Multi-scale Fusion Model Based on Gated Recurrent Unit for Enhancing Prediction Accuracy of State-of-charge in Battery Energy Storage Systems

Hao Liu, Fengwei Liang, Tianyu Hu, *Member*, *IEEE*, Jichao Hong, *Member*, *IEEE*, and Huimin Ma, *Member*, *IEEE*

***Abstract*—Accurate prediction of the state-of-charge (SOC) of battery energy storage system (BESS) is critical for its safety and lifespan in electric vehicles. To overcome the imbalance of existing methods between multi-scale feature fusion and global feature extraction, this paper introduces a novel multi-scale fusion (MSF) model based on gated recurrent unit (GRU), which is specifically designed for complex multi-step SOC prediction in practical BESSs. Pearson correlation analysis is first employed to identify SOC-related parameters. These parameters are then input into a multi-layer GRU for point-wise feature extraction. Concurrently, the parameters undergo patching before entering a dual-stage multi-layer GRU, thus enabling the model to capture nuanced information across varying time intervals. Ultimately, by means of adaptive weight fusion and a fully connected network, multi-step SOC predictions are rendered. Following extensive validation over multiple days, it is illustrated that the proposed model achieves an absolute error of less than 1.5% in real-time SOC prediction.**

***Index Terms*—Electric vehicle, battery energy storage system (BESS), state-of-charge (SOC) prediction, gated recurrent unit (GRU), multi-scale fusion (MSF).**

NOMENCLATURE

$\odot$	Hadamard product operator
$\sigma(\cdot)$	Sigmoid function
ω_1, ω_2	Learnable weight parameters
b_r	Learnable bias parameter of reset gate in gated recurrent unit (GRU)
b_z	Learnable bias parameter of update gate in GRU
b_h	Learnable bias parameter of candidate hidden state in GRU
C	Number of features in input information
h_{t-1}	Previous hidden state of a GRU
$\tilde{h}_t$	Candidate hidden state of a GRU
L	Length of input information
n	Number of samples
N	Number of input information patches
$P_i, \hat{P}_i$	Actual and predicted values
P	Length of a patch
r_d	Pearson correlation coefficient for driving
r_p	Pearson correlation coefficient for parked-charging
r_t	Reset gate of a GRU
S	Length of non-overlapping region between two consecutive patches
t	Time spot
W_{xr}, W_{hr}	Learnable weight parameters of reset gate in GRU
W_{xz}, W_{hz}	Learnable weight parameters of update gate in GRU
W_{xh}, W_{hh}	Learnable weight parameters of candidate hidden state in GRU
x_t	Input of a GRU
$\mathcal{X}$	(Explanatory) information
$\mathcal{X}_p$	Output after applying patching to $\mathcal{X}$
$\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{X}_h$	Outputs of a multi-layer GRU
X_i, Y_i	Values of the i^{th} datapoint
$\bar{X}, \bar{Y}$	Sample means of variables X and Y
z_t	Update gate of a GRU

I. INTRODUCTION

WITH the increasing global concern for environmental protection and sustainable development, electric vehicles (EVs) have become increasingly important in the energy

Manuscript received: September 29, 2023; revised: January 31, 2024; accepted: February 26, 2024. Date of CrossCheck: February 26, 2024. Date of online publication: March 30, 2024.

This work was supported in part by the National Natural Science Foundation of China (No. 62172036).

H. Liu, T. Hu (corresponding author), and H. Ma are with the School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, China (e-mail: m202220884@xs.ustb.edu.cn; tianyu@ustb.edu.cn; mhmpub@ustb.edu.cn).

F. Liang and J. Hong are with the School of Mechanical Engineering, University of Science and Technology Beijing, Beijing, China (e-mail: lfw20000316@163.com; qdbithong@163.com).

DOI: 10.35833/MPCE.2023.000726

sector [1]-[3]. Battery energy storage systems (BESSs) are core components in EVs that are essential for providing sustained power and reducing emissions. Thus, state-of-charge (SOC) prediction for lithium-ion batteries (LIBs), known for their fast charging, long life, and high energy density, has been widely studied [4]-[6]. The SOC of a BESS, which is monitored by a battery management system (BMS), is crucial for safe operation and risk prevention. Due to a variety of operational environments, accurate SOC prediction for LIBs remains challenging. SOC prediction methods generally fall into three categories: definition-based, model-based, and data-driven [7]-[11].

1) Definition-based methods. Definition-based methods for SOC prediction primarily encompass open circuit voltage (OCV) and ampere-hour integration methods, among which OCV methods are extensively utilized due to their accuracy [12]-[14]. Reference [15] showed that using an incremental OCV-based test improves SOC tracking accuracy under varying temperatures. Reference [16] further enhanced the estimation accuracy by incorporating a temperature model and an OCV-SOC table. Reference [17] applied recursive least squares and an adaptive extended Kalman filter based on Thevenin's model for precise SOC prediction, and they achieved robust noise resistance [17]. The ampere-hour integration method is commonly used for SOC estimation. However, it assumes a constant battery capacity and disregards the actual capacity fluctuations derived from varying usage and environmental conditions, which lead to estimation inaccuracies. To address this issue, [18] introduced an enhanced ampere-hour integration technique based on long short-term memory (LSTM) to reduce SOC estimation errors. And [19] considered factors such as temperature and aging in the new algorithm to improve precision.

2) Model-based methods. Model-based methods for SOC prediction mainly include filtering and equivalent circuit method (ECM) [20]-[22]. To address cumulative errors in traditional SOC estimation, [23] effectively mitigated estimation errors derived from Gaussian noise using an LIB Thevenin's model and an extended Kalman filtering (EKF) algorithm [23]. Reference [24] combined adaptive forgetting factor least-squares online identification with unscented Kalman filtering to enhance the accuracy of SOC estimation based on the ECM. Battery ECMs are also widely used in BMSs because of their ease of implementation. In general, ECM parameters vary with operating conditions; therefore, handling the dependence of parameters greatly affects the accuracy of the ECM over a wide operating range. Reference [25] alleviated the difficulty in specifying the parameter-functional SOC dependence of model parameters by converting the SOC-dependent ECM into a linear parameter variation input-output model, and proposed a nonparametric sparse Gaussian process regression (GPR) method [25]. Reference [26] achieved high SOC prediction accuracy by combining the ECM with an adaptive untraceable Kalman filter (AUKF) [26]. Model-based methods have the advantages of stability and high accuracy, but their complex model parameters make the prediction difficult.

3) Data-driven methods. Data-driven methods for SOC estimation encompass machine learning and neural networks [27]-[29], which have been applied extensively to create advanced SOC prediction methods without requiring additional details about battery chemistry, internal properties, or extra filters. Reference [30] demonstrated the outstanding performance of optimized machine-learning techniques for enhancing battery SOC prediction. Reference [31] introduced an adaptive SOC prediction method based on GPR, directly mapping battery parameters such as voltage, capacity, and temperature to the corresponding models. Low error rates were demonstrated with various types of batteries at 25 °C [31]. In addition to machine-learning methods, many studies have been conducted on neural-network-based SOC prediction methods. Reference [32] proposed an LSTM-based multi-step SOC prediction method that utilized actual vehicle data while accounting for weather and driver behaviors as influential factors [32]. Similarly, [33] proposed a novel multi-step SOC prediction method for real-world BESS by employing a gated recurrent unit (GRU) [33]. However, when deep learning techniques such as LSTM and GRU are employed for direct prediction, the predictive performance may exhibit limitations in capturing long-term dependency relationships.

