IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 25, 2021, accepted February 3, 2021, date of publication February 8, 2021, date of current version February 18, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3058018

A Dynamic Spatial-Temporal Attention-Based GRU Model With Healthy Features for State-of-Health Estimation of Lithium-Ion Batteries

SHENGMIN CUI AND INWHEE JOE
Department of Computer Science, Hanyang University, Seoul 04763, South Korea

Corresponding author: Inwhee Joe (iwjoe@hanyang.ac.kr)

This work was supported by the Institute for Information and Communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (Development of the technology to automate the recommendations for big data analytic models that define data characteristics and problems), under Grant 2020-0-00107.

ABSTRACT A proper battery management system (BMS) plays a vital role in ensuring the safety and reliability of electric vehicles (EVs) and other electronic products. Accurate State-of-Health (SOH) estimation of Lithium-ion (Li-ion) batteries is a key factor in a BMS. It is difficult to determine SOH because of the complexity of the electrochemical reactions within the battery. To improve the accuracy of SOH estimation, a dynamic spatial-temporal attention-based gated recurrent unit (DSTA-GRU) model is proposed in this paper. First, we extract six features from the battery's charging and discharging processes that can reflect the aging degree of the battery to some extent. Second, this paper proposes a model to combine spatial attention and temporal attention that can not only consider the effects of states at different time step on the results, but also consider the effects of different features in the space domain. Third, the proposed model is trained and tested on NASA battery datasets and compared with other conventional models. Experiments carried on these data sets demonstrate that our model achieves higher accuracy than other conventional models.

INDEX TERMS Lithium-ion battery, state of health, dynamic spatial attention, temporal attention, gated recurrent unit.

I. INTRODUCTION

Due to the characteristics of high energy density and long lifetime, battery stacks based on Lithium-ion (Li-ion) batteries have been used in many fields, such as hybrid electric vehicles (HEVs), electric vehicles (EVs), ships and satellites [1]–[3]. However, improper use of Li-ion batteries such as overcharging or overdischarging may cause batteries to heat up, which will further affect the battery life and cause safety risks. Besides, the performance of Li-ion batteries will degrade as their capacity decrease and impedance increase, causing failures to electronic devices and systems [3]. In order to optimize power utilization, extend battery life, improve battery performance, and ensure the safe operation of electronic devices, Li-ion batteries must be managed more effectively [4]. A battery management system (BMS) can supervise power, maintain their proper use, and avoid potential accidents [5]. State-of-health (SOH) estimation which can provide accurate knowledge about the internal state of batteries is one of the primary roles of a BMS [6]–[8]. SOH is an indicator of battery aging, which is usually referred to as capacity or power degradation. It provides very useful information for when to remove or replace batteries. If the performance of the battery is abnormal, we should be able to tell from this parameter. Although relatively accurate SOH can be obtained by measuring the data of the discharge process under laboratory conditions using special high-precision equipment, this method has special limitations on the current and temperature of the discharge. These settings are not realistically available with commercial BMS in general. It is important to use proper methods for SOH estimation since SOH cannot be measured directly with

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Jiang.

commercially available sensors [4]. As a result, a great deal of research in the field of BMS has been focused on accurately estimating SOH.

The interior of a Li-ion battery is a very complex system and there are several factors that affect the health and aging process of the Li-ion battery. Power fading and capacity decrease are not results of a single cause, but of many different processes and their interactions [9]. Therefore, when studying the mechanisms of aging, we cannot study those processes in isolation. Factors that affect the SOH of batteries can be divided into internal and external factors. Internal factors include battery materials and the growth of internal resistance. It has been studied that the dominant aging mechanism on the anode is ascribable to the formation of the solid electrolyte interface (SEI), which significantly increases the impedance. In addition, when the loss of lithium in the activated carbon occurs, it causes self-discharge and capacity degradation. External factors include overcharging, overdischarging, operating temperature, inappropriate charge and discharge cycles, and high charge and discharge rate [10]–[12]. It becomes quite difficult to estimate the SOH of the battery since these unpredictable factors affect the health of the battery. In order to solve this problem and achieve more accurate SOH estimation, researchers have made numerous contributions in recent years. In general, existing methods can be divided into two main categories: model-based methods and data-driven methods [6].

Model-based methods require a lot of prior knowledge about the battery (e.g., the physical and chemical principles of the battery), and then establish physical or mathematical models to estimate SOH. The electrochemical model (EM) [13] and equivalent circuit model (ECM) [14] are the main two types of model-based methods. An EM is used to estimate the SOH by studying the processes of electrochemical reaction inside the battery. This method is difficult to apply in practice since it is too complex to model. Fuller *et al.* [15] proposed a pseudo-two-dimensional (P2D) model which utilizing a first principles based derivation of governing equations for describing the microscopic behavior of Li-ion batteries. Single particle (SP) model [16] derived from P2D model is one of the simplest EMs. This model assumed that each electrode consists of only one spherical solid particle and ignores the diffusion of Li-ion diffusion in the liquid phase. However, under high current conditions, the error of this method is much higher than the P2D model. In order to strike a compromise between model complexity and accuracy Torai *et al.* [17] proposed a model for SOH estimation by exploiting the differential capacity characteristics of the LiFePO4 (LFP) / graphite battery. Their model was based on parameters that are correlated with the phase transition behavior of the active LFP and graphite materials. To further simplify SOH estimation model, model-based methods have a tendency to adopt ECM which uses external characteristics of the battery to describe the dynamic behavior of batteries. Considering practicality, ECMs are more widely used [18], [19]. Wang *et al.* [20] adopted the Thevenin ECM to analyze parameters of battery and employed density-based spatial clustering of applications with noise (DBSCAN) based method to evaluate inconsistency of a battery system. Kim and Cho [21] developed an ECM and used extended Kalman filter (EKF) combined with a per-unit system to predict state of charge (SOC) and SOH of Li-ion batteries. Feng *et al.* [22] added moving average noise to the one resistor-capacity circuit model. Yu *et al.* [23] proposed a quantum particle swarm optimization-based particle filter (PF) method for SOH estimation. Their method requires fewer parameters to control, thus reduced the computational complexity and made this method easier to be applied. To balance the accuracy of the model with computational complexity, Ouyang *et al.* [24] adopted internal resistance ECM to independently learn the dynamics of each battery cell in the battery pack. However, methods based on ECM are susceptible to environmental interference and noise. As data becomes more available and accessible in recent years, data-driven methods can be utilized to estimate SOH. Data-driven methods have recently become popular since they rely only on historical data and avoid establishing complex physical or mathematical models.

Data-driven methods are directly based on the battery's monitorable data (e.g., current, voltage, temperature, etc.) or healthy features (HFs) that can be extracted from charging and discharging processes and do not take into account the battery's failure mechanism and electrochemical reactions. The basic idea of methods of this kind is to establish a mapping between monitorable data (or HFs) and SOH of batteries. Support vector machine (SVM) is a well-known machine learning method that can be used for regression and capturing the characteristics of nonlinear systems [25]. Nuhic *et al.* [26] collected several observable features under various aging conditions and applied SVM as an SOH estimator. Klass *et al.* [27] not only collected some observable features, but also added the SOC calculated by coulomb counting and temperature dependence to the input of the SVM based estimation system. Patil *et al.* [28] collected and analyzed Li-ion batteries data of discharge cycles and extracted several HFs. Then, they estimated the SOH and remaining useful life (RUL) with regression and classification version of SVM respectively. Gaussian process regression (GPR) is also commonly utilized for modeling battery degradation. Yang *et al.* [29] extracted four features from charging curves which can reflect the aging phenomenon of the battery from different perspectives, analyzed the correlation degree of features, and estimated SOH with GPR. Wang *et al.* [30] made use of the incremental capacity (IC) curve when selecting features, and extracted peak value and the corresponding position as the input of GPR to estimate SOH. Zheng and Deng [31] extracted several features from the charging processes, then analyzed the correlation between features and SOH, and selected the three most relevant features. Finally, they proposed multiple GPR model for SOH estimation. In addition, other machine learning methods such

as relevance vector machine (RVM) [32], autoregressive integrated moving average (ARIMA) [33], random forest [34], and XGBoost [35] are also investigated for SOH estimation. Recently, artificial neural network (ANN) methods have been drawing attention because of their powerful ability to learn nonlinear patterns. Bai *et al.* [36] integrated ANN with dual EKF for SOH estimation. SOC, current, and capacity were selected as input features. Wu *et al.* [37] presented the group method of data handling polynomial ANN model for establishing the relationship between properties of battery voltage curves and SOH. She *et al.* [38] applied radial basis function neural network (RBFNN) for learning the relationship between extracted features and battery aging level. Feng *et al.* [39] developed an electrochemical-thermal-neural-network (ETNN) model and integrated it with the unscented Kalman filter (UKF) to estimate battery parameters. Their model can quickly eliminate initial errors. Although classic ANNs are powerful in capturing patterns of nonlinear systems, they are not effective enough for systems related to time series. Particularly for SOH estimation, mapping relationship between features extracted from a single charge or discharge cycle and SOH is not appropriate. It would be better to use data within a specific time interval to estimate SOH instead of using a single time point. Therefore, the method used to estimate SOH must be able to effectively process time series data. A Recurrent neural network (RNN) that can process sequential inputs with its internal state is a natural approach to solve this problem. Chaoui and Ibe-Ekeocha [40] built a dynamically driven RNN for SOH estimation. Nonetheless, a basic RNN model is easily prone to long-term dependency problem, which is caused by the vulnerability of gradients to either vanish or explode during the training phase. Long short-term memory (LSTM) and Gated recurrent unit (GRU), as variants of RNN, were often employed to learn long-term dependency by using gating system. Wu *et al.* [41] extracted several HFs during charging and discharging processes and used LSTM based method for SOH estimation. GRU which has fewer parameters than LSTM can reach or exceed the performance of LSTM in several areas, thus it has also been used to study battery degradation problems [42], [43].

In general, SOH estimation can be considered as a multivariate time series prediction problem and LSTM and GRU naturally become the preferred models. However, attention mechanisms introduced by Bahdanau *et al.* [44] achieved good results on the task of translation. Inspired by this work, researchers have presented various attention mechanisms and performed well in many tasks, such as speech recognition [45], neural machine translation [46], and image captioning [47]. Although these attention based encoder-decoder models somewhat alleviate long-term dependency problem of LSTM and GRU, they can not effectively select the relevant input feature when dealing with multivariate time series prediction. We believe that it is helpful to improve the accuracy of estimation because different features have distinct influences on the final results.

