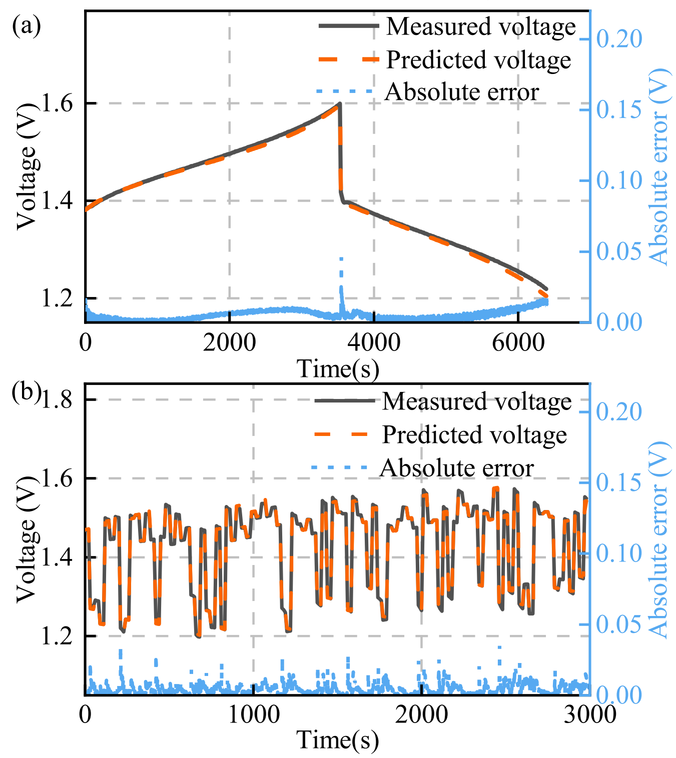


Fig. 12. Prediction results and errors of battery Cell C. (a) Constant current conditions; (b) Variable current conditions.

C. Comparison of Different Models

The performance metrics of LSTM, GRU, BiGRU, and CNN-BiGRU under different operating conditions are presented in Table VI. The results show that BiGRU is superior to GRU, particularly under complex and variable current conditions. The improvement is attributed to BiGRU's capability to process battery information bidirectionally and adjust weights comprehensively, resulting in more accurate voltage estimates. Furthermore, CNN-BiGRU achieves the highest accuracy, as the CNN effectively captures spatial features in the data, which enables more efficient weight adjustment in the BiGRU layer and improved prediction

This article has been accepted for publication in IEEE Transactions on Energy Conversion. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TEC.2026.3686467

accuracy. Specifically, although BiGRU or LSTM models are effective at capturing temporal dependencies within individual input features, they struggle to handle the coupled, locally spatial interactions among current, flow rate, and SOC in VRB systems [2]. Conversely, CNN is designed to extract local spatial patterns from the above features over short time windows. By performing convolution operations, CNNs learn localized feature representations, which are then passed to capture longer-term temporal dependencies [27]. This hierarchical feature learning (spatial via CNN, temporal via BiGRU) allows the model to accurately represent the coupled electrochemical and hydraulic dynamics of VRBs.

TABLE VI PERFORMANCE METRICS FOR DIFFERENT MODELS

Model	Op.	MAE (mV)	Mean MAE	RMSE (mV)	Mean RMSE
LSTM	1	4.46	9.13	6.13	11.8
	2	9.23		12.05	
	3	8.58		11.61	
	4	11.17		13.99	
	5	12.31		15.42	
GRU	1	4.57	9.53	6.18	12.5
	2	9.34		12.89	
	3	9.68		13.16	
	4	11.12		14.15	
	5	12.92		15.86	
BiGRU	1	4.34	8.70	6.01	11.2
	2	9.13		11.06	
	3	7.459		10.03	
	4	11.01		13.73	
	5	11.59		14.93	
CNN-BiGRU	1	4.04	7.95	5.77	10.7
	2	8.57		10.05	
	3	6.63		9.84	
	4	10.99		13.7	
	5	9.5		13.2	

D. Results of Pump Fault Detection

In this section, Cell C is used for pump fault detection experiments. Charging and discharging experiments are performed on Cell C under two sets of 1800 seconds of simulated operating conditions, which we named Cases 1 and 2, respectively. Case 1 is a random positive and negative pulse charging condition, while Case 2 is a random positive and negative pulse discharging condition. Three types of pump faults are set up in each set of conditions, namely, positive pump fault (F_P), negative pump fault (F_N), and bilateral pump fault (F_B). Furthermore, we selected a sliding window step of 1 second, predicting the voltage every second. A sampling rate of 1 Hz is practically feasible, and the framework can be adapted to other rates supported by the target BMS hardware. The pump fault detection results for Case 1 are shown in Fig. 13, and those for Case 2 are presented in Fig. 13.