The existing literature shows that although SOC prediction has been extensively studied, studies still remain limited. In addition, most existing works rely on laboratory conditions, leaving a gap in studies based on real-world EV data. Whereas some existing studies including [34]-[37] share methodological similarities with this paper, they exhibit an imbalance between multi-scale feature fusion and global information extraction from complex data. In response, this paper proposes an innovative model that utilizes multi-layer GRUs for multi-scale fusion, and it is specifically designed for SOC prediction. The innovation of this paper has promising implications in SOC prediction for real-world EVs. The contributions of this paper are summarized as follows.

1) To address the complexities of real-world BESSs, we introduce a novel multi-step SOC prediction model that leverages a GRU. The robustness and effectiveness of the proposed model in predicting the SOC are empirically validated through an exhaustive evaluation on a real-world driving dataset.

2) The proposed model captures multi-scale features by amalgamating a dual-stage patch GRU with a conventional GRU. It not only excels in identifying local characteristics but also enhances the model's ability to grasp long-term contextual nuances, leading to significant improvements in SOC prediction accuracy.

3) By leveraging learnable weights to fuse features at two levels, the model autonomously adjusts the significance of each feature based on the data. This not only enhances the prediction accuracy but also bolsters the robustness of the model.

The remainder of this paper is organized as follows. Section II describes the data and preprocessing. Section III introduces the methodology for SOC prediction. Section IV presents case studies. Section V concludes this paper.

II. Data Description and Preprocessing

A. Data Description

The dataset utilized in this paper consists of driving data from an EV collected from January to March with 10 s sampling intervals. The raw data encompass various attributes such as timestamps, charging modes, and driving status. In this paper, we initially select 15 data attributes for comprehensive data representation. Table I lists the detailed specifications of the selected attributes.

TABLE I
Detailed Specifications of Selected Attributes

Parameter	Description
Time	Data collection time with 10 s sampling intervals
Vehicle status	1: driving; 2: parking
Charging status	1: charging; 3: not charging; 4: fully charged
Mileage	35051-100335 km
Speed	0-135 km/h
Total voltage	315-410 V
Total current	−210-335 A
DC-DC status	1: operation; 2: disconnection
Gear	8-bit binary number
Insulation resistance	0-60000 kΩ
The maximum cell voltage	3.3-4.3 V
The minimum cell voltage	3.2-4.3 V
The maximum temperature	0-42 ℃
The minimum temperature	−2-41 ℃
SOC	0-100%

B. Data Preprocessing

In real-world BESSs, the parameters often engage in intricate nonlinear interactions, adding layers of complexity to the task of delineating the relationships between these parameters and the SOC of BESSs. To navigate this intricate landscape, a series of systematic steps are implemented to fine-tune the data analysis, as illustrated in Fig. 1. Initially, irrelevant data are filtered out from the original data. Then, data cleansing is performed to ensure data quality and consistency. Under both driving and parked-charging conditions, Pearson correlation analyses are conducted, revealing parameters directly linked to SOC changes including total current, total voltage, the maximum and minimum cell voltages, and the maximum and minimum temperatures. Finally, based on these key parameters, a new dataset is constructed using a sliding-window method. The dataset encompasses the data from January to February and is allocated to the training, testing, and validation sets at a ratio of 6:2:2. This preprocessing aims to enhance the understanding of nonlinear interactions in BESSs, thereby improving the accuracy of the SOC prediction model.

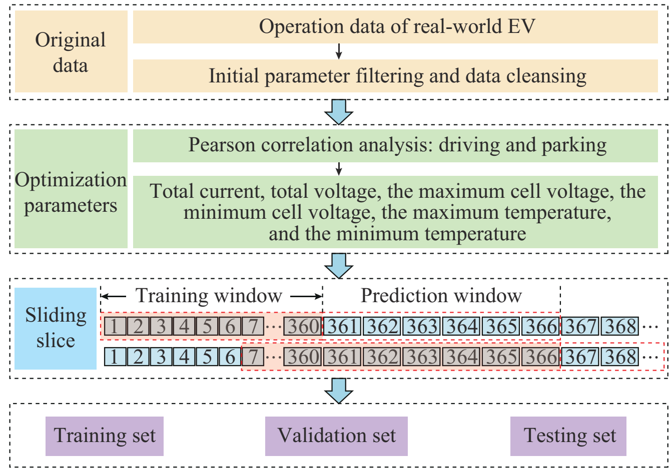

Fig. 1. Procedure for data preprocessing.

III. Methodology

A. GRU

A GRU [38] is a specialized form of recurrent neural network architecture designed to tackle the vanishing gradient problem and to model sequential data more effectively. To achieve these, the GRU integrates gating mechanisms that allow for enhanced learning of both short- and long-term sequence dependencies.

The core of the GRU architecture lies in its two gating mechanisms, i.e., reset and update gates. These mechanisms enable the GRU to manage information flow and control the balance between retaining past information and integrating new inputs during sequence processing. Figure 2 illustrates the structure of a multi-layer GRU and internal structure of a GRU cell.

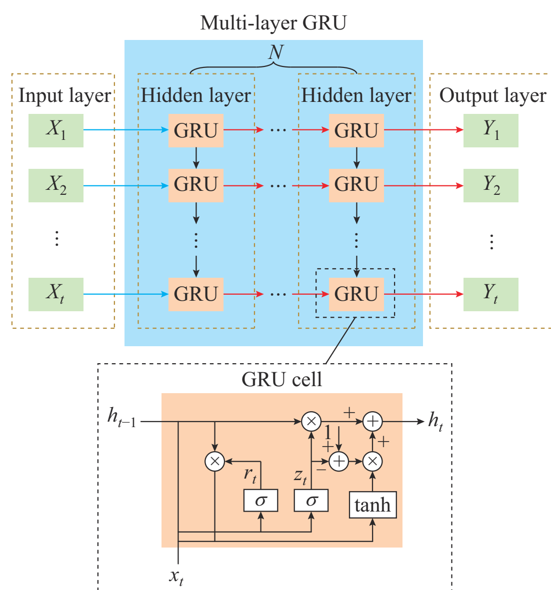

Fig. 2. Structure of a multi-layer GRU and internal structure of a GRU cell.

The reset gate r_t determines the extent to which the previous hidden state must be reset, while the current input is processed. r_t is calculated as:

$$r_t=\sigma(W_{xr}x_t+W_{hr}h_{t-1}+b_r) \tag{1}$$

The update gate z_t governs the degree to which a new input affects the updated hidden state. The calculations are as follows:

$$z_t=\sigma(W_{xz}x_t+W_{hz}h_{t-1}+b_z) \tag{2}$$

$$\tilde{h}_t=\tanh(W_{xh}x_t+W_{hh}(r_t\odot h_{t-1})+b_h) \tag{3}$$

$$h_t=z_t\odot h_{t-1}+(1-z_t)\odot\tilde{h}_t \tag{4}$$

B. Framework of Multi-scale Fusion Gated Recurrent Unit (MSFGRU)

The MSFGRU is a carefully engineered neural network architecture primarily aimed at boosting the predictive accuracy of the SOC of a BESS by improving its sequential modeling capabilities. To achieve this goal, the proposed architecture combines several key components and techniques.