To address these aforementioned problems, we propose a dynamic spatial and temporal attention-based gated recurrent unit (DSTA-GRU) model for SOH estimation, with novel structures to better capture spatial and temporal patterns. Our idea first employs a 1D convolutional layer as embedding for information extraction, then calculates attention for each convolutional kernel and aggregates them using attention weights. Attention weights are calculated dynamically based on input by sharing parameters of the embedding block between input of each time step. This operation, which we called dynamic spatial attention, takes into account spatial correlations. Then we apply GRU for capturing dependencies for time steps and assign attention weight for GRU output at each time step based on its correlation with the last output. This temporal attention idea is based on our belief that the last output of GRU should contribute the most to the final result, thus we present this method instead of commonly used encoder-decoder attention methods. In this way, our proposed model combines dynamic spatial attention which can selectively emphasize informative convolutional kernels and suppress less helpful ones, and temporal attention which can learn the relative importance of hidden states of each time step. We use this model to estimate SOH of Li-ion batteries and extract six health features from charging and discharging processes. Furthermore, Adam optimizer is applied to optimize the weights and bias of the model. Experiments and analysis on NASA datasets show that the proposed model can achieve higher accuracy than other baseline methods.

The rest of this paper is organized as follows. Firstly, Section II presents the definition of SOH and battery datasets. The HFs extraction and dynamic spatial and temporal attention model are elaborated in Section III. Section IV displays and discusses the SOH estimation results. Finally, conclusions are drawn in Section V.

II. PRELIMINARIES

In this section, we first introduce the definition of SOH to indicate the problem we are addressing more clearly and then describe the source and experimental conditions of datasets we selected.

A. DEFINITION OF SOH

Since the SOH of the battery is an important indicator in a BMS, it must be clearly and concisely defined. The purpose of the SOH is to demonstrate the degree of battery deterioration compared with new batteries. Therefore, the SOH is defined as the ratio of the current maximum available capacity to the initial value and expressed as:

$$SOH = \frac{C_t}{C_0} \tag{1}$$

where C_t represents the capacity of t-th cycle and C_0 represents the initial capacity. The SOH curve shows a trend of overall degradation as the battery ages due to continuous charging and discharging.

TABLE 1. Experimental conditions of batteries.

Battery	Charging			Discharging	
	Charging current (A)	Voltage upper limit (V)	Cut-off current (mA)	Discharging current (A)	Cut-off voltage (V)
B0005	1.5	4.2	20	2	2.7
B0006	1.5	4.2	20	2	2.5
B0007	1.5	4.2	20	2	2.2

B. DATASETS

In this work, data of Li-ion batteries obtained from NASA Ames Prognostics Center of Excellence [48] are used for validating our methods. We choose three cycle datasets of 18650 Li-ion rechargeable batteries named B0005, B0006, and B0007. Each dataset records information about the battery that runs through three different operations (e.g. charging, discharging, and impedance) at room temperature (24°C). The charging and discharging processes of these three batteries are as follows. A complete charging process has two phases: the first phase is the constant current (CC) phase, during which the Li-ion battery is charged with a current of 1.5A until the voltage reaches 4.2V; the next phase is the constant voltage (CV) phase, in which continues to charge the battery in CV mode until the current drops to 20mA. During the discharging process, the Li-ion battery is discharged with a constant current of 2A until the voltage reaches the discharge cut-off voltage. The specific experimental conditions and cut-off voltage of these batteries are enumerated in Table 1, and the capacity aging curves are shown in Figure 1.

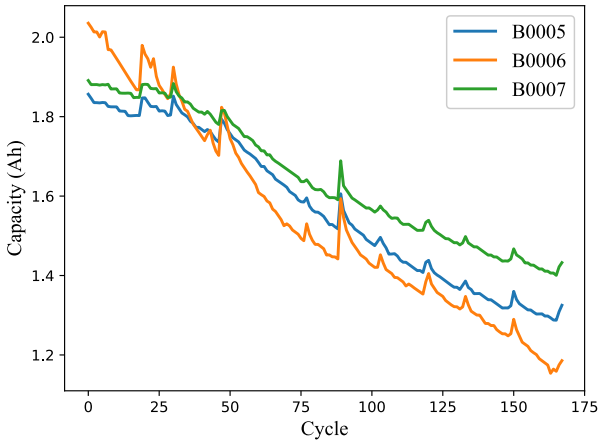


FIGURE 1. Battery capacity decay curves as functions of the number of cycles.

III. METHODOLOGY

Data-driven methods, as we've discussed in Section I, are to create a mapping between cyclic observable data (or HFs) and SOH. Therefore, we first introduce HFs we extracted to learn the mechanisms of battery degradation. Then, we present details of dynamic spatial and temporal attention based model for estimating SOH.

A. EXTRACTION OF HEALTHY FEATURES

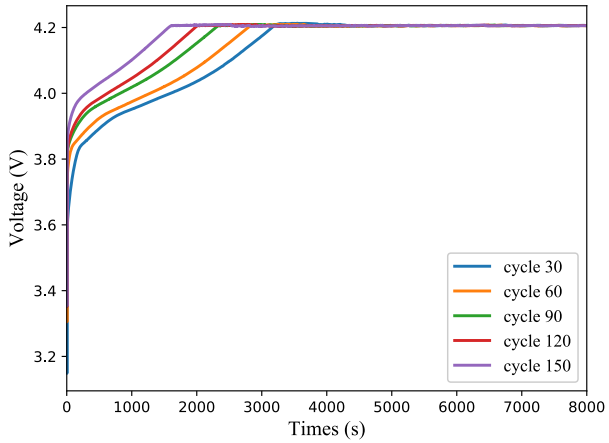
Several studies directly employ observable data such as voltage, current, and temperature or capacity for SOH estimation, but the trouble is that these methods require a sampling method to reduce the size of data since the size of the observed voltage, current and temperature in a charge cycle or a discharge cycle can be very large, and data such as capacity is difficult to measure in real-world applications. In addition, although the IC curve has recently been used frequently to estimate SOH, the noise of the original IC curve can lead to the extraction of inappropriate implied features. We believe that the IC curve needs to be de-noised in the form of filtering to achieve a smooth one, which does not meet our purpose of making the process of extracting features as simple as possible. Despite the fact that the selected feature is not directly related to SOH, it must be able to indirectly reflect the aging degree of the battery. In this paper, six HFs extracted from the charging and discharging processes of batteries are reproduced below.

1) F1: THE DURATION OF CC PHASE

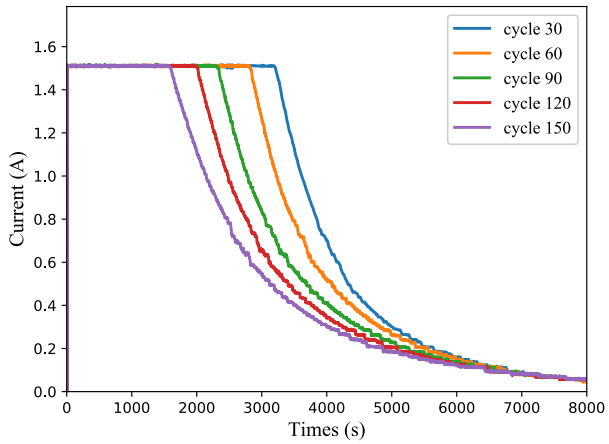
As mentioned in Section II-B, a complete charging process is divided into CC and CV stages. The CC and CV charging processes are suitable for analyzing battery aging problem since they are relatively stable. Figure 2(a) shows the battery voltage curve during the charging processes at different charging cycles. It is evident from Figure 2(a) that the end time of the CC phase decreases as the number of charging cycles increases. The capacity of the battery primarily depends on the duration of the CC phase. This can manifest the battery polarization phenomenon, that is, as the battery polarization becomes more serious, the end time of the CC phase is gradually reduced. Therefore, we define the end time of the CC phase as the first feature $F1$. In addition, we find that the CC duration of the first cycle in the three battery datasets is much shorter than the next few cycles, so the charge and discharge data of the first cycle in these three battery datasets are excluded.

2) F2: THE RATIO OF CC PHASE

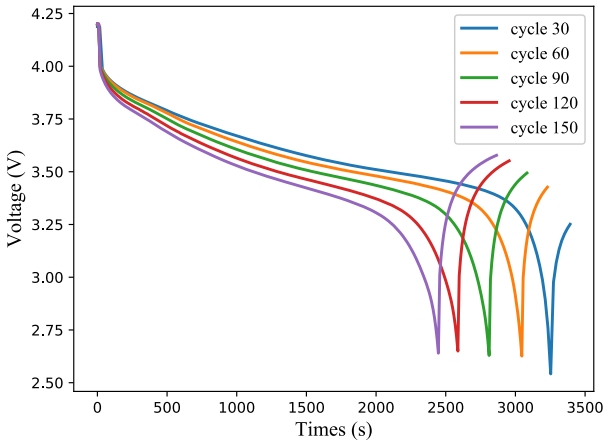
In addition, as the number of cycles increases, the duration of the CV phase increases, and the complete charging time gradually decreases. However, the change of the duration of the CV phase is less pronounced than the change of the CC phase. Therefore, compared to directly use the duration of



(a)



(b)



(c)

FIGURE 2. Charge and discharge profiles of B0005 with different cycles: (a) charging voltage curves; (b) charging current curves; (c) discharging voltage curves.

the CV phase, the ratio of the CC phase to the complete charging process can be more suitable for reflecting the degree of battery aging. It is selected as the second HF:

$$F2 = \frac{t_{CC}}{t_{CC} + t_{CV}} \tag{2}$$

where t_{CC} and t_{CV} represent the duration of the CC and CV phase in each cycle, respectively.

3) F3: THE AREA UNDER THE CC PHASE

It is also obvious from Figure 2(a) that the voltage area at the CC stage decreases as the battery ages. The third HF is expressed as:

$$F3 = \int_0^{t_{CC}} v(t)dt \tag{3}$$

where $v(t)$ denotes voltage of battery at time t during the CC phase.

4) F4: THE TIME REQUIRED FOR EQUAL VOLTAGE RISE INTERVAL

As the battery decays, the CC phase gradually decreases resulting in a difference in the time required for the same voltage change in the charging process. Therefore, we define the time it takes for the voltage rise from 2.7V to 4.2V as the fourth HF:

$$F4 = t_{v_2} - t_{v_1} \tag{4}$$

where t_{v_2} and t_{v_1} are time points corresponding to 2.7V and 4.2V respectively.