For Cell C, the value of $\mu+3\sigma$ in the non-fault data is 20 mV. However, as shown in Figs. 9 to 11, model errors during large current jumps or sudden flow rate change can exceed this threshold due to transient voltage instability caused by uneven electrolyte concentration [32]. This results in an additional maximum error of 20 mV caused by the load condition. Furthermore, to account for the accelerated degradation of small batteries, a 10 mV margin should be added to the threshold. Therefore, the final margin is set to 30 mV, leading to a preferred threshold of 50 mV, according to (18). Detailed calculation procedures are provided in Appendix A. Given manufacturer-specific variations in VRB performance, this threshold may need to be further adjusted based on actual conditions.

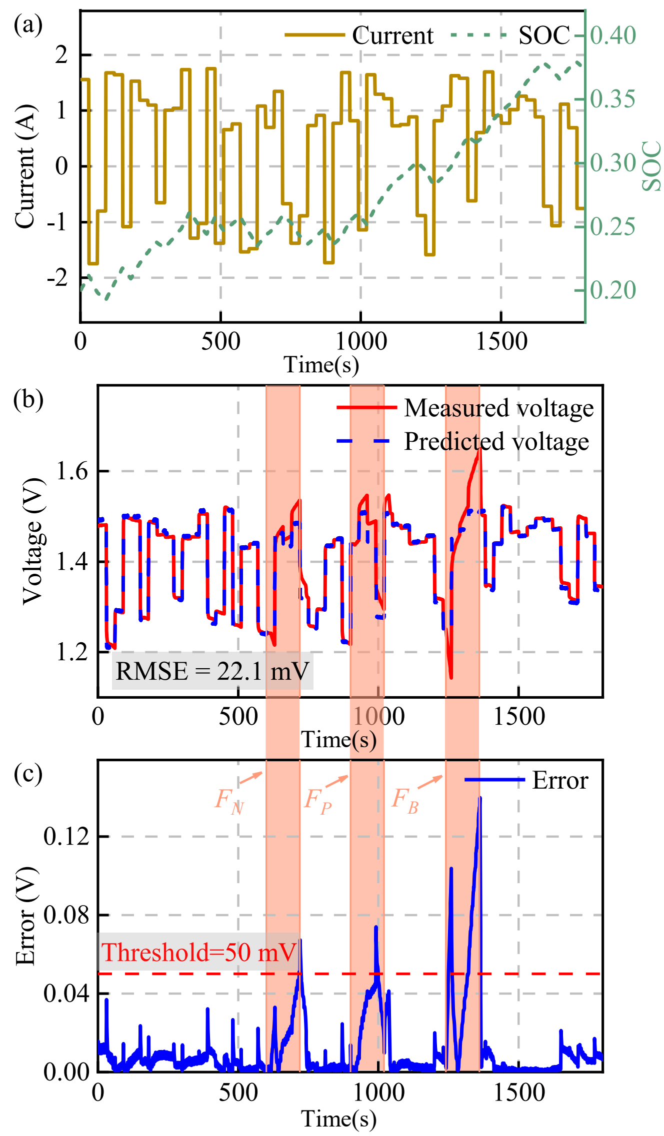


Fig. 13. Pump fault detection results in random positive and negative pulse charging conditions (Case 1). (a) Current and SOC information; (b) Predicted and measured voltage; (c) Predicted error.

As illustrated in Fig. 13 and Fig. 14, the bilateral pump fault in Case 1 and the negative pump and the bilateral pump faults in Case 2 cause rapid voltage changes, leading to significant differences between predicted and measured voltages. In contrast, for the other three faults, the measured voltage initially changes but returns to normal as the current adjusts. For instance, in the negative pump fault in Case 1, the current is negative at the onset of the fault, causing a voltage drop. When

This article has been accepted for publication in IEEE Transactions on Energy Conversion. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TEC.2026.3686467

the current shifts to positive, the voltage rises rapidly, causing the fault voltage to approach normal levels. Consequently, the voltage change due to the pump fault takes longer to become noticeable. This delay in voltage deviation highlights the subtle nature of pump faults under complex operating conditions. Relying solely on charging and discharging cut-off voltages for fault detection risks prolonged operation under fault conditions. In contrast, the proposed method consistently exhibits significantly larger model errors during faulty conditions compared to normal conditions. As illustrated in Table VII, VRB pump faults can be detected in an average of only 62.67 s under complex operation conditions, which corresponds to 57.85% of the fault duration. With a 50 mV threshold, all six pump fault scenarios are detected promptly, demonstrating the model's high sensitivity and robustness.