1) Multi-layer GRU. Within the multi-layer GRU, each GRU layer independently extracts and learns features at its designated level and subsequently transmits these abstracted features to the subsequent layer. The lower layers are focused predominantly on capturing fundamental patterns, whereas the deeper layers are adept at discerning complex phenomena. This hierarchical feature-extraction mechanism provides a comprehensive understanding of the data. The cascaded architecture of a multi-layer GRU not only augments the model's capacity for feature extraction and data modeling but also facilitates a profound comprehension of the complex patterns and inherent interrelations present within the time-series data.

2) Patching. Extracting a specific semantic context is critical when analyzing correlations within time-series data. A majority of earlier research used only individual input elements, whereas recent studies have demonstrated enhanced results by adopting a segment-based input method known as "patching" [39]-[41]. This novel method involves aggregating time steps into subseries-level patches. This process substantially enriches immediate contextual comprehension, enabling the capture of comprehensive semantic information that is otherwise difficult to obtain when concentrating solely on isolated data points. The patching process generates a new sequence consisting of patches with a total length of N, where $N=\lfloor (L-P)/S \rfloor+1$, and $\lfloor\cdot\rfloor$ represents the floor function.

3) Multi-scale method. Within the framework of MSFGRU, we employ a sophisticated multi-scale method that handles input data across multiple layers of granularity. This nuanced method enables the model to extract temporal information from both individual data points and aggregated segments. Consequently, the model captures both short- and long-term dependencies inherent in the sequences.

4) Adaptive fusion method. We employ an adaptive fusion method that skillfully combines data from various scales to provide a comprehensive representation of an input sequence. This fusion process is dynamic and regulated by two learnable weight parameters, ω_1 and ω_2. This feature enhances the ability of the network to capture intricate temporal relationships, thereby boosting the overall accuracy of the model.

The proposed model takes as input a tensor $\mathcal{X}\in\mathbb{R}^{L\times C}$, where $L=360$ denotes the length of the input sequence and $C=6$ is the number of feature dimensions. Initially, $\mathcal{X}$ is fed into a multi-layer GRU, yielding a feature vector denoted by $\mathcal{Y}_1\in\mathbb{R}^{1\times H}$. In this context, $H=128$ specifies the number of hidden units within the GRU. Simultaneously, we subject $\mathcal{X}$ to a patching operation that generates a new segmented sequence represented as $\mathcal{X}_p\in\mathbb{R}^{N\times P\times C}$. Here, $P=60$ is the length of each segment, $N=\lfloor (L-P)/S \rfloor+1=\lfloor (360-60)/6 \rfloor+1=51$ indicates the total number of patches, and $S=6$ is the non-overlapping region between two consecutive patches. Subsequently, $\mathcal{X}_p$ is processed through another multi-layer GRU, generating a feature vector labeled $\mathcal{X}_h\in\mathbb{R}^{N\times H}$. Next, this new feature vector $\mathcal{X}_h$ is treated as a novel sequence and is fed into yet another multi-layer GRU, resulting in feature vector $\mathcal{Y}_2\in\mathbb{R}^{1\times H}$. Finally, $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are weighted by parameters ω_1 and ω_2, respectively, and their weighted sums produce the final feature representation. This aggregated feature representation is then passed through a fully connected prediction head to generate the desired SOC sequence. The complete learning process is formalized through the following equations.

$$\mathcal{Y}_1=MGRU_1(\mathcal{X}) \tag{5}$$

$$\mathcal{X}_p=Patching(\mathcal{X}) \tag{6}$$

$$\mathcal{Y}_2=MGRU_3(MGRU_2(\mathcal{X}_p)) \tag{7}$$

$$SOC=Linear(\omega_1\mathcal{Y}_1+\omega_2\mathcal{Y}_2) \tag{8}$$

where $MGRU_1(\cdot)$-$MGRU_3(\cdot)$ are multi-layer GRU networks; $Patching(\cdot)$ is a patching operation; and $Linear(\cdot)$ is a fully connected network. The above MSFGRU structure based on a multi-layer GRU is shown in Fig. 3.

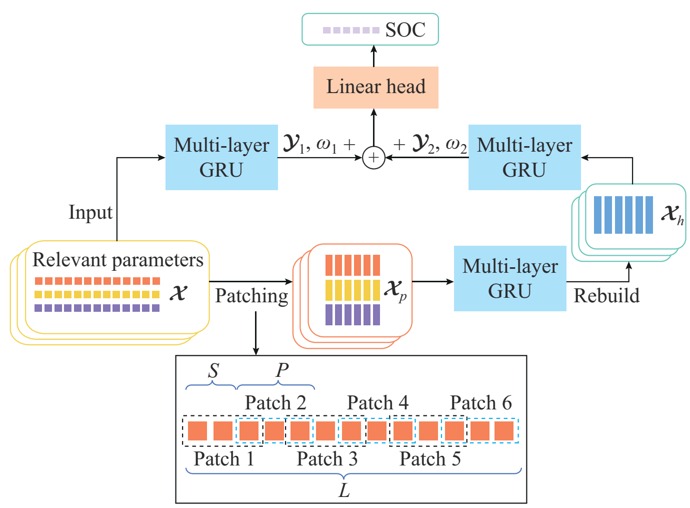

Fig. 3. MSFGRU structure based on a multi-layer GRU.

IV. Case Study

A. Evaluation Metrics

The mean squared error (MSE) and mean relative error (MRE) are instrumental in assessing the prediction performance of the proposed model, and their respective formulae are delineated as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(P_i - \hat{P}_i)^2 \tag{9}$$

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{P_i}\left|P_i - \hat{P}_i\right| \times 100\% \tag{10}$$

B. Objective of MSFGRU Optimization

The objective of MSFGRU optimization is to minimize the average squared error between the predicted and actual sequences, which is expressed as:

$$\min \frac{1}{n}\sum_{i=1}^{n}(P_i - \hat{P}_i)^2 \tag{11}$$

C. Pearson Correlation Analysis

The Pearson correlation coefficient is a statistical measure used to assess the strength of the linear relationship between two variables, which can be calculated by:

$$\rho_{X,Y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{12}$$

The calculated value falls within the range of −1 to 1, indicating the strength and direction of the linear connection between the two factors. If the value exceeds 0.4, a noticeable positive correlation is observed. Therefore, variables within this range are selected as relevant variables.

Based on the dataset information on the vehicle status and charging conditions, the data can be categorized into two primary scenarios: active driving and parked-charging. In this paper, Pearson correlation analyses are performed separately for these two scenarios, and the results are presented in Table II.