5) F5: THE TIME REQUIRED FOR EQUAL CURRENT DROP INTERVAL

The CV phase is also helpful for us to analyze battery degradation. Similar to $F4$ and can be seen in Figure 2(b) the time required for equal current drop interval during the CV phase also varies with battery ages. The time required from 1.5A drops to 0.3A is defined as the fifth HF:

$$F5 = t_{c_2} - t_{c_1} \tag{5}$$

where t_{c_1} and t_{c_2} denote the time corresponding to the 1.5A and 0.3A, respectively.

6) F6: THE TIME REQUIRED FOR EQUAL VOLTAGE DROP INTERVAL

We not only considered the charging process, but also the impact of the discharging process with the degradation of the battery. The voltage curves of the discharge process under several different cycles are shown in the Figure2(c). As can be seen from this figure, the discharge voltage drops faster as the number of cycles increases. We, therefore, picked the time it takes to drop from 3.7V to 2.7V as the sixth HF:

$$F6 = t_{v_4} - t_{v_3} \tag{6}$$

where t_{v_3} and t_{v_4} denote the instant at the voltages of 3.7V and 2.7V in the discharging process, respectively.

In this study, the relationships between SOH and six extracted HFs are evaluated by Pearson and Spearman's rank

correlation coefficient. Pearson correlation, which measures the linear relationship between two variables, is calculated directly from the raw data. Spearman's rank correlation is calculated by using the rankings of the two variables and it assesses the monotonic relationship between the two variables. They are denoted by:

$$Pearson = \frac{E(ab) - E(a)E(b)}{\sqrt{E(a^2) - E^2(a)}\sqrt{E(b^2) - E^2(b)}} \quad (7)$$

$$Spearman = \frac{\sum_i (s_{a_i} - \bar{s_a})(s_{b_i} - \bar{s_b})}{\sqrt{\sum_i (s_{a_i} - \bar{s_a})^2}\sqrt{\sum_i (s_{b_i} - \bar{s_b})^2}} \quad (8)$$

where a and b represent the SOH and the HF, respectively. s_{a_i} and s_{b_i} denote the ranks of a_i and b_i, respectively.

The absolute value of the Pearson correlation coefficient to 1 indicates a perfect linear relationship between the two variables, while the coefficient close to 0 means that two variables have no linear relationship. Similarly, the absolute value of the Spearman's rank correlation coefficient is close to 1, indicating that two variables have a strong monotonic correlation. As shown in Table 2 and 3, the absolute of the Pearson correlation coefficients of the six extracted features and SOH on the three data sets are all greater than 0.96, and the absolute of Spearman correlation coefficients are all greater than 0.93, indicating that the extracted features have a strong linear and monotonic relationship with SOH.

TABLE 2. Pearson correlation coefficient between SOH and HFs.

Battery	F1	F2	F3	F4	F5	F6
B0005	0.9975	0.9906	0.9977	0.9975	-0.9860	0.9983
B0006	0.9934	0.9898	0.9934	0.9934	-0.9610	0.9937
B0007	0.9933	0.9846	0.9926	0.9933	-0.9669	0.9982

TABLE 3. Spearman's rank correlation coefficient between SOH and HFs.

Battery	F1	F2	F3	F4	F5	F6
B0005	0.9931	0.9821	0.9931	0.9931	-0.9814	0.9980
B0006	0.9949	0.9920	0.9950	0.9949	-0.9398	0.9991
B0007	0.9906	0.9743	0.9901	0.9906	-0.9523	0.9972

B. PROPOSED DYNAMIC SPATIAL-TEMPORAL ATTENTION-BASED GRU MODEL FOR SOH ESTIMATION

The main contribution of our research is developing a model for estimating SOH with dynamic spatial and temporal attention mechanisms. The architecture of the proposed model is presented in Figure 3. First, a convolutional layer is applied to learn the internal representation of data. After that, we propose a novel dynamic spatial attention mechanism to learn the spatial correlations. Then, we process the embedded input with a GRU layer for identifying long term dependencies. Finally, the estimation result is made by temporal attention mechanism over outputs of the GRU layer.

For battery data, we employ $\mathbf{x}_t = [x_t^1, x_t^2, \ldots, x_t^n] \in \mathbb{R}^n$ to denote a vector of n exogenous (HFs) at cycle t. Our model aims to learn a mapping between historical $[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T]$ and target value y_T, where T is the length of sliding window size.

1) DYNAMIC SAPTIAL ATTENTION

Convolutional neural networks (CNNs) have achieved good results in many computer vision tasks because of their ability to learn spatial patterns [49], [50]. By using a series of convolutional layers, CNN models can capture only the attributes of images that are most salient for the target, thereby improving performance. Similarly in natural language processing (NLP) tasks, the 1D convolutional layer is often used as an embedding layer. The 1D convolutional layer is expected to learn the locality when kernels slide across input features. Given the vector of features at each step t, the first step of our model is to generate an embedding function for capturing the dependencies between different variables. To this end, we adopt a 1D convolutional layer to obtain the embedding of the information at each step. A 1D convolutional layer is a computational operation that can map $\mathbf{x}_t$ of our input data to a feature map $\mathbf{U}_t \in \mathbb{R}^{n' \times K}$. Let us denote the set of filter kernels as $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_K]$, where $\mathbf{v}_k$ is the k-th filter. Then, the output $\mathbf{U}_t = [\mathbf{u}_t^1, \mathbf{u}_t^2, \ldots, \mathbf{u}_t^K]$ can be calculated with

$$\mathbf{u}_t^k = \mathbf{v}_k \circledast \mathbf{x}_t + b_k \quad (9)$$

where $\circledast$ denotes convolution operation and $b_k \in \mathbb{R}$ is bias. $\mathbf{u}_t^k \in \mathbb{R}^{n'}$ is the k-th channel of $\mathbf{U}_t$ and can give the responses of the filter $\mathbf{v}_k$ at every space position. Intuitively, learned filters can activate when capturing some type of patterns. Given an input, it is common for a convolutional layer learns with multiple filters in parallel. This gives the model multiple ways to extract features of interest from input or to look at input from multiple perspectives. Therefore, it is necessary to find hyperparameters such as kernel size and number of filters for better results. In computer vision, CNN models capture hierarchical patterns by using a series of convolutional layers and downsampling operations, resulting in making models deeper and computationally more expensive for better performance. In this task, the input data are not as big as image data, so the model should be not too large. Moreover, if only one convolutional layer is applied and directly flatten the feature map as the input of the subsequent layer (the layer that handles temporal relationship), the size is too large that the next layer is difficult to handle. Therefore, we expect the embedding block to have a powerful representation without using too many convolutional layers and provide informative features for the next layer.

Inspired by Squeeze-and-Excitation (SE) networks [51] and dynamic convolution [52], we proposed a novel attention method for our embedding. Different from SE networks which attention calculated based on channel information and dynamic convolution that attention computed based on average pooling results of input data, we calculate attention

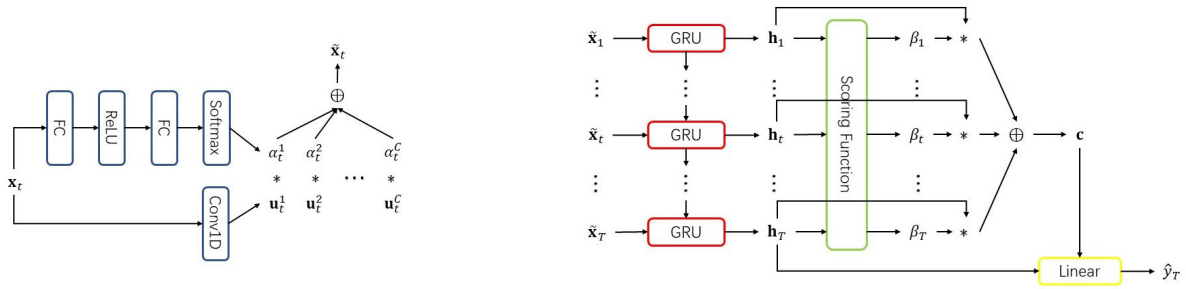


FIGURE 3. The overview of our proposed DSTA-GRU model: (a) dynamic space attention mechanism: first, a 1D convolutional layer is used to embed input data of time step t, then using two FC layers with ReLU and sotfmax functions to generate attention weight α_t^c for each channel of feature map $\mathbf{U}_t$, finally, aggregate feature map with attention weights to generate adjusted input $\tilde{\mathbf{x}}_t$. (b) temporal attention mechanism: first employ GRU to learn long-term dependency and then computes attention weight β_t for hidden states $\mathbf{h}_t$ based on scoring and softmax functions, then the final estimation result is made by context $\mathbf{c}$ and the last hidden state $\mathbf{h}_T$.

directly based on input data. Our main idea is to model channel interdependencies to enhance the representation power of feature map that is the output of the convolution layer, and aggregate all the channels of feature map by using attention weights as shown in Figure 3(a). Attention is calculated to capture channel-wise dependencies by using input data. Therefore, we use the two fully connection (FC) layers with ReLU and softmax function to generate the attention weight for each channel of the feature map. The formulations are as follows:

$$\mathbf{e}_t = \mathbf{W}_2 \mathrm{ReLU}(\mathbf{W}_1 \mathbf{x}_t) \tag{10}$$

$$\alpha_t^k = \frac{\exp(e_t^k)}{\sum_{i=1}^{K} \exp(e_t^i)} \tag{11}$$

where $\mathbf{W}_1 \in \mathbb{R}^{n'' \times n}$ and $\mathbf{W}_2 \in \mathbb{R}^{K \times n''}$ are trainable parameters of the first and second FC layers, respectively, and $\mathbf{e}_t = [e_t^1, e_t^2, \ldots, e_t^K] \in \mathbb{R}^K$ is the output of the second FC layer. In order to limit the complexity of the embedding part, two FC layers are applied and ReLU function is to make this operation non-linear. The output size of the first FC layer is designed as n'' to form a transition between the input feature n and the number of channels K. The final output of this embedding block is obtained by aggregating $\mathbf{U}_t$ with attention weights:

$$\tilde{\mathbf{x}}_t = \sum_{k=1}^{K} \alpha_t^k \mathbf{u}_t^k \tag{12}$$

This block only considers spatial patterns and attention weights are generated dynamically by sharing parameters of this block between inputs at different time steps.