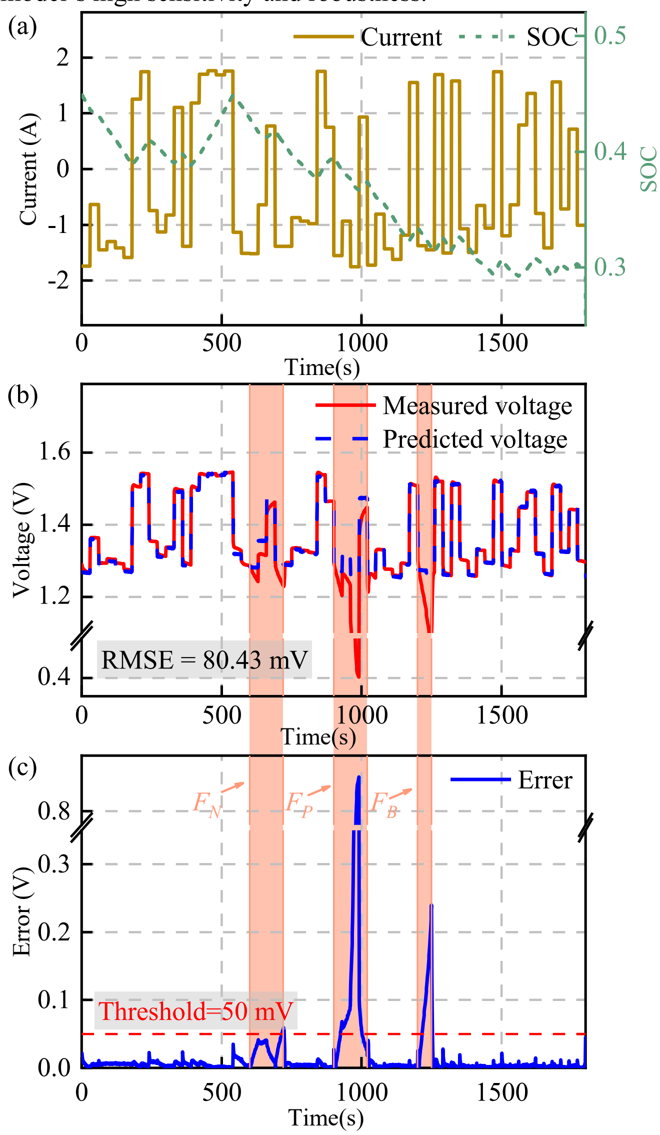


Fig. 14. Pump fault detection results in random positive and negative pulse discharging conditions (Case 2). (a) Current and SOC information; (b) Predicted and measured voltage; (c) Predicted error.

TABLE VII PERFORMANCE ON THE PUMP FAULT DETECTION IN TWO CASES

Faults categories	Fault duration	Detect? and moment	Detection time and efficiency
Case 1_F_N	[601, 720]	√, 718th s	118 s (98.33%)
Case 1_F_P	[901,1020]	√, 991st s	91 s (75.83%)
Case 1_F_P	[1241,1360]	√, 1250th s	10 s (8.33%)
Case 2_F_N	[601,720]	√, 715th s	115 s (95.83%)
Case 2_F_P	[901,1020]	√, 925th s	25 s (20.83%)
Case 2_F_P	[1201,1250]	√, 1217th s	17 s (34%)

E. Analysis of Aging and Fine-tuning Contribution

To demonstrate that the proposed model can adapt to battery aging and to validate the necessity of fine-tuning for transfer learning, Cell C is subjected to 50 charge–discharge cycles (about 87% SOH) and used for fault detection of an aged battery. Battery aging is most directly manifested as capacity loss and increased internal resistance. During charge and discharge, the terminal voltage reaches the upper/lower cutoff limits more quickly, leading to slower voltage rise and faster voltage decay over the same time interval. At the instant the current undergoes a step change during complex operating conditions, the magnitude of the voltage jump directly reflects the internal resistance. Consequently, aged cells exhibit larger voltage jumps, which increases the difficulty of model prediction and reduces its accuracy. Taking Case 1 as an example, as illustrated in Fig. 15(a), the fine-tuned model achieves an RMSE of 23.98 mV on the aged cell, which is 1.88 mV higher than the 22.1 mV observed for the new cell.

The pre-trained model fails to capture the dynamic characteristics and internal resistance features of Cell C, resulting in poor voltage prediction performance for both the new and aged cells, as shown in Fig. 16. The fault detection rate (FDR) and false alarm rate (FAR) are commonly used to evaluate the performance of fault diagnosis methods. As presented in Table VIII, although the pre-trained model achieves an FDR of 100% for both the new and aged cells, the FAR is excessively high, making it unsuitable for practical application. The calculation of FAR is illustrated using the performance of the pre-trained model on Case 1 of the aged Cell C (i.e., Fig. 15(d)). During the entire operating period, a total of 181 data points exceeded the threshold of 50 mV. Among them, 2, 36, and 34 points are correctly identified during the F_N, F_P, and F_B, respectively, yielding 72 points in total. Therefore, the FAR in this scenario is calculated as $(181 - 72) / 181 = 60.22\%$.