TABLE II
PEARSON COEFFICIENTS BETWEEN VARIABLES AND SOC

Feature	r_d	r_p	Feature	r_d	r_p
Time	NAN	NAN	DC-DC status	NAN	NAN
Vehicle status	NAN	NAN	Gear	0.026	NAN
Charging status	NAN	NAN	Insulation resistance	0.022	0.017
Mileage	−0.049	−0.045	The maximum cell voltage	0.980	0.980
Speed	−0.077	NAN	The minimum cell voltage	0.970	0.980
Total voltage	0.980	0.980	The maximum temperature	0.510	0.490
Total current	−0.015	0.500	The minimum temperature	0.480	0.490

In the active driving scenario, the parameters that exhibit significant correlations with the SOC include the total voltage, the maximum cell voltage, the minimum cell voltage, the maximum temperature, and the minimum temperature. By contrast, in the parked-charging scenario, the total current shows a notable correlation with the SOC. Consequently, these observations lead to the identification of the optimal input variables for the prediction model, including the total voltage, the total current, the maximum cell voltage, the minimum cell voltage, the maximum temperature, and the minimum temperature.

D. Training of MSFGRU

A grid search method has been employed to determine the optimal configuration of the MSFGRU. Within this framework, the parameter combination of {batch size, layers, hidden size, patch length} yielding the minimum average loss in the validation set is selected as the optimal solution. Table III lists additional details of the hyperparameters and results of the grid search method.

TABLE III
HYPERPARAMETERS AND RESULTS OF GRID SEARCH METHOD

Hyperparameter	Parameter value	Determination
Epoch	Initial number is 100	Early stopping
Learning rate	Initial value is 0.0001	Exponential decay
Batch size	[32, 64, 128, 256]	64
Layer	[1, 2, 3, 4, 5]	4
Hidden size	[64, 128, 256, 512]	128
Patch length	[12, 24, 36, 48, 60, 72]	60

To enhance the training efficacy, a dual method of early stopping and adaptive learning rate reduction is adopted. As Fig. 4 shows, the training period is initially set to be 100 epochs. However, with a continuous decline in performance over 10 consecutive epochs on the validation set, the training is terminated to prevent model overfitting. Concurrently, at an initial learning rate of 0.0001, a gradual reduction in the learning rate is implemented throughout the training process to ensure effective convergence of the model.

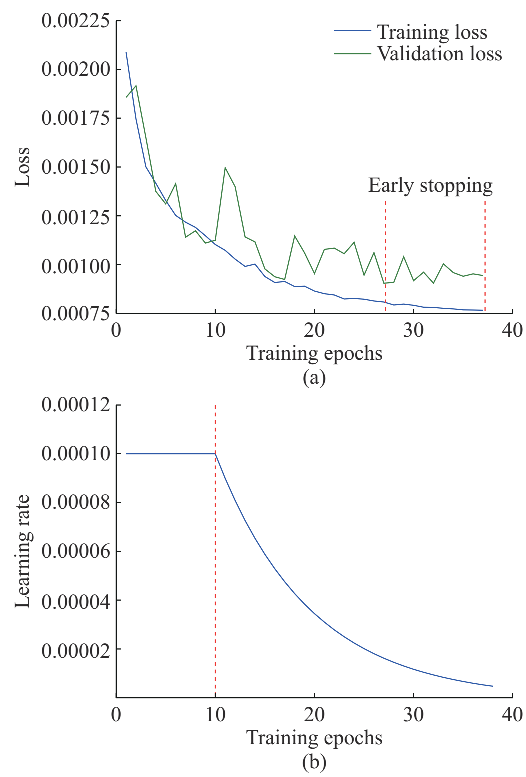

Fig. 4. Early stopping and adaptive learning rate reduction. (a) Training and validation losses. (b) Learning rate.

E. Comparison with Test Set

For a rigorous assessment of the efficacy of the proposed MSFGRU model, we benchmark it against the well-established baseline models such as linear regression (LR), LSTM [32], GRU with dropout [33], Transformer [42], and PatchTST [40]. As Table IV illustrates, LR exhibits the least favorable performance in the prediction tasks, followed by LSTM. These two models underperform when dealing with data characterized by high dynamism and complexity. By contrast, GRU demonstrates superior predictive performance. However, whereas Transformer and PatchTST underperform as compared with the GRU in short-term prediction, their performance progressively improves with the extension of the prediction horizon, eventually matching or surpassing that of the GRU. This enhancement is primarily attributed to the self-attention mechanisms that proficiently address long-term dependencies. The proposed MSFGRU model achieves an average improvement of 16% in MSE and 5% in MRE as compared with the best-performing baseline model. In addition, as Fig. 5 shows, the model consistently maintains an absolute error of less than 1.5% for 1 min predictions.

TABLE IV
TEST PERFORMANCE OF MSFGRU COMPARED WITH BASELINE MODELS

Model	Error type	Performance with different lead steps (%)			
		1 min	2 min	3 min	5 min
LR	MRE	0.628	0.647	0.673	0.741
	MSE	0.318	0.339	0.367	0.456
LSTM	MRE	0.567	0.596	0.613	0.666
	MSE	0.225	0.263	0.282	0.349
GRU	MRE	0.523	0.551	0.577	0.625
	MSE	0.223	0.240	0.266	0.328
Transformer	MRE	0.544	0.565	0.578	0.635
	MSE	0.232	0.246	0.260	0.330
PatchTST	MRE	0.539	0.552	0.569	0.623
	MSE	0.229	0.238	0.254	0.322
MSFGRU	MRE	0.496	0.521	0.548	0.606
	MSE	0.181	0.196	0.221	0.294

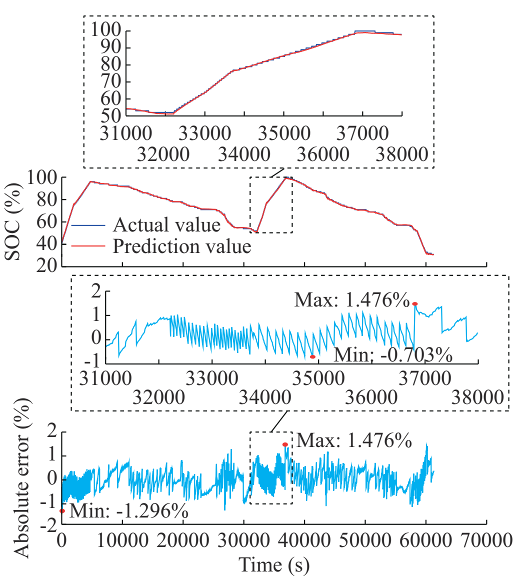

Fig. 5. Performance of MSFGRU: 1 min predictions on test dataset.

Selected results from the baseline experiments are obtained, with a particular focus on two scenarios: parked-charging, and post-charging driving. As shown in Figs. 5 and 6, in the parked-charging phase, the proposed MSFGRU and conventional GRU both exhibit exceptional performances, maintaining an absolute error of approximately 1%. By contrast, the LR, LSTM, Transformer, and PatchTST display significantly higher absolute errors. During the transition to the post-charging driving phase, the absolute error for the GRU increases significantly, whereas that for the MSFGRU experiences only a slight increase. This further confirms the distinct advantage of the MSFGRU in maintaining high stability under varying operational conditions.