2) TEMPORAL ATTENTION

Recently, a multitude of disparate temporal attention mechanisms have been proposed by researchers and utilized in multivariate time series tasks [53]–[55]. They have one thing in common: they are based on an encoder-decoder structure. Therefore, these temporal attention mechanisms are used in the decoder and to adaptively select relevant hidden states of encoder across all steps. These methods require at least two RNN layers, one for encoding and the other for decoding. Although these methods perform well in some data sets, the data set of our task is relatively small, so the model for this task should be not too complicated. Hence, different from these methods, we propose a novel temporal attention mechanism, which only use one RNN layer and compute attention weights between the hidden states of this RNN layer. Intuitively, for the target estimation at the current time, we believe that the hidden state of the RNN at the current time step has the most effect on the final prediction result. Consequently, we generate attention weight for each state by exploiting the correlation between it and the last state. This approach can not only reduce the complexity of the model and make it easier to train the model but also can represent information relevant to the current state.

For learning temporal relationship, given the input adjusted by our spatial attention $[\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_T]$ with $\tilde{\mathbf{x}}_t \in \mathbb{R}^{n'}$, we first employ GRU based RNN for capturing long-term dependencies since it not only can solve gradients exploding and vanishing problem but also has simpler structure than other gated RNNs. A GRU cell consists of an update gate and a reset gate that can decide how much previous information should be ignored and new information should be added to the state. The operations of a GRU cell can be formulated as:

$$\mathbf{z}_t = \sigma(\mathbf{W}_z[\mathbf{h}_{t-1}; \tilde{\mathbf{x}}_t] + \mathbf{b}_z) \tag{13}$$

$$\mathbf{r}_t = \sigma(\mathbf{W}_r[\mathbf{h}_{t-1}; \tilde{\mathbf{x}}_t] + \mathbf{b}_r) \tag{14}$$

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_h[\mathbf{r}_t \odot \mathbf{h}_{t-1}; \tilde{\mathbf{x}}_t] + \mathbf{b}_h) \tag{15}$$

$$\mathbf{h}_t = \mathbf{z}_t \odot \mathbf{h}_{t-1} + (1 - \mathbf{z}_t) \odot \tilde{\mathbf{h}}_t \tag{16}$$

where $\mathbf{z}_t \in \mathbb{R}^m$ and $\mathbf{r}_t \in \mathbb{R}^m$ denote update gate and reset gate, respectively. $\mathbf{W}_z, \mathbf{W}_r, \mathbf{W}_h \in \mathbb{R}^{m \times (m+n')}$ and $\mathbf{b}_z, \mathbf{b}_r, \mathbf{b}_h \in \mathbb{R}^m$ are trainable parameters. σ and $\odot$ are sigmoid function and element wise multiplication, respectively. $\mathbf{h}_t \in \mathbb{R}^m$ and $\tilde{\mathbf{x}}_t \in \mathbb{R}^{n'}$ are hidden state and adjusted input at time step t, respectively.

Then, after GRU, a temporal attention as shown in Figure 3(b) is used to adaptively select the relevant hidden states across all time steps. First, a scoring function is applied to evaluate relevance between each state and the last state. The score of hidden state at time step t can be calculated as:

$$score(\mathbf{h}_t, \mathbf{h}_T) = \mathbf{h}_t^T \mathbf{W}_s \mathbf{h}_T, \quad 1 \le t \le T \tag{17}$$

where $\mathbf{W}_s \in \mathbb{R}^{m \times m}$ is trainable parameter. Then, the attention weight of $\mathbf{h}_t$ can be obtained by using Softmax function. The attention weight can be calculated as:

$$\beta_t = \frac{\exp(score(\mathbf{h}_t, \mathbf{h}_T))}{\sum_{j=1}^{T} \exp(score(\mathbf{h}_j, \mathbf{h}_T))} \tag{18}$$

The attention weight β_t indicates the importance of the t-th hidden state for the SOH estimation. Next, we aggregate hidden states with attention weights to form the context vector:

$$\mathbf{c} = \sum_{t=1}^{T} \beta_t \mathbf{h}_t \tag{19}$$

Finally, the estimation output $\hat{y}_T$ can be calculated by using a linear function with concatenation result of the context vector and the last hidden state:

$$\hat{y}_T = \mathbf{W}_o^T [\mathbf{c}; \mathbf{h}_T] + b_o \tag{20}$$

where $[\mathbf{c}; \mathbf{h}_t] \in \mathbb{R}^{2m}$ is a concatenation of context vector $\mathbf{c}$ and the last hidden state $\mathbf{h}_T$. $\mathbf{W}_o \in \mathbb{R}^{2m}$ and $b_o \in \mathbb{R}$ are parameters to learn.

IV. EXPERIMENTAL RESULTS

In order to verify the proposed DSTA-GRU model, real world Li-ion batteries cycle data collected from NASA are applied for performing experiments. Details of experimental data have been described in Section II-B. We first compare the estimation performance of our model with different hyperparameters, aiming to find the most appropriate hyperparameter setting for our proposed model. Then we verify the impact of the components of our model on the results of SOH estimation. After that, we carry out SOH estimation on the three batteries respectively and compare to other baseline models. Finally, we study the robustness of DSTA-GRU model and compare the performance of our model with different HFs as input. The mean absolute error (MAE) and root mean squared error (RMSE) are used as criteria for performance evaluation. The calculation formulae of MAE and RMSE are as:

$$\text{MAE} = \frac{1}{N} \sum_{t=1}^{N} |y_t - \hat{y}_t| \tag{21}$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2} \tag{22}$$

where y_t and $\hat{y}_t$ are real value and estimated value of SOH at cycle t, respectively.

A. HYPERPARAMETERS ANALYSIS

For data-driven methods, the performance of a model depends to a large extent on the setting of hyperparameters. The setting of hyperparameters is defined prior to the training process. Here we employ B0005 dataset and use 5-fold cross-validation (CV) to analyze hyperparameters. We choose Pytorch library written in Python to implement our model and train the model on a single Nvidia GTX 1080 Ti GPU. Adam with an initial learning rate 0.01 is utilized as optimizer to train our model. The number of maximum iteration is set as 3000 and the learning rate is reduced by 10% if the loss value of validation set does not drop in 100 iterations. The hyperparameters that our model needs to examine include the number of filters K of the convolutional layer, sliding window T, and the hidden state size m of the GRU layer. For our model, we set K varying among [64, 128, 256], T varying among [5, 10, 15], and m varying among [64, 128]. Other hyperparameters such as the stride, zero-padding size, and kernel size of the convolutional layer are set to 1, 1, and 3, respectively. The output size of the first FC layer in the embedding block and the batch size are set to 16 and 32, respectively. The average RMSE and MAE results of 5-fold CV for different hyperparameter combinations are summarized in Table 4. First, let us focus on T. It can be seen from Table 4 that when $T = 5$ or $T = 15$, MAE results are all greater than 0.0025 and RMSE values are all greater than 0.0035, but when $T = 10$, there are several settings of hyperparameters achieve MAE values less than or equal to 0.0025 and RMSE results less than 0.0035. Therefore, for our SOH estimation model, $T = 10$ is the best choice. Then

TABLE 4. 5-fold CV performances on B0005 dataset with different settings of hyperparameters.

Hyperparameters			Criteria	
T	m	K	average MAE	average RMSE
5	64	64	0.0027	0.0036
		128	0.0026	0.0037
		256	0.0027	0.0035
	128	64	0.0038	0.0052
		128	0.0037	0.0054
		256	0.0040	0.0055
10	64	64	0.0025	0.0034
		128	0.0023	0.0031
		256	0.0025	0.0034
	128	64	0.0035	0.0050
		128	0.0040	0.0053
		256	0.0043	0.0060
15	64	64	0.0028	0.0039
		128	0.0034	0.0044
		256	0.0028	0.0038
	128	64	0.0038	0.0055
		128	0.0031	0.0048
		256	0.0034	0.0047

we look for the most appropriate m. After determining that T is set to 10, it is clear from Figure 4 that when $m = 64$, the performances of our model are significantly better than $m = 128$. We believe that the performance decline when $m = 128$ is due to overfitting caused by too many hidden neurons in the GRU cell. In addition, it can be seen from this figure that the value of m has a greater impact on the result than the value of K. Finally, when $T = 10$ and $m = 64$, it is clear that $K = 128$ is the best choice. Therefore, the setting of hyperparameters for our model is finally determined as $\{T = 10, m = 64, K = 128\}$. Moreover, since the size of the B0006 and B0007 datasets and the experimental conditions are relatively close to B0005, we also use the above hyperparameters combination when training these two datasets.

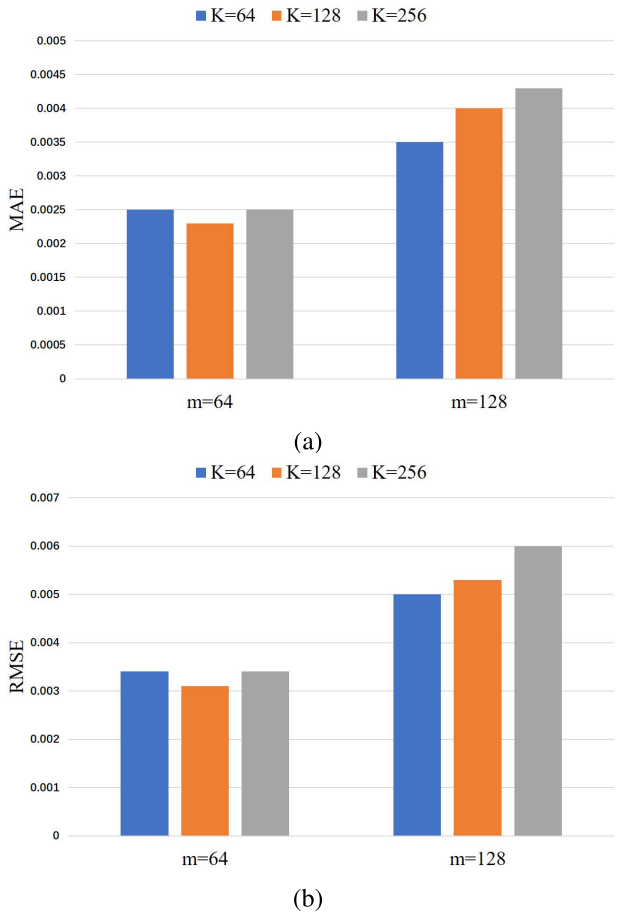


FIGURE 4. Criteria results of different settings of hyperparameters when $T = 10$ on B0005 dataset: (a) MAE; (b) RMSE.