The proposed fine-tuned model has learned the system characteristics of the new Cell C. Therefore, in the same voltage prediction task, it achieves better RMSE performance. Although the RMSE slightly increases after aging, it remains within an acceptable range. Since the key parameter thresholds are closely related to the voltage prediction error, the aging effect has already been considered during threshold design. Consequently, even in the pump fault diagnosis task for the aged Cell C, the FDR and FAR under Cases I and II remain the same as those for the new battery, maintaining 100% and 0%, respectively, as shown in Table VIII and Fig. 15(b).

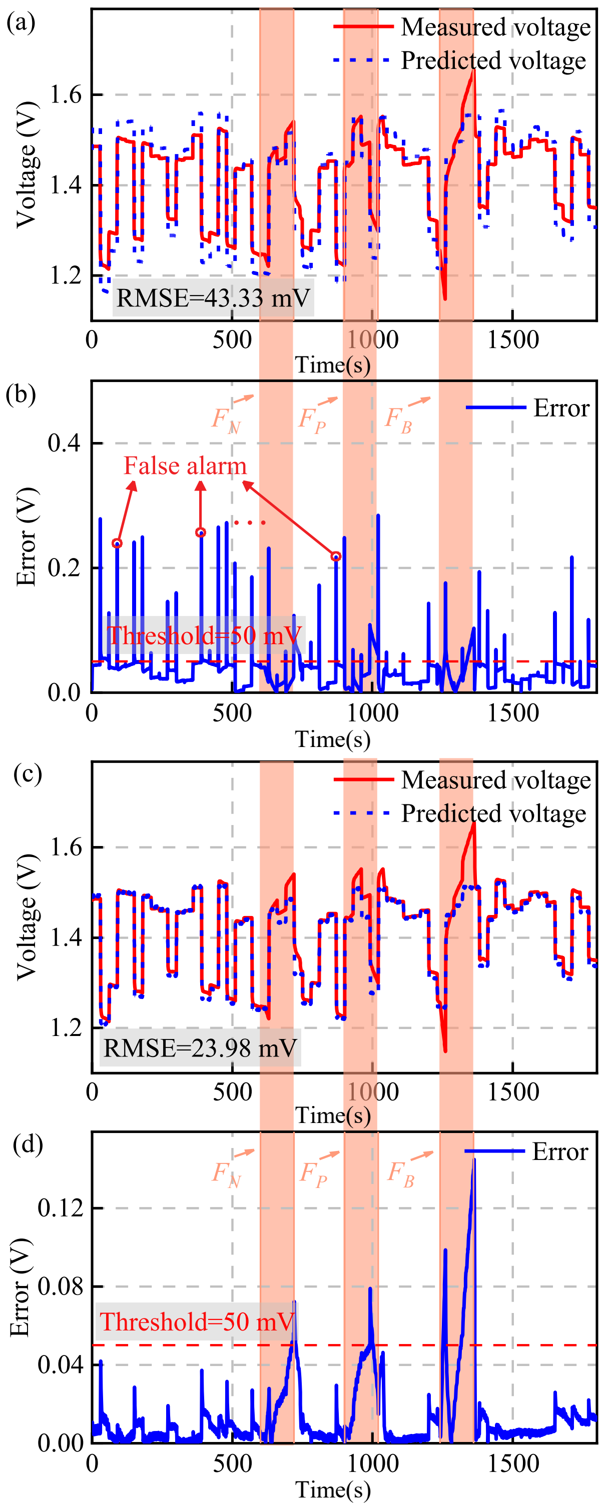


Fig. 15. Pump fault detection results for aged Cell C (87% SOH) in Case I. (a) Predicted and measured voltage under fine-tuned model; (b) Predicted absolute error under fine-tuned model; (c) Predicted and measured voltage under pre-trained model; (d) Predicted absolute error under pre-trained model.