Under various real-world driving conditions, the proposed model notably outperforms other baseline models chiefly due to its multi-scale input and adaptive fusion mechanisms. First, the proposed model utilizes a standard GRU to capture short-term dependencies while concurrently employing a two-stage GRU with patching operations to effectively grasp long-term dependencies. This method, which integrates both short-term and long-term dependencies, endows the model with a more comprehensive insight into the data. Second, the adaptive fusion mechanism enables the proposed model to dynamically adjust the weights assigned to short- and long-term dependencies, allowing it to adapt to different driving scenarios. Finally, the proposed model incorporates a bidirectional Kalman filter to smooth the output results, further enhancing its prediction accuracy and stability.

F. Ablation Experiments

To corroborate the effectiveness and necessity of each individual component within the MSFGRU, we have conducted additional ablation experiments, the results of which are shown in Table V. We label the individual multi-layer GRU as Module I and the patching GRU as Module P, with learnable weights denoted by W. Here, K signifies the bidirectional Kalman filter used for output smoothing. It is noted that when I and P are not used simultaneously, W becomes ineffective; when I and P take effect while W remains inactive, $\omega_1 = \omega_2 = 0.5$. From Table V, the following remarks are obtained.

1) In contrast to the traditional GRU, which relies solely on point-wise operations, the enhanced GRU, which utilizes patching operations, exhibits remarkable proficiency. It captures both short- and long-term dependencies embedded in the data more effectively. This augmentation significantly improves the overall performance of the proposed model.

2) By combining the features gathered from both traditional point-wise operations and advanced patching techniques, we enhance the ability to predict the SOC of a BESS accurately. In addition, our method outperforms basic weighted averaging by capturing complex inter-feature relationships, thereby improving the overall accuracy of SOC prediction.

G. Further Confirmation of Temporal Stability

In the task of predicting the SOC of a BESS for the upcoming 1-5 min period, the MSFGRU model demonstrates remarkable temporal stability. The training, validation, and testing phases are conducted using data from January to February.

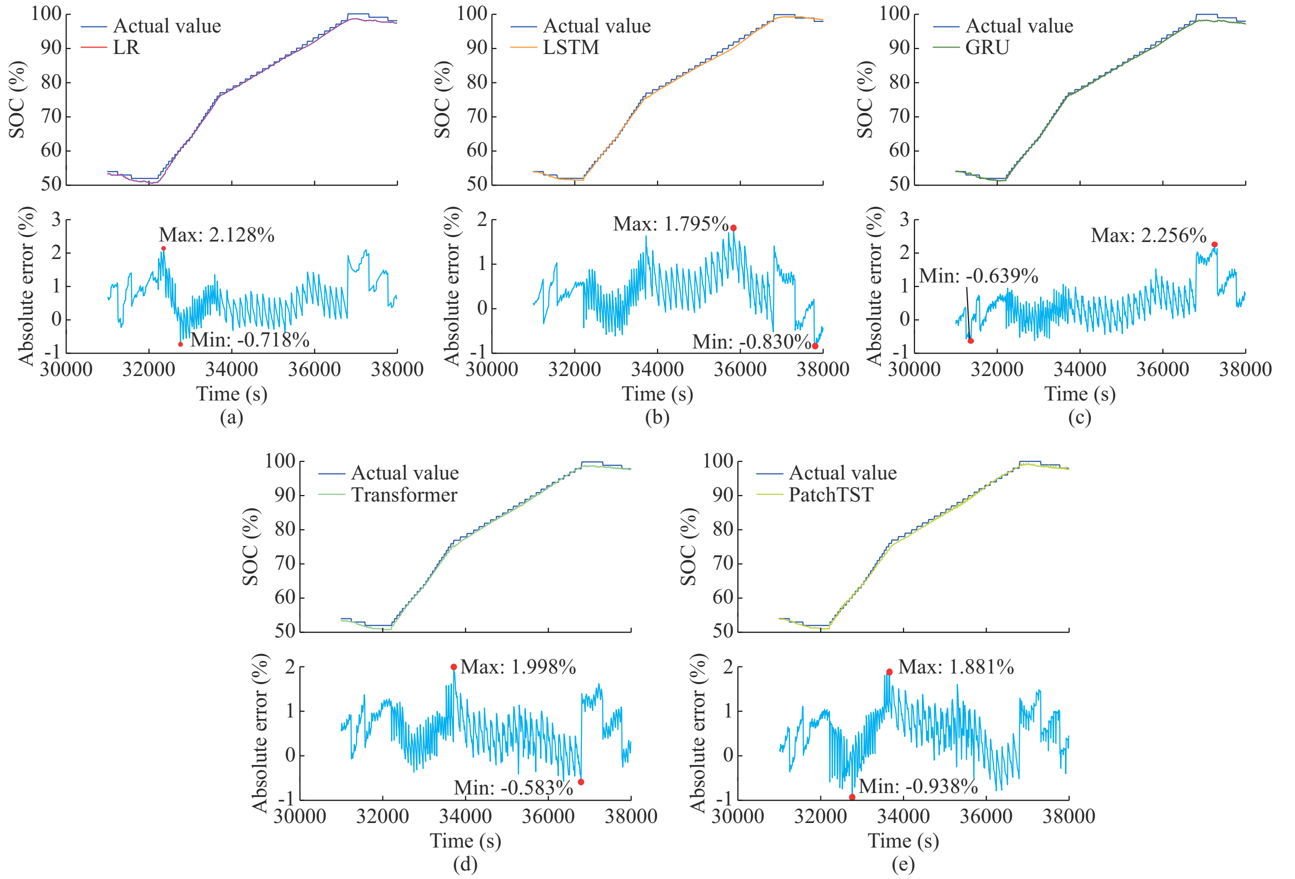

Fig. 6. Performance comparison of different models: 1 min predictions on test dataset. (a) LR. (b) LSTM. (c) GRU. (d) Transformer. (e) PatchTST.

TABLE V
ABLATION EXPERIMENT RESULTS FOR MSFGRU

Module				Test set	
I	P	W	K	MRE (%)	MSE (%)
√				0.523	0.232
	√			0.529	0.225
√	√			0.517	0.200
√	√	√		0.505	0.186
√	√	√	√	0.496	0.181

To further validate the temporal robustness of the proposed model, supplemental tests are executed over nine nonconsecutive days in March, when the data are relatively complete. The performance results are presented in Table VI.

Figure 7 further elucidates the performance of the MSFGRU in predicting SOC for the next 1 min interval, with all observed absolute errors falling within a 1.5% margin. This not only robustly validates the inherent temporal stability of the proposed model but also underscores its capacity for high-accuracy SOC prediction across different time frames in the future.

H. Effects of Hyperparameters

The results vary with adjustments to the optimized hyperparameters. Based on the optimal solution determined through a grid search, we individually analyze the effect of each hyperparameter, as shown in Fig. 8.

TABLE VI
PERFORMANCE TESTED OVER NINE NONCONSECUTIVE DAYS IN MARCH

Lead step (min)	MRE (%)	MSE (%)
1	0.569	0.204
2	0.600	0.222
3	0.620	0.241
5	0.653	0.292

1) Batch size. To utilize computational resources efficiently and to accelerate the training process, a large dataset is segmented into smaller batches. The optimal batch size is determined through a grid search considering the conventional range of {32, 64, 128, 256}. As Fig. 8(a) shows, the model performs optimally under a batch size of 64. We should note that excessively small batches may lead to unstable training and slower convergence, whereas excessively large batches could increase memory demands and predispose the training to converge to local optima.