B. COMPONENTS ANALYSIS

In order to examine the effectiveness of dynamic spatial attention and temporal attention mechanism within our model we train and test several variants of our model on B0005, B0006, B0007 datasets. For each dataset 50% of the data is used as training set and the rest of the data is used as test set. In addition, 20% of the data in training set is randomly selected as validation set during each training. To demonstrate the effectiveness of each attention mechanism, we first consider two variants of our model: Dynamic Spatial Attention GRU (DSA-GRU) is a variant of our DSTA-GRU model with temporal attention removed and only GRU and dynamic spatial attention retained; Temporal Attention GRU (TA-GRU) is DSTA-GRU with dynamic spatial attention removed. Then, in order to verify the validity of GRU in our model, we consider two additional variants: DSTA-RNN and DSTA-LSTM that are variants with replaced GRU in our model with the standard RNN and LSTM, respectively. For the sake of fair comparison, the hyperparameters of each model are set as follows. For the GRU, DSA-GRU, and TA-GRU models, we set $\{T = 10, m = 64\}$, $\{T = 10, m = 64, K = 128\}$, and $\{T = 10, m = 64\}$, respectively, while for other models, the hybperparameters settings refer to the previous experiment. The best performance of each model was obtained through multiple experiments and the results are tabulated in Table 5 and depicted in Figure 5. From the table and figure, it can be seen that DSA-GRU outperforms the GRU model on all three datasets. For example, MAE and RMSE are reduced by 35.3% and 31.7%, respectively, on the B0005 dataset. This can illustrate the effectiveness of dynamic spatial attention. Similarly, TA-GRU performs better than GRU on all three data sets. In particular, we can observe from Figure 5 that TA-GRU not only performs better than GRU but also has a slightly higher improvement effect than dynamic spatial attention. Furthermore, our DSTA-GRU combines the two attention mechanisms leading to further improvements in performance on the three datasets. Then, we consider DSTA with different RNN cells. As can be seen, DSTA-GRU outperforms DSTA-RNN and DSTA-LSTM in the terms of MAE and RMSE. The DSTA-RNN is the worst performer of the three, which can be attributed to the RNN's lack of a gating system and its simple cell structure. The performance of DSTA-LSTM is slightly worse than DSTA-GRU since the gating system of GRU is more concise and efficient for many small datasets and less complex tasks. Finally, compared with the GRU model in B0005, B0006, and B0007, the RMSE results of the DSTA-GRU model proposed by us are reduced by 45.1%, 53.1%, and 52.3%, respectively, and the MAE results are reduced by 54.4%, 61.3%, and 61.4%, respectively. This experiment can indicate that the accuracy

TABLE 5. Performances on B0005, B0006, and B0007 datasets with different variants of our proposed model.

Models	B0005		B0006		B0007	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
GRU	0.0068	0.0082	0.0080	0.0098	0.0070	0.0084
DSA-GRU	0.0044	0.0056	0.0045	0.0070	0.0045	0.0054
TA-GRU	0.0043	0.0054	0.0058	0.0069	0.0041	0.0051
DSTA-RNN	0.0049	0.0066	0.0049	0.0081	0.0050	0.0067
DSTA-LSTM	0.0039	0.0058	0.0037	0.0069	0.0037	0.0050
DSTA-GRU	0.0031	0.0045	0.0031	0.0046	0.0027	0.0040

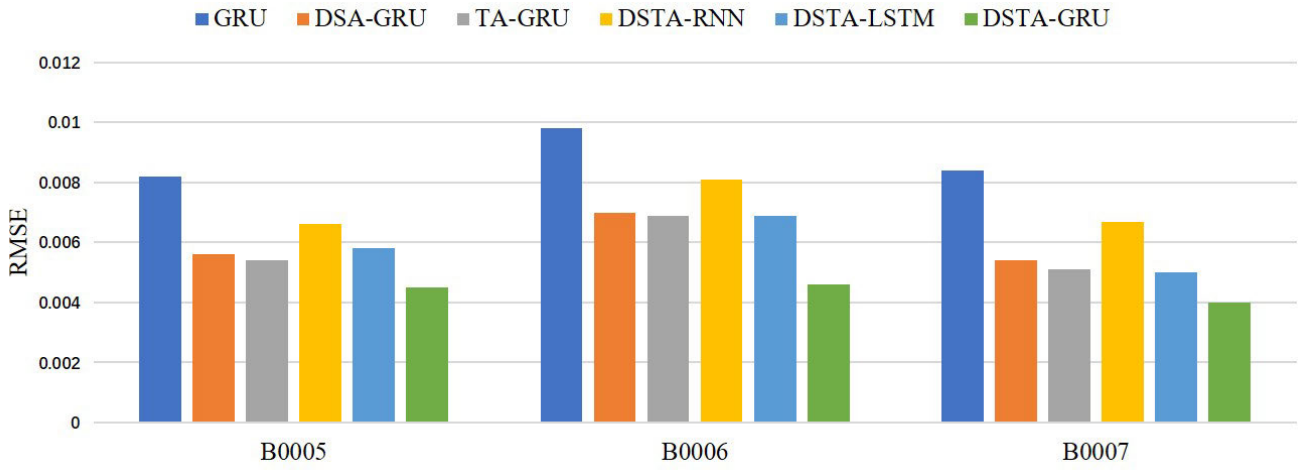


FIGURE 5. RMSE comparison with different variants of our model on B0005, B0006, and B0007 datasets.

TABLE 6. Performances of different models.

Models	B0005				B0006				B0007			
	50%		70%		50%		70%		50%		70%	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
SVM	0.0298	0.0316	0.0144	0.0150	0.0086	0.0115	0.0133	0.0165	0.0092	0.0102	0.0045	0.0052
GPR	0.0192	0.0205	0.0076	0.0084	0.0104	0.0119	0.0042	0.0058	0.0062	0.0069	0.0037	0.0049
RNN	0.0293	0.0389	0.0070	0.0090	0.0091	0.0165	0.0071	0.0091	0.0114	0.0174	0.0054	0.0069
LSTM	0.0069	0.0096	0.0028	0.0044	0.0073	0.0081	0.0046	0.0052	0.0090	0.0105	0.0049	0.0055
GRU	0.0068	0.0082	0.0029	0.0042	0.0080	0.0098	0.0050	0.0056	0.0070	0.0084	0.0037	0.0049
DSTA-GRU	0.0031	0.0045	0.0024	0.0035	0.0031	0.0046	0.0024	0.0033	0.0027	0.0040	0.0023	0.0032

of GRU can be improved by adding our proposed dynamic spatial attention and temporal attention in the SOH estimation task.

C. PERFORMANCE COMPARISON WITH BASELINE METHODS

In this section, we discuss the performance of DSTA-GRU by comparing it with other conventional methods. In this experiment, machine learning algorithms SVM and GPR, which are often used in regression problems, and RNN, LSTM, and GRU, which are often used to process time series data in the field of deep learning, are selected as baseline models. For RNN, LSTM and GRU, the hyperparameter T and m are determined by conducting grid search over $\{5, 10, 15\}$ and $\{64, 128\}$ respectively. These three models all get the best results when $T = 10$ and $m = 64$. In addition, 50% and 70% of cycles data are employed to train the models, and the remaining data is used for testing respectively. For SVM and GPR we chose the Radial Basis Function (RBF) kernel. Besides, as in the previous experiment, 20% of training data are randomly selected as validation set. The MAE and RMSE values of each model on B0005, B0006, and B0007 datasets and different start cycles are tabulated in Table 6 and the estimation results and absolute error are depicted in Figure 6, Figure 7, and Figure 8, respectively. First we discuss the estimation results on the B0005 dataset. As can be seen from Figure 6, although all models can estimate a gradual downward trend, SVM, GPR, and RNN models all have obvious errors. In contrast, the estimation results of LSTM, GRU, and the DSTA-GRU model are closer to the real SOH. Among these three models, according to Figure 6(c) and (d), we can find that the DSTA-GRU attains smaller estimation errors. In addition, these methods have improved their performance after obtaining more training data. As shown in Table 6, the best baseline model on B0005 dataset is GRU. LSTM performs slightly worse than GRU. The DSTA-GRU decreases by 54.4%, 17.2% in MAE, and by 45.1%, 16.7% in RMSE compared to GRU method with 50% and 70% training and validation data. Similarly, for the B0006 data set, it can be found from Figure 7 that the performances of SVM, GPR, and RNN models are worse than the remaining three models. In contrast to the results for the B0005 dataset, the best performing baseline model on the B0006 dataset is LSTM. Compared with LSTM on B0006 dataset, the DSTA-GRU