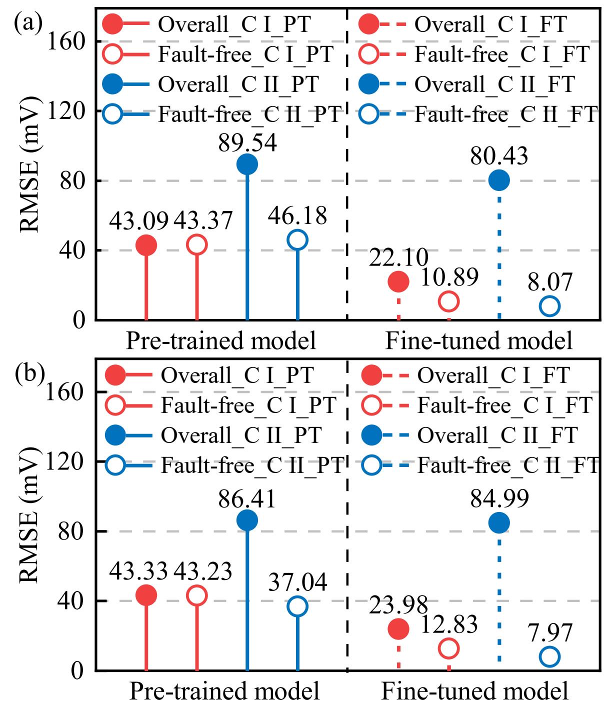


Fig. 16. RMSE comparison: overall vs fault-free scenarios for pre-trained and fine-tuned models. (a) New cell (100% SOH); (b) Aged cell (87% SOH).

TABLE VIII FDR AND FAR OF DIFFERENT AGING STATES FOR PRE-TRAINED AND FINE-TUNED MODELS

Aging state	Scenario	Pre-trained model		Fine-tuned model	
		FDR	FAR	FDR	FAR
100% SOH	Case I	100%	64.58%	100%	0%
	Case II	100%	43.75%	100%	0%
87% SOH	Case I	100%	60.22%	100%	0%
	Case II	100%	29.83%	100%	0%

VI. CONCLUSION

An analysis of fault mechanisms shows that pump failure disrupts electrolyte circulation and causes rapid voltage fluctuations. Common basic statistical models or simple baseline approaches struggle to perform accurate dynamic predictions under multivariate and non-stationary operating conditions. Based on this insight, a CNN-BiGRU voltage prediction model is developed to accurately capture VRB voltage dynamics using measured data without the need for complex electrochemical modeling. To improve adaptability across varying operating conditions, an offline tail fine-tuning strategy is introduced, which enables the pre-trained model to incorporate real-time operational information and maintain high prediction accuracy. Experimental validation under various operating scenarios demonstrated that the proposed framework consistently limits the prediction error within 40 mV, except during abrupt voltage changes at the end of charging or discharging processes. Furthermore, a real-time fault detection scheme is implemented using a sliding window approach with adjustable step sizes, which offers adaptability for deployment across platforms with different computational capabilities. Finally, the framework is validated through

This article has been accepted for publication in IEEE Transactions on Energy Conversion. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TEC.2026.3686467

experiments involving actual pump faults under two complex operating scenarios, successfully detecting six pump fault scenarios in a timely and reliable manner.

APPENDIX A

The load condition component in the key parameter *margin* is determined by minimizing the average detection time of the proposed pump fault detection method under two cases, i.e.,

$$\min(\text{ave. time}) = \frac{1}{2}\sum_{i}^{2}(t_{i,N} + t_{i,P} + t_{i,B}), \tag{A.1}$$

subject to

$$\begin{cases} t_{1,N} \in [601,900] \\ t_{1,P} \in [901,1240] \\ t_{1,B} \in [1241,1800] \\ t_{1,N} \in [601,900] \\ t_{1,P} \in [901,1200] \\ t_{1,B} \in [1201,1800] \end{cases}, \tag{A.2}$$

where $t_{i,N}$, $t_{i,P}$, and $t_{i,B}$ (i=1 and 2) denote detection time of negative, positive, and bilateral pump faults in VRBs under Cases 1 and 2, respectively.

The particle swarm optimization (PSO) algorithm, combined with a backpropagation method, is employed to solve the above optimization problem. After randomly initializing α particles in the initial stage, the PSO iteration is first performed for each particle until convergence. A population size of 50 particles is used for the PSO algorithm to ensure sufficient search diversity within a controllable computational cost. Results are considered invalid if missed alarms or premature false alarms occur, in which case the process automatically proceeds to the next iteration and is recorded accordingly. In the kth iteration, the bth particle is updated as follows:

$$v_b^k = \omega v_b^{k-1} + c_1 r_1(\alpha_g^{k-1} - \alpha_b^{k-1}) + c_2 r_2(\alpha_l^{k-1} - \alpha_b^{k-1}) \tag{A.3}$$

$$\alpha_b^k = \alpha_b^{k-1} + v_b^k \tag{A.4}$$

where v is the iteration coefficient, ω is the inertia coefficient, c_1 and c_2 represent the step sizes for the local best position and the global best position, both set to 2. r_1 and r_2 are samples drawn from a uniform distribution between 0 and 1. In addition, the subscripts g and l denote global and local, respectively. The initial value of the inertia weight is set to 0.9 and decreases linearly to 0.4 to balance exploration and exploitation. The cognitive and social acceleration coefficients are both set to 2. Velocity updates are bounded by a maximum speed per dimension to avoid instability.