2) Number of hidden layers. The number of layers significantly affects the ability to discern complex data patterns. By conducting a grid search within the range of {1, 2, 3, 4, 5} layers, we discover that the four-layer structure yields optimal performance, as illustrated in Fig. 8(b). Insufficient layers may impede the ability to model intricate relationships, whereas an excessive number of layers can increase the risk of vanishing gradients and overfitting. An appropriate number of layers is crucial for maximizing the efficacy of the model.

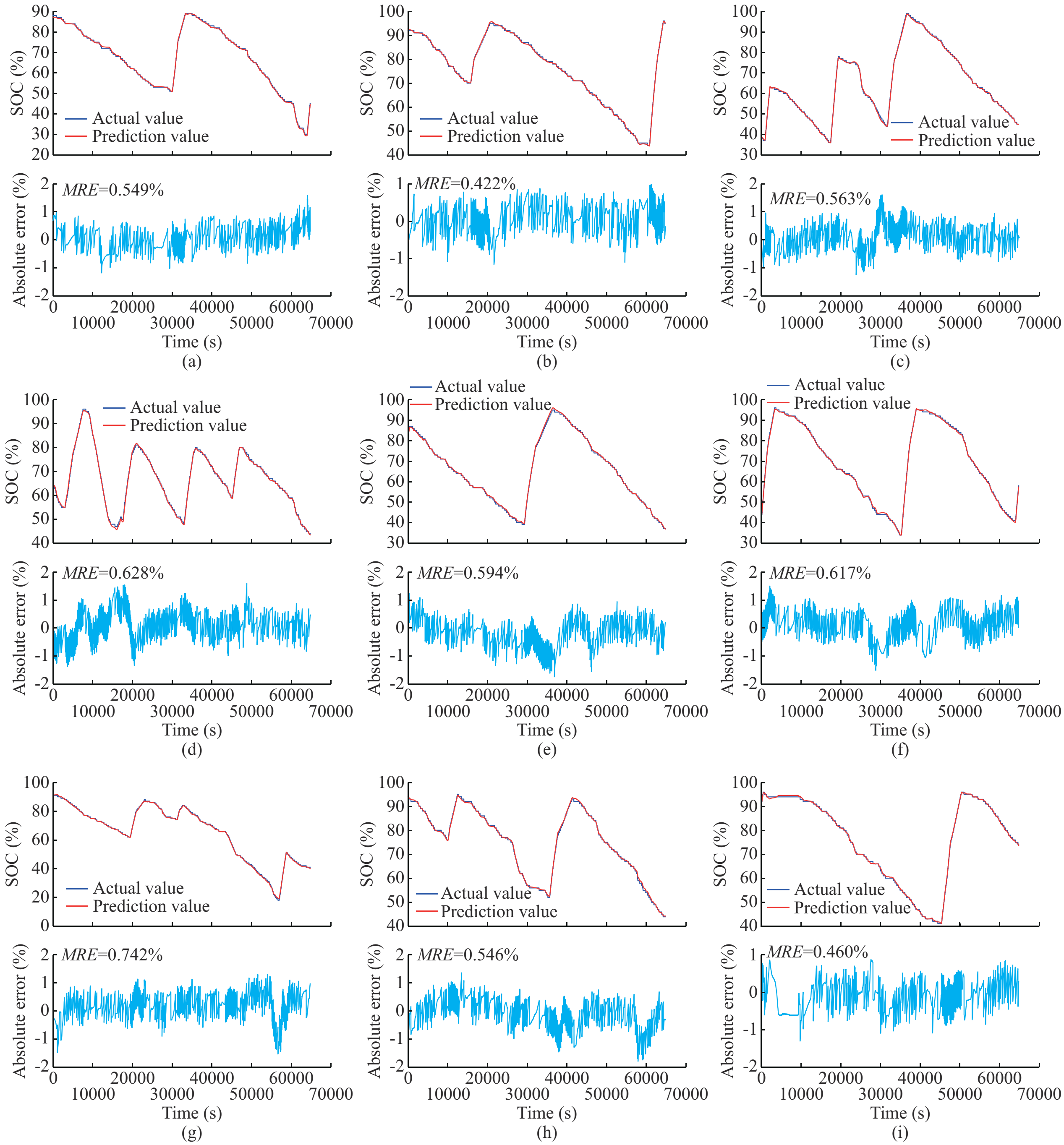

Fig. 7. Performance of MSFGRU in predicting SOC for next 1 min interval. (a) Day 1. (b) Day 2. (c) Day 3. (d) Day 4. (e) Day 5. (f) Day 6. (g) Day 7. (h) Day 8. (i) Day 9.

3) Number of hidden units. The number of hidden units in the GRU significantly affects the model performance and training process. Our utilization of the grid search ranges from {64, 128, 256, 512}. As Fig. 8(c) shows, the model performs optimally with 128 hidden units. Fewer hidden units may struggle to capture temporal relationships effectively, whereas a greater number increases the risk of overfitting and computational complexity, thereby affecting the prediction performance.

4) Patching length. To effectively capture local patterns in time-series data, the dataset is segmented into overlapping fragments. Considering a data collection frequency of six times per minute, a non-overlapping region of six data points between adjacent segments is established, with the minimum segment length set to be 12 data points. During the grid search, the range of the segment lengths is set to be {12, 24, 36, 48, 60, 72}. As shown in Fig. 8(d), the model performed optimally with a segment length of 60 data points. Although increasing the segment length generally improves the predictive performance, excessive overlap beyond an optimal length introduces redundancy and consequently diminishes the performance.