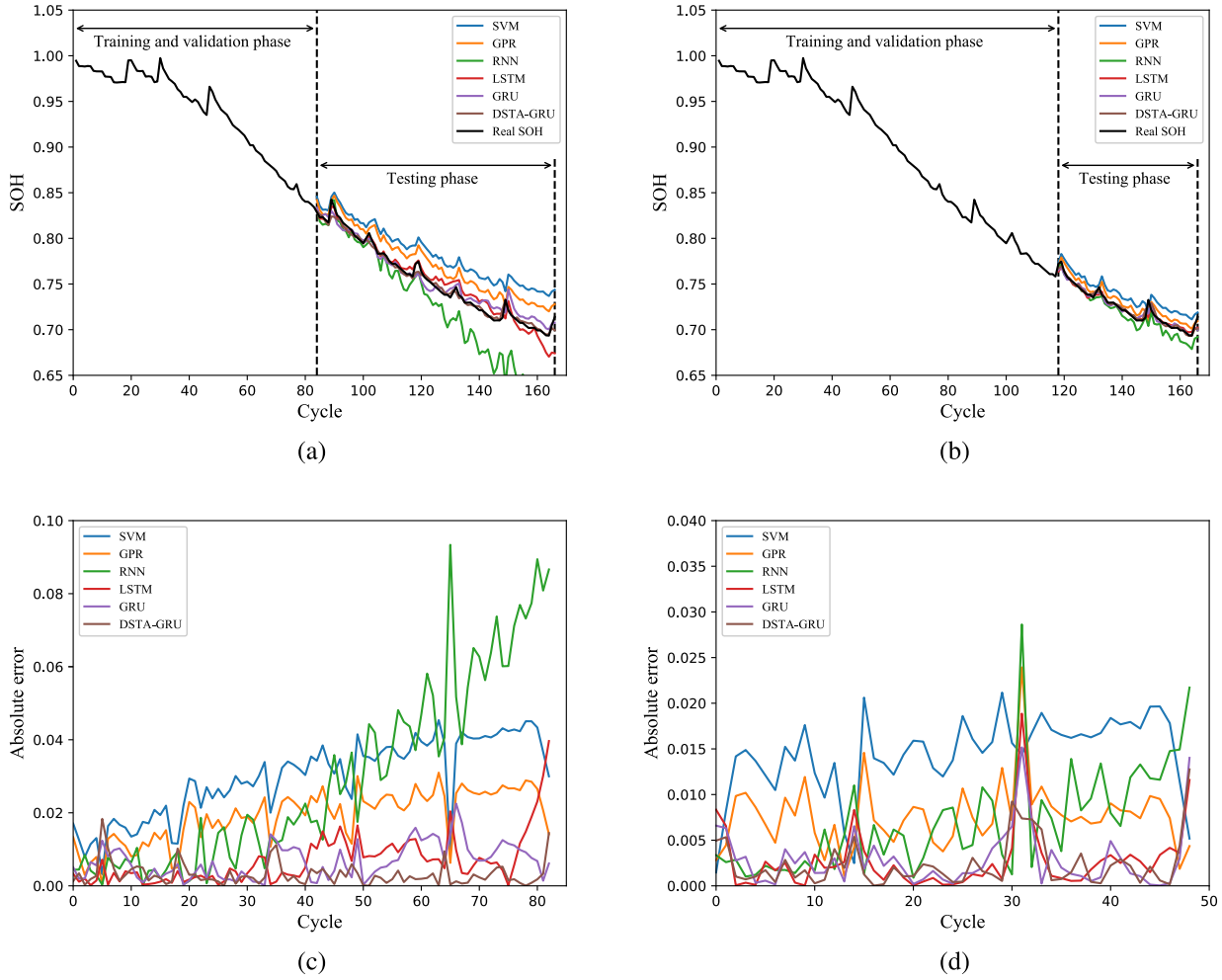


FIGURE 6. SOH estimation of B0005: (a) estimation curves with 50% of the cycle data as training and validation sets; (b) estimation curves with 70% of the cycle data as training and validation sets; (c) absolute errors with 50% of the cycle data as training and validation sets; (d) absolute errors with 70% of the cycle data as training and validation sets.

reduces MAE from 0.0073 to 0.0031 with 50% training and validation cycle data and 0.0046 to 0.0024 with 70% training and validation cycle data. For the B0007 dataset, the GPR performs best among baseline models. The errors plotted in Figure 8(c) and (d) show that the error of DSTA-GRU is more stable than other baseline models. Meanwhile, DSTA-GRU performs better than GPR in both evaluation metrics with MAE of 0.0027 and 0.0023 and RMSE of 0.0040 and 0.0032. A synthesis of all three datasets confirmed that GRU and LSTM perform better than other baseline methods. Although GPR performs better than LSTM in B0007 data sets, GPR performs much worse than LSTM in B0005 and B0006. Based on these analyses, it can be found that estimation results of DSTA-GRU are closer to the real SOH and the performance outperforms all other baseline models on B0005, B0006, and B0007 datasets in terms of MAE and RMSE.

D. VERIFICATION WITH ARTIFICIAL NOISE

When training a deep learning model, using more noise makes it more difficult to fit the data and reducing noise reduces the regularization effect [56]. Since both model fitting and regularization are important for model performance, maintaining an appropriate noise level during the training process is necessary to develop an effective model. The data we use are real-life data, which are inherently noisy, hence we did not carry out noise reduction processing on the features before training the model. To verify the robustness of our model, we add Gaussian white noise to all HFs in the test set. Then, take the noise-added data as input and use the trained model to estimate SOH. In this experiment, we tested with data sets B0005, B0006, and B0007, and 50% of the data from each dataset were used as the test set. Gaussian white noise with standard deviations of 0.01, 0.05, and 0.1 was added in the normalized HFs to examine the impact of noise on the model. For each noise level we tested 10 times and the average RMSE and MAE are tabulated in Table 7. As can be seen from Table 7, when the standard deviation of noise is 0.01, the test results on the three data sets have almost no change, and when the standard deviation of noise is 0.05, the results are comparable with those without artificial noise. Although

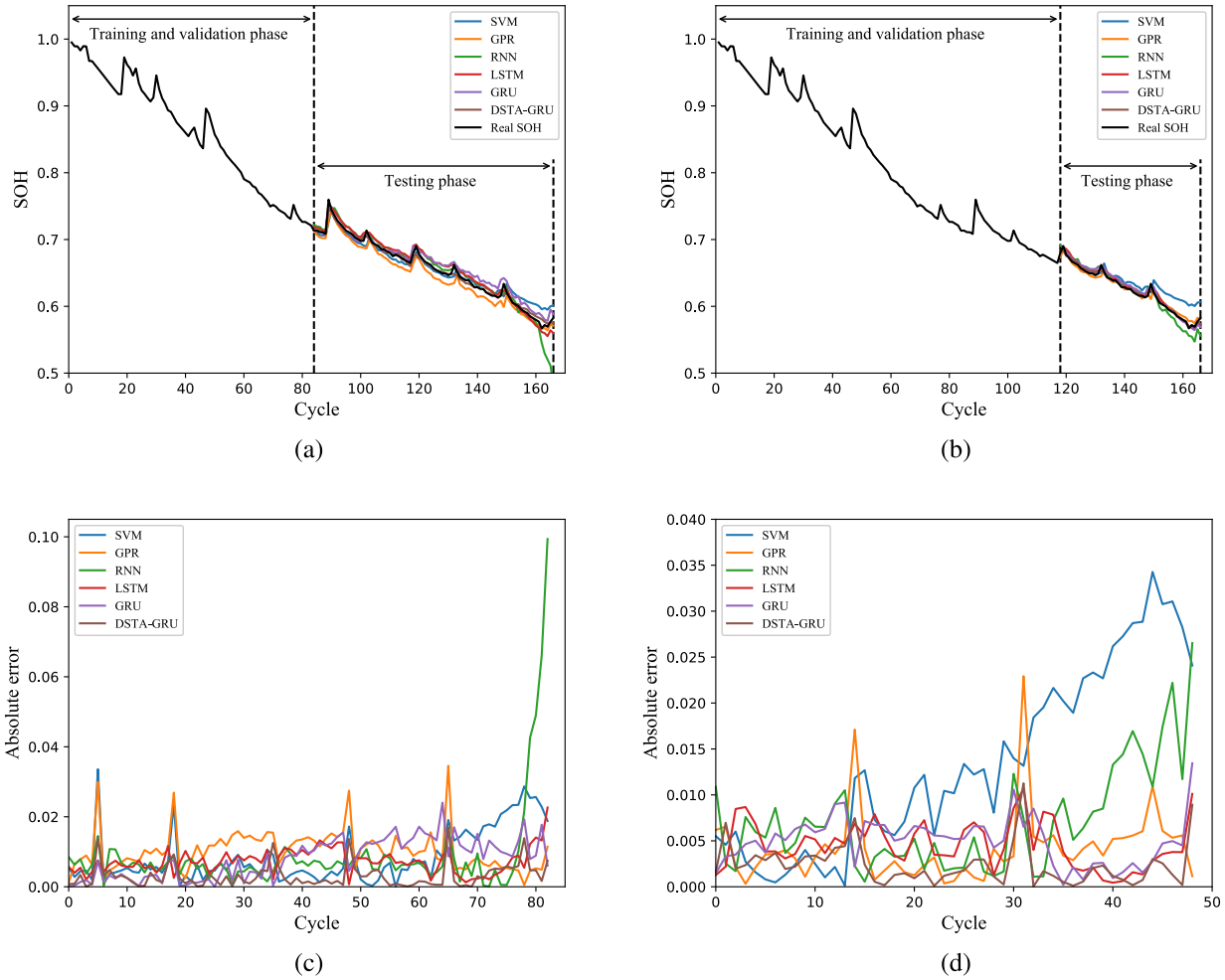


FIGURE 7. SOH estimation of B0006: (a) estimation curves with 50% of the cycle data as training and validation sets; (b) estimation curves with 70% of the cycle data as training and validation sets; (c) absolute errors with 50% of the cycle data as training and validation sets; (d) absolute errors with 70% of the cycle data as training and validation sets.

TABLE 7. Performances of DSTA-GRU with artificial noises.

Noise standard deviation	B0005		B0006		B0007	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
0	0.0031	0.0045	0.0031	0.0046	0.0027	0.0040
0.01	0.0033	0.0046	0.0032	0.0047	0.0028	0.0040
0.05	0.0044	0.0057	0.0041	0.0055	0.0030	0.0042
0.1	0.0066	0.0082	0.0060	0.0076	0.0035	0.0047

both RMSE and MAE increase considerably at a standard deviation of 0.1 of the noise, the results are comparable to the best results of the other basic methods without adding artificial noise. These results indicate that DSTA-GRU is robust to noisy HFs.

E. PERFORMANCE COMPARISON WITH DIFFERENT HFs

In this paper, we extract six HFs for SOH estimation. Among these features, F1, F3, and F4 are obtained from the CC stage of the charging process; F5 from the CV stage of the charging process; F2 from the entire charging process; and F6 from the partial discharge process. In summary, extracting all six features requires obtaining data from complete charging cycle and part of the discharging cycle. Although using all these features to estimate SOH can obtain good estimation results, the extraction process has certain limitations. Therefore, in this experiment, we designed two sets of input combinations to verify DSTA-GRU model, with the purpose of simplifying the way of feature extraction i.e. reducing the conditions required for extracting HFs. The first combination