The PSO optimization results indicate an optimal margin of 20 mV when only the load condition component is calculated. Considering the potential influence of degradation factors on the battery's external characteristics, the degradation effect is set to 10 mV. In summary, the final optimal margin is determined to be 30 mV.

REFERENCES

[1] B. Xiong *et al.*, 'Design of A Two-Stage Control Strategy of Vanadium Redox Flow Battery Energy Storage Systems for Grid Application', *IEEE Trans. Sustain. Energy*, vol. 13, no. 4, pp. 2079–2091, Oct. 2022, doi: 10.1109/TSTE.2022.3181751.

[2] Y. Liu *et al.*, 'A novel vanadium redox flow battery modelling method using honey badger optimization assisted CNN-BiLSTM', *J. Power Sources*, vol. 558, p. 232610, Feb. 2023, doi: 10.1016/j.jpowsour.2022.232610.

[3] R. Yang, R. Xiong, S. Ma, and X. Lin, 'Characterization of external short circuit faults in electric vehicle Li-ion battery packs and prediction using artificial neural networks', *Appl. Energy*, vol. 260, p. 114253, Feb. 2020, doi: 10.1016/j.apenergy.2019.114253.

[4] R. Yang, R. Xiong, and W. Shen, 'On-board diagnosis of soft short circuit fault in lithium-ion battery packs for electric vehicles using an extended Kalman filter', *CSEE J. Power Energy Syst.*, vol. 8, no. 1, pp. 258–270, Jan. 2022, doi: 10.17775/CSEEJPES.2020.03260.

[5] Y. Xu, X. Ge, W. Shen, and R. Yang, 'A Soft Short-Circuit Diagnosis Method for Lithium-Ion Battery Packs in Electric Vehicles', *IEEE Trans. Power Electron.*, vol. 37, no. 7, pp. 8572–8581, Jul. 2022, doi: 10.1109/TPEL.2022.3151620.

[6] X. Kong *et al.*, 'Pseudo-two-dimensional model and impedance diagnosis of micro internal short circuit in lithium-ion cells', *J. Energy Storage*, vol. 27, p. 101085, Feb. 2020, doi: 10.1016/j.est.2019.101085.

[7] T. Roy and S. Dey, 'Fault Detectability Conditions for Linear Deterministic Heat Equations', *IEEE Control Syst. Lett.*, vol. 3, no. 1, pp. 204–209, Jan. 2019, doi: 10.1109/LCSYS.2018.2872215.

[8] T. Lin, Z. Chen, C. Zheng, D. Huang, and S. Zhou, 'Fault Diagnosis of Lithium-Ion Battery Pack Based on Hybrid System and Dual Extended Kalman Filter Algorithm', *IEEE Trans. Transp. Electrification*, vol. 7, no. 1, pp. 26–36, Mar. 2021, doi: 10.1109/TTE.2020.3006064.

[9] S. Sepasiahooyi and F. Abdollahi, 'Fault detection of new and aged lithium-ion battery cells in electric vehicles', *Green Energy Intell. Transp.*, vol. 3, no. 3, p. 100165, Jun. 2024, doi: 10.1016/j.geits.2024.100165.

[10] 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, p. 118866, Jan. 2021, doi: 10.1016/j.energy.2020.118866.

[11] T. P. Banerjee and S. Das, 'Multi-sensor data fusion using support vector machine for motor fault detection', *Inf. Sci.*, vol. 217, pp. 96–107, Dec. 2012, doi: 10.1016/j.ins.2012.06.016.

[12] G. Ma, S. Xu, and C. Cheng, 'Fault detection of lithium-ion battery packs with a graph-based method', *J. Energy Storage*, vol. 43, p. 103209, Nov. 2021, doi: 10.1016/j.est.2021.103209.

[13] Z. Qin, Y. Li, J. Tang, S. Zhang, C. Xie, and B. Xiong, 'Pump Fault Detection Method for Vanadium Redox Flow Batteries Without Flow Rate Sensors', in *2022 IEEE Sustainable Power and Energy Conference (iSPEC)*, Dec. 2022, pp. 1–5. doi: 10.1109/iSPEC54162.2022.10033051.

[14] J. Hong, Z. Wang, and Y. Yao, 'Fault prognosis of battery system based on accurate voltage abnormity prognosis using long short-term memory neural networks', *Appl. Energy*, vol. 251, p. 113381, Oct. 2019, doi: 10.1016/j.apenergy.2019.113381.

[15] D. Li, Z. Zhang, P. Liu, Z. Wang, and L. Zhang, 'Battery Fault Diagnosis for Electric Vehicles Based on Voltage Abnormality by Combining the Long Short-Term Memory Neural Network and the Equivalent Circuit Model', *IEEE Trans. Power Electron.*, vol. 36, no. 2, pp. 1303–1315, Feb. 2021, doi: 10.1109/TPEL.2020.3008194.