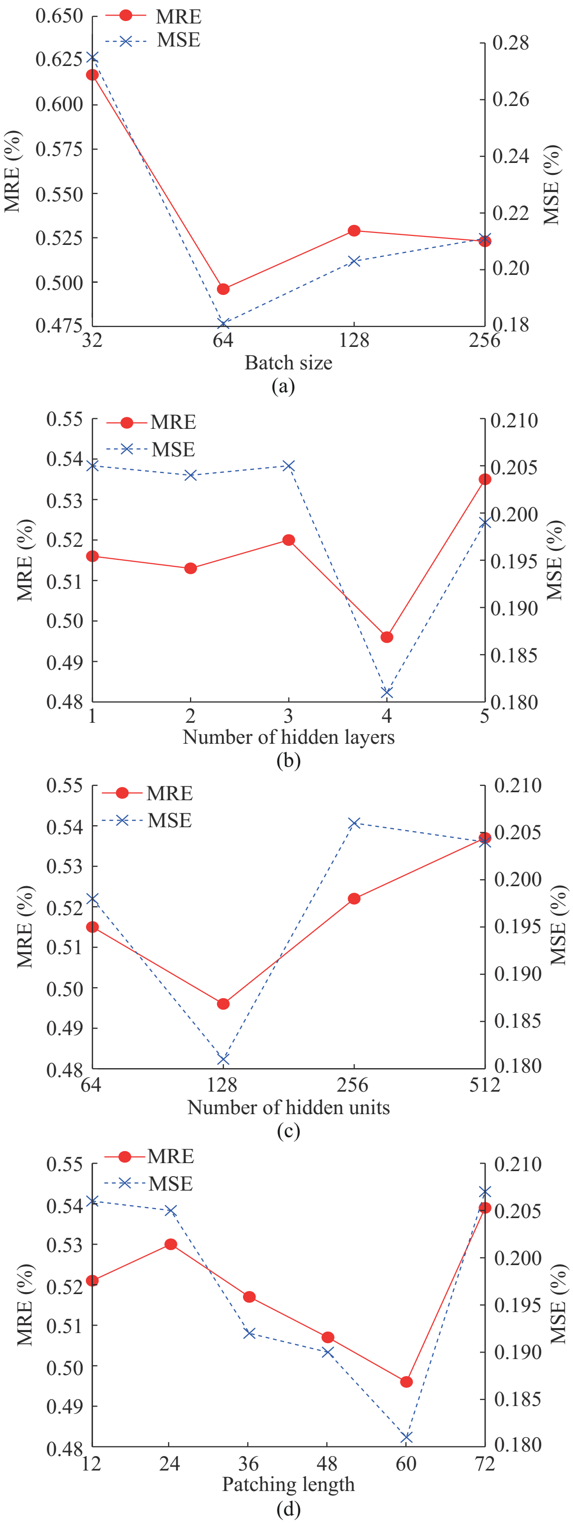

Fig. 8. Effects of hyperparameters on experimental results. (a) Batch size. (b) Number of hidden layers. (c) Number of hidden units. (d) Patching length.

V. Conclusion

This paper has successfully developed an MSFGRU model that is specifically designed to provide accurate prediction for the SOC of EV batteries within a time span ranging from 1-5 min. By employing Pearson correlation analysis of authentic driving data, six key parameters are identified that have a positive correlation with the SOC. Feature extraction is conducted at both point- and patch-wise levels, and this is followed by the adaptive fusion of these features using dynamically adjusted weights. These integrated features are subsequently fed into a fully connected network for multi-step predictions. In further testing conducted over nine complete days in March, the model consistently shows high accuracy, which confirms its temporal stability. In summary, the MSFGRU model excels in terms of temporal stability, generalization capabilities, and prediction accuracy, making it a highly promising tool for future BESSs in EVs and extending its potential for broad practical applications.

References

[1] X. Ren, S. Liu, X. Yu *et al.*, "A method for state-of-charge estimation of lithium-ion batteries based on PSO-LSTM," *Energy*, vol. 234, p. 121236, Nov. 2021.

[2] J. Hong, Z. Wang, T. Zhang *et al.*, "Research on integration simulation and balance control of a novel load isolated pure electric driving system," *Energy*, vol. 189, p. 116220, Dec. 2019.

[3] I. Babaeiyazdi, A. Rezaei-Zare, and S. Shokrzadeh, "State of charge prediction of EV Li-ion batteries using EIS: a machine learning approach," *Energy*, vol. 223, p. 120116, May 2021.

[4] Z. Li, J. Huang, B. Y. Liaw *et al.*, "On state-of-charge determination for lithium-ion batteries," *Journal of Power Sources*, vol. 348, pp. 281-301, Apr. 2017.

[5] Z. Zhang, X. Cheng, Z. Lu *et al.*, "SoC estimation of lithium-ion battery pack considering balancing current," *IEEE Transactions on Power Electronics*, vol. 33, no. 3, pp. 2216-2226, May 2017.

[6] V. Chandran, C. K. Patil, A. Karthick *et al.*, "State of charge estimation of lithium-ion battery for electric vehicles using machine learning algorithms," *World Electric Vehicle Journal*, vol. 12, no. 1, p. 38, Mar. 2021.

[7] J. P. Rivera-Barrera, N. Muñoz-Galeano, and H. O. Sarmiento-Maldonado, "SOC estimation for lithium-ion batteries: review and future challenges," *Electronics*, vol. 6, no. 4, p. 102, Nov. 2017.

[8] M. A. Hannan, M. H. Lipu, A. Hussain *et al.*, "A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: challenges and recommendations," *Renewable and Sustainable Energy Reviews*, vol. 78, pp. 834-854, Oct. 2017.

[9] J. Meng, M. Ricco, G. Luo *et al.*, "An overview and comparison of online implementable SOC estimation methods for lithium-ion battery," *IEEE Transactions on Industry Applications*, vol. 54, no. 2, pp. 1583-1591, Nov. 2017.

[10] D. N. How, M. Hannan, M. H. Lipu *et al.*, "State of charge estimation for lithium-ion batteries using model-based and data-driven methods: a review," *IEEE Access*, vol. 7, pp. 136116-136136, Sept. 2019.

[11] W. Zhou, Y. Zheng, Z. Pan *et al.*, "Review on the battery model and SOC estimation method," *Processes*, vol. 9, no. 9, p. 1685, Sept. 2021.

[12] K. Zhang, R. Xiong, Q. Li *et al.*, "A novel pseudo-open-circuit voltage modeling method for accurate state-of-charge estimation of $LiFePO_4$ batteries," *Applied Energy*, vol. 347, p. 121406, Oct. 2023.

[13] U. N. Jibhkate and U. B. Mujumdar, "Development of low complexity open circuit voltage model for state of charge estimation with novel curve modification technique," *Electrochimica Acta*, vol. 429, p. 140944, Oct. 2022.

[14] Y. Wang, Y. Cheng, Y. Xiong *et al.*, "Estimation of battery open-circuit voltage and state of charge based on dynamic matrix control-extended Kalman filter algorithm," *Journal of Energy Storage*, vol. 52, p. 104860, Aug. 2022.

[15] F. Zheng, Y. Xing, J. Jiang *et al.*, "Influence of different open circuit voltage tests on state of charge online estimation for lithium-ion batteries," *Applied energy*, vol. 183, pp. 513-525, Dec. 2016.

[16] Y. Xing, W. He, M. Pecht *et al.*, "State of charge estimation of lithium-ion batteries using the open-circuit voltage at various ambient temperatures," *Applied Energy*, vol. 113, pp. 106-115, Jan. 2014.

[17] R. Xiao, Y. Hu, X. Jia *et al.*, "A novel estimation of state of charge for the lithium-ion battery in electric vehicle without open circuit voltage experiment," *Energy*, vol. 243, p. 123072, Mar. 2022.