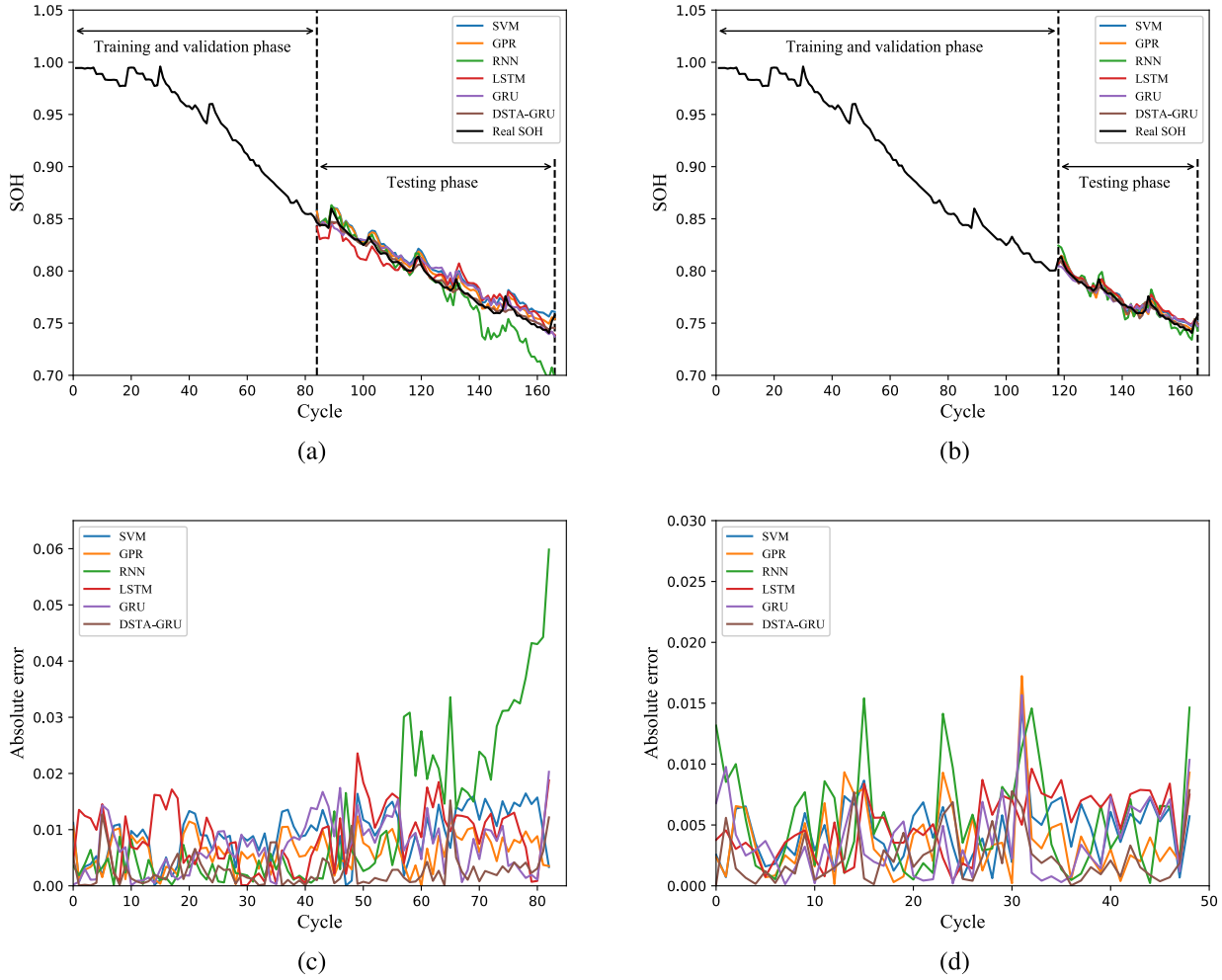


FIGURE 8. SOH estimation of B0007: (a) estimation curves with 50% of the cycle data as training and validation sets; (b) estimation curves with 70% of the cycle data as training and validation sets; (c) absolute errors with 50% of the cycle data as training and validation sets; (d) absolute errors with 70% of the cycle data as training and validation sets.

is $\{F1, F3, F4, F6\}$, which allows us to estimate SOH merely based on the CC phase of the charging process and partial discharge process. The second combination is based on the first one with the removal of $F6$, thus this combination can estimate based on the CC stage of the charging process only. In this experiment, we conducted training and test on data sets B0005, B0006, and B0007 respectively, and 50% of the data was used as the training set and the rest as the test set. For the model, we employed the 5-fold CV approach as in Section IV-A to find the best hyperparameters setting and found that the minimum average RMSE and MAE were achieved at $\{T = 10, m = 64, K = 64\}$ using both input combinations. The training was set up and conducted in the same way as the process in Section IV-B. The results of two input settings are tabulated in Table 8. According to Table 8, it can be seen that using more HFs in our selected features can lead to better estimation results. While using fewer features results in a decrease in accuracy, it requires fewer constraints when extracting features and improves practicality.

TABLE 8. Performances of DSTA-GRU with different HFs.

Input	B0005		B0006		B0007	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
F1,F2,F3,F4,F5,F6	0.0031	0.0045	0.0031	0.0046	0.0027	0.0040
F1,F3,F4,F6	0.0033	0.0053	0.0036	0.0051	0.0030	0.0043
F1,F3,F4	0.0040	0.0056	0.0044	0.0053	0.0027	0.0044

F. FURTHER DISCUSSION

This paper focuses on estimating the SOH of Li-ion batteries using a data-driven approach. In order to make better use of the data-driven approach, we extracted six HFs from the charging and discharging processes and assessed the correlation between HFs and SOH by Pearson and Spearman's rank correlation coefficient. The results show that HFs we extracted are strongly correlated with SOH. The extraction of features is easy to implement since these features can be obtained by observing the voltage and current during charging and discharging and using a timer. Our proposed

DSTA-GRU model, which combines spatial and temporal attention, not only captures local spatial patterns, but also alleviates the long-term dependency problem to some extent, thus resulting in a better performance. This method achieves reliable and robust SOH estimation.

It is worth noting that in practice, the uncertainty of the estimated value is helpful in making an informed decision. However, our estimation approach is point value estimation and does not provide uncertainty quantification. A powerful estimation model should not only be able to estimate values, but also exhibit the associated uncertainty [57]. In general, methods that provide quantitative estimates of uncertainty are basically based on the Bayesian framework. Inspired by reference [58] and [59], our future research interests will focus on designing a deep learning model for estimating the SOH that not only has high accuracy and robustness but also provides certainty quantification. In addition, we design two sets of input combinations. Although the estimation accuracy is reduced, the process of feature extraction is simplified and requires fewer conditions. However, our approach requires at least the CC phase of the charging process. So far, our research topic has focused on cyclic SOH estimation, which is to extract features from the charging or discharge cycles and establish mapping between them and SOH, but in practical applications the conditions are more complicated and there are situations where the complete CC phase data are not available. Therefore, our future research direction will also focus on specific practical situations, such as the situation of starting and ending charging at any time when using mobile phones and the solution of EV battery packs containing multiple cells.

V. CONCLUSION

SOH estimation plays a very important role in BMS of Li-ion batteries. In this paper, we proposed a novel data-driven DSTA-GRU method for SOH estimation. The main contributions of our work are as follows: (1) In order to make better use of the advantages of the data-driven approach, we extracted six characteristics that can indirectly reflect the degree of battery aging and analyzed their correlation; (2) To improve the accuracy of the estimation, we add a dynamic spatial attention mechanism and a temporal attention mechanism to the GRU; (3) The effectiveness of dynamic spatial attention and temporal attention has been proved through experiments. The proposed DSTA-GRU method is trained and tested on NASA Li-ion batteries datasets. The experimental results indicate that the proposed DSTA-GRU model outperforms SVM, GPR, RNN, LSTM, and GRU in SOH estimation of Li-ion batteries.

REFERENCES

[1] B. Saha, K. Goebel, S. Poll, and J. Christophersen, "Prognostics methods for battery health monitoring using a Bayesian framework," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 2, pp. 291–296, Feb. 2009.

[2] I.-S. Kim, "A technique for estimating the state of health of lithium batteries through a dual-sliding-mode observer," *IEEE Trans. Power Electron.*, vol. 25, no. 4, pp. 1013–1022, Apr. 2010.

[3] J. Liu and Z. Chen, "Remaining useful life prediction of lithium-ion batteries based on health indicator and Gaussian process regression model," *IEEE Access*, vol. 7, pp. 39474–39484, Mar. 2019.

[4] K. Liu, K. Li, Q. Peng, and C. Zhang, "A brief review on key technologies in the battery management system of electric vehicles," *Frontiers Mech. Eng.*, vol. 14, no. 1, pp. 47–64, Apr. 2018.

[5] Y. Li, S. Zhong, Q. Zhong, and K. Shi, "Lithium-ion battery state of health monitoring based on ensemble learning," *IEEE Access*, vol. 7, pp. 8754–8762, Jan. 2019.

[6] R. Xiong, L. Li, and J. Tian, "Towards a smarter battery management system: A critical review on battery state of health monitoring methods," *J. Power Sources*, vol. 405, pp. 18–29, Nov. 2018.

[7] H. Rahimi-Eichi, U. Ojha, F. Baronti, and M.-Y. Chow, "Battery management system: An overview of its application in the smart grid and electric vehicles," *IEEE Ind. Electron. Mag.*, vol. 7, no. 2, pp. 4–16, Jun. 2013.

[8] X. Zhang, Y. Wang, C. Liu, and Z. Chen, "A novel approach of battery pack state of health estimation using artificial intelligence optimization algorithm," *J. Power Sources*, vol. 376, pp. 191–199, Feb. 2018.

[9] M. Berecibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, and P. Van den Bossche, "Critical review of state of health estimation methods of li-ion batteries for real applications," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 572–587, Apr. 2016.

[10] M. S. H. Lipu, M. A. Hannan, A. Hussain, M. M. Hoque, P. J. Ker, M. H. M. Saad, and A. Ayob, "A review of state of health and remaining useful life estimation methods for lithium-ion battery in electric vehicles: Challenges and recommendations," *J. Cleaner Prod.*, vol. 205, pp. 115–133, Dec. 2018.

[11] B. Balagopal and M. Y. Chow, "The state of the art approaches to estimate the state of health (SOH) and state of function (SOF) of lithium Ion batteries," in *Proc. INDIN*, Cambridge, U.K., 2015, pp. 1302–1307.

[12] L. Zhang, W. Fan, Z. Wang, W. Li, and D. U. Sauer, "Battery heating for lithium-ion batteries based on multi-stage alternative currents," *J. Energy Storage J.*, vol. 32, Dec. 2020, Art. no. 101885.

[13] J. Li, K. Adewuyi, N. Lotfi, R. G. Landers, and J. Park, "A single particle model with chemical/mechanical degradation physics for lithium ion battery state of health (SOH) estimation," *Appl. Energy*, vol. 212, pp. 1178–1190, Feb. 2018.

[14] Z. B. Wei, J. Y. Zhao, D. X. Ji, and K. J. Tseng, "A multi-timescale estimator for battery state of charge and capacity dual estimation based on an online identified model," *Appl. Energy*, vol. 204, pp. 1264–1274, Oct. 2017.

[15] T. F. Fuller, M. Doyle, and J. Newman, "Simulation and optimization of the dual lithium ion insertion cell," *J. Electrochem. Soc.*, vol. 141, no. 1, pp. 1–10, 1994.