[16] O. Ojo, H. Lang, Y. Kim, X. Hu, B. Mu, and X. Lin, 'A Neural Network Based Method for Thermal Fault Detection in Lithium-Ion Batteries', *IEEE Trans. Ind. Electron.*, vol. 68, no. 5, pp. 4068–4078, May 2021, doi: 10.1109/TIE.2020.2984980.

[17] O. J. Ojo, X. Lin, H. Lang, and X. Hu, 'A Voltage Fault Detection Method Enabled by A Recurrent Neural Network and Residual Threshold Monitor for Lithium-ion Batteries', in *2021 IEEE Transportation Electrification Conference & Expo (ITEC)*, Jun. 2021, pp. 813–820. doi: 10.1109/ITEC51675.2021.9490102.

[18] D. Li *et al.*, 'Battery Thermal Runaway Fault Prognosis in Electric Vehicles Based on Abnormal Heat Generation and Deep Learning Algorithms', *IEEE Trans. Power Electron.*, vol. 37, no. 7, pp. 8513–8525, Jul. 2022, doi: 10.1109/TPEL.2022.3150026.

[19] D. Luder *et al.*, 'Big data generation platform for battery faults under real-world variances', *Green Energy Intell. Transp.*, vol. 4, no. 3, p. 100282, Jun. 2025, doi: 10.1016/j.geits.2025.100282.

[20] H. Zhao *et al.*, 'Online surface temperature prediction and abnormal diagnosis of lithium-ion batteries based on hybrid neural network and fault threshold optimization', *Reliab. Eng. Syst. Saf.*, vol. 243, p. 109798, Mar. 2024, doi: 10.1016/j.ress.2023.109798.

[21] Y.-H. Sun, H.-L. Jou, and J.-C. Wu, 'Auxiliary diagnosis method for lead–acid battery health based on sample entropy', *Energy Convers. Manag.*, vol. 50, no. 9, pp. 2250–2256, Sep. 2009, doi: 10.1016/j.enconman.2009.05.001.

[22] 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, doi: 10.1016/j.apenergy.2016.12.143.

[23] X. Li, K. Dai, Z. Wang, and W. Han, 'Lithium-ion batteries fault diagnostic for electric vehicles using sample entropy analysis method', *J. Energy Storage*, vol. 27, p. 101121, Feb. 2020, doi: 10.1016/j.est.2019.101121.

[24] Z. Sun *et al.*, 'Detection of voltage fault in the battery system of electric vehicles using statistical analysis', *Appl. Energy*, vol. 307, p. 118172, Feb. 2022, doi: 10.1016/j.apenergy.2021.118172.

[25] H. Chen, X. Li, H. Gao, J. Liu, C. Yan, and A. Tang, 'Numerical modelling and in-depth analysis of multi-stack vanadium flow battery module incorporating transport delay', *Appl. Energy*, vol. 247, pp. 13–23, Aug. 2019.

[26] C. Tian, Y. Zhang, W. Zuo, C.-W. Lin, D. Zhang, and Y. Yuan, 'A heterogeneous group CNN for image super-resolution', Sep. 26, 2022, *arXiv*: arXiv:2209.12406. doi: 10.48550/arXiv.2209.12406.

[27] O. Rubasinghe, X. Zhang, T. K. Chau, Y. H. Chow, T. Fernando, and H. H.-C. Iu, 'A Novel Sequence to Sequence Data Modelling Based CNN-LSTM Algorithm for Three Years Ahead Monthly Peak Load Forecasting', *IEEE Trans. Power Syst.*, vol. 39, no. 1, pp. 1932–1947, Jan. 2024, doi: 10.1109/TPWRS.2023.3271325.

[28] F. Wu, S. Wang, D. Liu, W. Cao, C. Fernandez, and Q. Huang, 'An improved convolutional neural network-bidirectional gated recurrent unit algorithm for robust state of charge and state of energy estimation of new energy vehicles of lithium-ion batteries', *J. Energy Storage*, vol. 82, p. 110574, Mar. 2024, doi: 10.1016/j.est.2024.110574.

[29] Y. Liu, Q. Li, and K. Wang, 'Revealing the degradation patterns of lithium-ion batteries from impedance spectroscopy using variational auto-encoders', *Energy Storage Mater.*, vol. 69, p. 103394, May 2024, doi: 10.1016/j.ensm.2024.103394.

[30] B. Min *et al.*, 'Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey', Nov. 01, 2021, *arXiv*: arXiv:2111.01243. doi: 10.48550/arXiv.2111.01243.