[18] X. Zhang, J. Hou, Z. Wang *et al.*, "Study of SOC estimation by the ampere-hour integral method with capacity correction based on

LSTM," *Batteries*, vol. 8, no. 10, p. 170, Oct. 2022.
[19] D. Liu, B. Huang, Y. Wang *et al.*, "Inflection point Ah-total integration method for real-time integration to correct lithium battery SOC," *Energy Storage Science and Technology*, vol. 8, no. 5, p. 850, Jun. 2019.
[20] P. Shrivastava, T. K. Soon, M. Y. I. B. Idris *et al.*, "Model-based state of *X* estimation of lithium-ion battery for electric vehicle applications," *International Journal of Energy Research*, vol. 46, no. 8, pp. 10704-10723, Mar. 2022.
[21] L. Wu, H. Pang, Y. Geng *et al.*, "Low-complexity state of charge and anode potential prediction for lithium-ion batteries using a simplified electrochemical model-based observer under variable load condition," *International Journal of Energy Research*, vol. 46, no. 9, pp. 11834-11848, Apr. 2022.
[22] L. Wang, J. Ma, X. Zhao *et al.*, "Adaptive robust unscented Kalman filter-based state-of-charge estimation for lithium-ion batteries with multi-parameter updating," *Electrochimica Acta*, vol. 426, p. 140760, Sept. 2022.
[23] Z. Cui, W. Hu, G. Zhang *et al.*, "An extended Kalman filter based SOC estimation method for Li-ion battery," *Energy Reports*, vol. 8, pp. 81-87, Aug. 2022.
[24] H. Wang, Y. Zheng, and Y. Yu, "Lithium-ion battery SOC estimation based on adaptive forgetting factor least squares online identification and unscented Kalman filter," *Mathematics*, vol. 9, no. 15, p. 1733, Jul. 2021.
[25] K. Fan, Y. Wan, and B. Jiang, "State-of-charge dependent equivalent circuit model identification for batteries using sparse Gaussian process regression," *Journal of Process Control*, vol. 112, pp. 1-11, Apr. 2022.
[26] X. Lin, Y. Tang, J. Ren *et al.*, "State of charge estimation with the adaptive unscented Kalman filter based on an accurate equivalent circuit model," *Journal of Energy Storage*, vol. 41, p. 102840, Sept. 2021.
[27] M. H. Lipu, M. Hannan, A. Hussain *et al.*, "Data-driven state of charge estimation of lithiumion batteries: algorithms, implementation factors, limitations and future trends," *Journal of Cleaner Production*, vol. 277, p. 124110, Dec. 2020.
[28] Z. Deng, X. Hu, X. Lin *et al.*, "Data-driven state of charge estimation for lithium-ion battery packs based on Gaussian process regression," *Energy*, vol. 205, p. 118000, Aug. 2020.
[29] C. Pan, Z. Peng, S. Yang *et al.*, "Adaptive neural network-based prescribed-time observer for battery state-of-charge estimation," *IEEE Transactions on Power Electronics*, vol. 38, no. 1, pp. 165-176, Sept. 2022.
[30] M. A. Hannan, M. H. Lipu, A. Hussain *et al.*, "Toward enhanced state of charge estimation of lithium-ion batteries using optimized machine learning techniques," *Scientific Reports*, vol. 10, no. 1, p. 4687, Mar. 2020.
[31] O. Ali, M. K. Ishak, A. B. Ahmed *et al.*, "On-line WSN SOC estimation using Gaussian process regression: an adaptive machine learning approach," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 9831-9848, Dec. 2022.
[32] J. Hong, Z. Wang, W. Chen *et al.*, "Online joint-prediction of multi-forward-step battery SOC using LSTM neural networks and multiple linear regression for real-world electric vehicles," *Journal of Energy Storage*, vol. 30, p. 101459, Aug. 2020.
[33] R. Li, H. Wang, H. Dai *et al.*, "Accurate state of charge prediction for real-world battery systems using a novel dual-dropout-based neural network," *Energy*, vol. 250, p. 123853, Jul. 2022.
[34] L. Ren, X. Cheng, X. Wang *et al.*, "Multi-scale dense gate recurrent unit networks for bearing remaining useful life prediction," *Future Generation Computer Systems*, vol. 94, pp. 601-609, May 2019.
[35] X. Shao, C. S. Kim, and P. Sontakke, "Accurate deep model for electricity consumption forecasting using multi-channel and multi-scale feature fusion CNN-LSTM," *Energies*, vol. 13, no. 8, p. 1881, Apr. 2020.
[36] W. Xu, H. Miao, Z. Zhao *et al.*, "Multi-scale convolutional gated recurrent unit networks for tool wear prediction in smart manufacturing," *Chinese Journal of Mechanical Engineering*, vol. 34, no. 1, p. 53, Jun. 2021.
[37] T. Wang and Y. Qin, "A novel multi-scale convolutional network with attention-based bidirectional gated recurrent unit for atrial fibrillation discrimination," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 2, pp. 445-455, Apr. 2021.
[38] J. Chung, C. Gulcehre, K. Cho *et al.* (2014, Dec.). Empirical evaluation of gated recurrent neural networks on sequence modeling. [Online]. Available: https://arxiv.org/abs/1412.3555
[39] Y. Zhang and J. Yan, "Crossformer: transformer utilizing cross-dimension dependency for multivariate time series forecasting," in *Proceedings of The Eleventh International Conference on Learning Representations*, Kigali, Rwanda, May 2023, pp. 1-21.
[40] Y. Nie, N. H. Nguyen, P. Sinthong *et al.* (2022, Nov.). A time series is worth 64 words: long-term forecasting with transformers. [Online]. Available: https://arxiv.org/abs/2211.14730
[41] D. Du, B. Su, and Z. Wei, "Preformer: predictive transformer with multi-scale segment-wise correlations for long-term time series forecasting," in *Proceedings of 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece, Jun. 2023, pp. 1-5.
[42] A. Vaswani, N. Shazeer, N. Parmar *et al.*, "Attention is all you need," in *Proceedings of 31st Annual Conference on Neural Information Processing Systems (NIPS)*, Long Beach, USA, Dec. 2017, p. 30.

Hao Liu received the B.E. degree from the School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, China, in 2022, where he is currently pursuing the M.E. degree in computer technology. His research interests include deep learning and renewable energy prediction.

Fengwei Liang received the B.E. degree from Hebei Agricultural University, Baoding, China, in 2022. He is currently pursuing the M.E. degree with the Department of Vehicle Engineering, School of Mechanical Engineering, University of Science and Technology Beijing, Beijing, China. His research interests include state safety of electric vehicles.

Tianyu Hu received the B.E. degree from the School of Electrical Engineering, Shandong University, Jinan, China, in 2015, and the Ph.D. degree from Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, Shenzhen, China, in 2020. He joined the School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, China, in September 2020, as an Associate Professor. His research interests include optimization and cognitive decision-making, computer simulation, and few-shot learning.

Jichao Hong received the Ph.D. degree in automotive engineering from the Beijing Institute of Technology, Beijing, China, in 2020. He is currently an Associate Professor with the School of Mechanical Engineering, University of Science and Technology Beijing, Beijing, China. He is the Director of IEEE PES China Electric Vehicle Technical Committee Power Battery Technology Sub-Committee, and a Senior Member of the Youth Working Committee of Chinese Electrotechnical Society. His research interests include new energy power system, big data mining analysis, and intelligent safety control.

Huimin Ma received the Ph. D. degree in mechanical and electrical engineering from Beijing Institute of Technology, Beijing, China, in 2001. She was the Director of the 3D Imaging Lab, Department of Electronic Engineering, Tsinghua University, Beijing, China. She is currently a Professor with the School of Computer and Communication, University of Science and Technology Beijing, Beijing, China. She is also the Dean of the Department of Internet of Things and Electronic Engineering and the Vice President of the Institute of Artificial Intelligence. She is also the Secretary-General of the China Society of Image and Graphics. Her research interests include 3D image cognition and simulation.