[16] S. K. Rahimian, S. Rayman, and R. E. White, "Comparison of single particle and equivalent circuit analog models for a lithium-ion cell," *J. Power Sources*, vol. 196, no. 20, pp. 8450–8462, 2011.

[17] S. Torai, M. Nakagomi, S. Yoshitake, S. Yamaguchi, and N. Oyama, "State-of-health estimation of $LiFePO_4$/graphite batteries based on a model using differential capacity," *J. Power Sources*, vol. 306, pp. 62–69, Feb. 2016.

[18] X. Chen, W. Shen, Z. Cao, A. Kapoor, and I. Hijazin, "Adaptive gain sliding mode observer for state of charge estimation based on combined battery equivalent circuit model in electric vehicles," in *Proc. ICIEA*, Melbourne, VIC, Australia, 2013, pp. 114–123.

[19] H. He, R. Xiong, H. Guo, and S. Li, "Comparison study on the battery models used for the energy management of batteries in electric vehicles," *Energy Convers. Manage.*, vol. 64, pp. 113–121, Dec. 2012.

[20] Q. Wang, Z. Wang, L. Zhang, P. Liu, and Z. Zhang, "A novel consistency evaluation method for series-connected battery systems based on real-world operation data," *IEEE Trans. Transport. Electrific.*, early access, Aug. 20, 2020, doi: 10.1109/TTE.2020.3018143.

[21] J. Kim and B. H. Cho, "State-of-charge estimation and state-of-health prediction of a Li-Ion degraded battery based on an EKF combined with a per-unit system," *IEEE Trans. Veh. Technol.*, vol. 60, no. 9, pp. 4249–4260, Nov. 2011.

[22] T. Feng, L. Yang, X. Zhao, H. Zhang, and J. Qiang, "Online identifcation of lithium-ion battery parameters based on an improved equivalent-circuit model and its implementation on battery state-of-power prediction," *J. Power Sources*, vol. 281, pp. 192–203, May 2015.

[23] J. Yu, B. Mo, D. Tang, H. Liu, and J. Wan, "Remaining useful life prediction for lithium-ion batteries using a quantum particle swarm optimization-based particle filter," *Qual. Eng.*, vol. 29, no. 3, pp. 536–546, 2017.

[24] Q. Ouyang, Z. Wang, K. Liu, G. Xu, and Y. Li, "Optimal charging control for lithium-ion battery packs: A distributed average tracking approach," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3430–3438, May 2020.
[25] S. Mukherjee, E. Osuna, and F. Girosi, "Nonlinear prediction of chaotic time series using support vector machines," in *Proc. 7th Neural Netw. Signal Process. IEEE Signal Process. Soc. Workshop*, Amelia Island, FL, USA, Sep. 1997, pp. 511–520.
[26] A. Nuhic, T. Terzimehic, T. Soczka-Guth, M. Buchholz, and K. Dietmayer, "Health diagnosis and remaining useful life prognostics of lithium-ionbatteries using data-driven methods," *J. Power Sources*, vol. 239, pp. 680–688, Oct. 2013.
[27] V. Klass, M. Behm, and G. Lindbergh, "A support vector machine-based state-of-health estimation method for lithium-ion batteries under electric vehicle operation," *J. Power Sources*, vol. 270, pp. 262–272, Dec. 2014.
[28] M. A. Patil, P. Tagade, K. S. Hariharan, S. M. Kolake, T. Song, T. Yeo, and S. Doo, "A novel multistage support vector machine based approach for Li ion battery remaining useful life estimation," *Appl. Energy*, vol. 159, pp. 285–297, Dec. 2015.
[29] D. Yang, X. Zhang, R. Pan, Y. Wang, and Z. Chen, "A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve," *J. Power Sources*, vol. 384, pp. 387–395, Apr. 2018.
[30] Z. Wang, J. Ma, and L. Zhang, "State-of-health estimation for lithium-ion batteries based on the multi-island genetic algorithm and the Gaussian process regression," *IEEE Access*, vol. 5, pp. 21286–21295, Oct. 2017.
[31] X. Zheng and X. Deng, "State-of-Health prediction for lithium-ion batteries with multiple Gaussian process regression model," *IEEE Access*, vol. 7, pp. 150383–150394, Oct. 2019.
[32] H. Li, D. Pan, and C. L. P. Chen, "Intelligent prognostics for battery health monitoring using the mean entropy and relevance vector machine," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 44, no. 7, pp. 851–862, Jul. 2014.
[33] Y. Zhou and M. Huang, "Lithium-ion batteries remaining useful life prediction based on a mixture of empirical mode decomposition and ARIMA model," *Microelectron. Rel.*, vol. 65, pp. 265–273, Oct. 2016.
[34] Y. Li, C. Zou, M. Berecibar, E. Nanini-Maury, J. C.-W. Chan, P. Van Den Bossche, J. Van Mierlo, and N. Omar, "Random forest regression for online capacity estimation of lithium-ion batteries," *Appl. Energy*, vol. 232, pp. 197–210, Dec. 2018.
[35] S. Song, C. Fei, and H. Xia, "Lithium-ion battery SOH estimation based on XGBoost algorithm with accuracy correction," *Energies*, vol. 13, no. 4, p. 812, Jan. 2020.
[36] G. Bai, P. Wang, C. Hu, and M. Pecht, "A generic model-free approach for lithium-ion battery health management," *Appl. Energy*, vol. 135, pp. 247–260, Dec. 2014.
[37] J. Wu, Y. J. Wang, X. Zhang, and Z. H. Chen, "A novel state of health estimation method of Li-ion battery using group method of data handling," *J. Power Sources*, vol. 327, pp. 457–464, Sep. 2016.
[38] C. She, Z. Wang, F. Sun, P. Liu, and L. Zhang, "Battery aging assessment for real-world electric buses based on incremental capacity analysis and radial basis function neural network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3345–3354, May 2020.
[39] F. Feng, S. Teng, K. Liu, J. Xie, Y. Xie, B. Liu, and K. Li, "Co-estimation of lithium-ion battery state of charge and state of temperature based on a hybrid electrochemical-thermal-neural-network model," *J. Power Sources*, vol. 455, p. 227935, Apr. 2020.
[40] H. Chaoui and C. Christopher Ibe-Ekeocha, "State of charge and state of health estimation for lithium batteries using recurrent neural networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 8773–8783, Oct. 2017.
[41] Y. Wu, Q. Xue, J. Shen, Z. Lei, Z. Chen, and Y. Liu, "State of health estimation for lithium-ion batteries based on healthy features and long short-term memory," *IEEE Access*, vol. 8, pp. 28533–28547, Feb. 2020.
[42] B. Xiao, Y. Liu, and B. Xiao, "Accurate State-of-Charge estimation approach for lithium-ion batteries by gated recurrent unit with ensemble optimizer," *IEEE Access*, vol. 7, pp. 54192–54202, 2019.
[43] Y. Song, L. Li, Y. Peng, and D. Liu, "Lithium-ion battery remaining useful life prediction based on GRU-RNN," in *Proc. 12th Int. Conf. Rel., Maintainability, Saf. (ICRMS)*, Oct. 2018, pp. 317–322.
[44] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. ICLR*, 2015, pp. 1–15.
[45] S. Kim, T. Hori, and S. Watanabe, "Joint CTC-attention based end-to-end speech recognition using multi-task learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 4835–4839.
[46] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," 2015, *arXiv:1508.04025*. [Online]. Available: http://arxiv.org/abs/1508.04025
[47] J. Lu, C. Xiong, D. Parikh, and R. Socher, "Knowing when to look: Adaptive attention via a visual sentinel for image captioning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3242–3250.
[48] B. Saha and K. Goebel. (2007). Battery Data Set. NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA, USA. [Online]. Available: http://ti.arc.nasa.gov/project/prognostic-data-repository
[49] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1–9.
[50] X. Ma, Z. Chen, and J. Zhang, "Fully convolutional network with cluster for semantic segmentation," in *Proc. CVPR*, 2018, pp. 1–7.
[51] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[52] Y. Chen, X. Dai, M. Liu, D. Chen, L. Yuan, and Z. Liu, "Dynamic convolution: Attention over convolution kernels," in *Proc. CVPR*, 2020, pp. 11030–11039.
[53] Y. Qin, D. Song, H. Cheng, W. Cheng, G. Jiang, and G. W. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," in *Proc. IJCAI*, 2017, pp. 2627–2673.
[54] C. Fan, Y. Zhang, Y. Pan, X. Li, C. Zhang, R. Yuan, D. Wu, W. Wang, J. Pei, and H. Huang, "Multi-horizon time series forecasting with temporal attention learning," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 2527–2535.
[55] Y. Chen, W. Lin, and J. Z. Wang, "A dual-attention-based stock price trend prediction model with dual features," *IEEE Access*, vol. 7, pp. 148047–148058, Oct. 2019.
[56] H. Noh, T. You, J. Mun, and B. Han, "Regularizing deep neural networks by noise: Its interpretation and optimization," in *Proc. NIPS*, 2017, pp. 1–10.
[57] K. Liu, X. Hu, Z. Wei, Y. Li, and Y. Jiang, "Modified Gaussian process regression models for cyclic capacity prediction of lithium-ion batteries," *IEEE Trans. Transport. Electrific.*, vol. 5, no. 4, pp. 1225–1236, Dec. 2019.
[58] K. Liu, Y. Shang, Q. Ouyang, and W. D. Widanage, "A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery," *IEEE Trans. Ind. Electron.*, vol. 68, no. 4, pp. 3170–3180, Apr. 2021.
[59] K. Liu, T. R. Ashwin, X. Hu, M. Lucu, and W. D. Widanage, "An evaluation study of different modelling techniques for calendar ageing prediction of lithium-ion batteries," *Renew. Sustain. Energy Rev.*, vol. 131, Oct. 2020, Art. no. 110017.

SHENGMIN CUI received the B.S. degree in computer science and technology from Dalian Nationalities University, China, in 2010, and the master's degree in computer software and theory from Yunnan University, China, in 2013. He is currently pursuing the Ph.D. degree in computer science with Hanyang University, South Korea. His research interests include the Internet of things, wireless network resource management, machine learning, deep learning, and reinforcement learning.

INWHEE JOE received the B.S. degree in electronics engineering from Hanyang University, Seoul, South Korea, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 1998. Since 2002, he has been a Faculty Member with the Division of Computer Science and Engineering, Hanyang University, Seoul. His current research interests include the Internet of Things, cellular systems, wireless power communication networks, embedded systems, network security, machine learning, and performance evaluation.