[31] Z. Cui, L. Kang, L. Li, L. Wang, and K. Wang, 'A hybrid neural network model with improved input for state of charge estimation of lithium-ion battery at low temperatures', *Renew. Energy*, vol. 198, pp. 1328–1340, Oct. 2022, doi: 10.1016/j.renene.2022.08.123.

[32] S. Parsegov, M. Pugach, A. Polyakov, and F. Ibáñez, 'Analysis of flow factor control strategy in vanadium redox flow batteries', *IFAC-Pap.*, vol. 55, no. 9, pp. 187–192, Jan. 2022, doi: 10.1016/j.ifacol.2022.07.033.

Binyu Xiong (Senior Member, IEEE) received the B.E. degree in electrical engineering from Wuhan University, Wuhan, China, in 2010, the M.Sc. and Ph.D. degrees from the Nanyang Technological University, Singapore, in 2011 and 2016, respectively. He is currently a Professor with the School of Automation, Wuhan University of Technology, Wuhan, China. His research interests include electrical and thermal modeling of batteries, battery state of charge estimation, large-scale energy storage systems, and renewable energy generation.

Shaojin Wang received the M.S. degree in electrical engineering from the Hubei University of Technology, Wuhan, China, in 2023. He is currently working toward the Ph.D. degree with the School of Automation, Wuhan University of Technology, Wuhan. His research interest focuses on battery modeling.

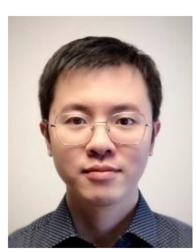
Jinsong Wang received his Ph.D. in system and control engineering from Case Western Reserve University, USA. He is presently a Research Scientist at HyperStrong Technology Co., LTD, Beijing, China. His research focuses on large-scale mathematical and convex optimization, particularly in the context of renewable energy and battery energy storage systems.

Peng Zhou received the B.E. degree in electrical engineering and its automation from Jiangnan University, Wuxi, China, in 2021, and the M.S degrees in control engineering from the Wuhan University of Technology, Wuhan, China, in 2025. He is presently a Research Scientist at Camel Group Co., LTD, Wuhan, China. His research interest focuses on battery fault diagnosis.

Quan Zhou received the Ph.D. degree in biomedical engineering from The Huazhong University of Science and Technology (HUST), Wuhan, China in 2021. He was a Postdoctoral Researcher with the Department of Mechanical Engineering, HUST, Wuhan, China, from 2021 to 2023. He is currently an Assistant Professor with the School of Automation, Wuhan University of Technology. His research interests include interdisciplinary field of artificial intelligence for medical applications, with expertise in medical image analysis and robotic surgery perception.

Changjun Xie (Member, IEEE) received the Ph.D. degree in vehicle engineering from the Wuhan University of Technology (WHUT), Wuhan, China, in 2009.

He is currently a Professor with the School of Automation, WHUT. He is the Dean of the Modern Industry College of Artificial Intelligence and New Energy Vehicles, WHUT. From 2012 to 2013, he was a Visiting Scholar with the University of California Davis, Davis, CA, USA. He has authored or coauthored more than 100 articles, and more than 60 articles are indexed by SCI or EI. His research interests include the fuel cell system, control strategy of new vehicles, and thermoelectric power generation.



Yang Li (Senior Member, IEEE) received the B.E. degree in electrical engineering from Wuhan University, Wuhan, China, in 2007, and the M.Sc. and Ph.D. degrees in power engineering from Nanyang Technological University (NTU), Singapore, in 2008 and 2015, respectively.

He was with the Energy Research Institute, NTU and the School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, QLD, Australia, School of Automation, Wuhan University of Technology, Wuhan, and Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden from

2016 to 2025. Since 2025, he has been a Professor with the School of Electrical Engineering and automation, Wuhan University, Wuhan, China. His research interests include modeling and control of energy storage systems in power and transport sectors.

Dr. Li serves as an Associate Editor for IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, and as a Guest Editor for IEEE JOURNAL OF SELECTED AND EMERGING TOPICS FOR INDUSTRIAL ELECTRONICS.

Peng Wang (Fellow, IEEE) received the B.Sc. degree in electronic engineering from Xian Jiaotong University, Xi'an, China, in 1978, the M.Sc. degree from Taiyuan University of Technology, Taiyuan, China, in 1987, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Saskatchewan, Saskatoon, SK, Canada, in 1995 and 1998, respectively. Currently, he is a full Professor with the School of Electrical and Electronic Engineering at Nanyang Technological University, Singapore. He is an Associate Editor or Guest Editor-in-Chief of IEEE Transactions on Smart Grid, IEEE Transactions on Power Delivery, Journal of Modern Power Systems and Clean Energy, CSEE Journal of Power and Energy Systems, and so on. His current research interests include power system planning and operation, renewable energy planning, solar/electricity conversion systems and power system reliability analysis.