IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 18 October 2024, accepted 4 November 2024, date of publication 7 November 2024, date of current version 18 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3493141

RESEARCH ARTICLE

State-of-Charge Estimation of Lithium–Ion Battery Integrated in Electrical Vehicle Using a Long Short-Term Memory Network

CHI NGUYEN VAN[1], MINH DUC NGO[2], CUONG DUONG DUC[1], LE QUANG THAO[3], AND SEON-JU AHN[4], (Member, IEEE)

[1]Department of Control and Measurement, Thai Nguyen University of Technology, Thai Nguyen 251750, Vietnam
[2]Department of Automation, Thai Nguyen University of Technology, Thai Nguyen 251750, Vietnam
[3]Faculty of Physics, VNU University of Science, Hanoi 100000, Vietnam
[4]Department of Electrical Engineering, Chonnam National University, Gwangju 61186, South Korea

Corresponding author: Seon-Ju Ahn (sjahn@chonnam.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) through Korean Government (MSIT) under Grant RS-2024-00454464; and in part by Thai Nguyen University of Technology (TNUT), Vietnam.

ABSTRACT Improving the accuracy of state-of-charge (SoC) estimation is crucial for electric vehicles (EVs) using Lithium-Ion batteries (LiBs). This helps users reliably predict driving range and optimize the charging process, thereby extending battery life and ensuring safety during use. However, due to temperature, driving mode, and charge-dependent electrochemical nonlinear dynamics, SoC estimation for LiB integrated with EVs remains a significant technical challenge. In particular, SoC estimation in the regions of SoC $< 30\%$ and SoC $> 80\%$ is often inaccurate due to nonlinearity and sensitivity to battery aging. Accurate estimation in these regions is crucial for making decisions regarding recharging and discharging to prolong battery life and prevent damage. To address this issue, this paper proposes a method for SoC estimation using a Long Short-Term Memory (LSTM) network, which is capable of retaining information on battery characteristics related to changes long term electrochemical parameter changes, such as the number of discharge cycles and the aging effects. The method utilizes practical data from 80,000 samples collected from pure electric vehicle testing under different driving modes, temperatures, and road conditions over a 30-day period. The LSTM network was optimized by adjusting the input data sequence and hidden size to minimize the number of hyperparameters. This makes it suitable for use on low-cost processors with moderate computing power. SoC estimation was evaluated across four SoC test regions: SoC $< 30\%$, SoC $> 80\%$, $30\% \leq$ SoC $\leq 80\%$, and $0\% \leq$ SoC $\leq 100\%$. The results were compared with feedforward neural network (FNN) and convolutional neural network (CNN). Despite having a configuration with a hidden size of 96 and a single layer, the LSTM model achieved estimation accuracy with RMSE $= 0.0106$, MAE $= 0.0077$, and MAPE $= 1.4116\%$.

INDEX TERMS Lithium-ion battery, SoC estimation, long short-term memory network, electric vehicles, feedforward neural network, convolutional neural network.

I. INTRODUCTION

Today, fossil fuel resources are gradually depleting due to overexploitation to meet the industrial and consumer demands, primarily for personal transportation means. This overexploitation is the main cause of greenhouse gas emissions and global warming, making Earth's climate more severe. Therefore, transitioning to the use of electric vehicles is an inevitable trend that has been accepted and adopted by countries worldwide, with each outlining individual transition pathways. The demand for electric vehicles and the growth of the electric vehicle market are influenced by various drawbacks. These include consumer worries about the remaining battery capacity during long-distance travel, the accuracy of battery parameter measurements, battery lifespan, replacement costs, and thermal safety [1], [2].

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

Currently, LiBs are widely used in electric vehicles due to their long lifespan, low self-discharge rate, high energy density, and usage in smart grid storage systems and renewable energy storage. To drive the development of the electric vehicle market, two trends are being implemented: 1) Developing new materials with better electrochemical properties to manufacture batteries with larger capacity, higher density, faster charging, and improved safety. 2) Advancing methods with higher accuracy to estimate electrochemical parameters inside cells such as SoC, State of Health (SoH), in order to optimize charging and discharging control actions, ensuring the battery remains in the healthiest state and ensuring absolute safety for users [3]. SoC is a crucial parameter of the battery, representing the remaining capacity within the cell at the time of measurement compared to the rated capacity of the cell, expressed by (1):

$$SoC\ (t) = \frac{Q(t)}{Q_f(t)} 100\% \tag{1}$$

where $Q_f\ (t)$ is the capacity of the cell when fully charged, this value changes slowly over time depending on the cell's aging level; $Q(t)$ is the remaining discharge capacity at the time of measurement. According to (1) above, the value of SoC varies in the range from 0% to 100%. If SoC = 100%, it means the battery is fully charged, and when SoC = 0%, the battery is considered fully discharged. Due to the characteristics of LiB, operating them when SoC is $< 20\%$ is not permitted in electric vehicles, requiring the battery to be recharged. SoC depends on several factors, with the temperature and aging (or usage time) being among the most prominent. As the temperature increases, the actual charge in the battery will decrease when compared to the same SoC value. LiBs are recommended for use within a temperature range of -10 to 50 degrees Celsius. Similarly, as the battery usage time increases, the battery will gradually age, and when this happens, 100% SoC would be equivalent to a 75% – 80% SoC of a new cell [4]. SoC is a parameter that cannot be directly measured in commercial applications such as electric vehicles because the measuring devices are bulky and expensive, and they can only be operated in laboratory settings. Therefore, in all electric vehicle applications, SoC needs to be estimated based on variables such as voltage, current, temperature, and other parameters. Currently, many researchers are focused on developing methods to estimate SoC. However, the nonlinearity of SOC and its dependence on temperature and aging make accurate SoC estimation very challenging, especially when SoC is $<30\%$ and SoC $>$ 85%, and when the battery has aged (old battery cell). There are three main methods for estimating SoC: traditional methods, model-based methods, and data-driven methods. For traditional methods, SoC is estimated based on the relationship between open-circuit voltage and SoC [5], Coulomb counting method [6], or using electrochemical impedance spectroscopy (EIS) [7], [8]. SoC estimated by traditional methods often suffers from cumulative errors, inaccuracies in the one-to-one relationship between open-circuit voltage (OCV) and SoC, and is therefore not widely used in the field of electric vehicles. For model-based methods, SoC estimation uses Kalman filters, Extended Kalman Filter, Sigma Point Kalman Filter, etc., for a mathematical model of the battery cell (usually an equivalent circuit model or electrochemical model) [9], [10], [11]. However, the errors in these methods will gradually increase as the battery cells age and sudden temperature variations occur during use, because the parameters in the model used to estimate SoC do not accurately reflect the actual cell values at the time of estimation. For data-driven methods, SoC estimation uses trainable neural networks to determine SoC based on the values of input data. The advantage of this method is that it does not require a model; however, selecting data to preserve the characteristics of the cell, especially temperature dependence and aging, is crucial for accurately estimating SoC. Additionally, the proper structure of the neural network and choice of learning algorithm also determine the success of accurately estimating SoC throughout the battery's usage. Complex, multi-parameter neural networks can offer accurate SoC estimation, but in practical deployment, they require powerful processors, leading to higher hardware costs and reducing the competitiveness of electric vehicles. Authors in the literature have used deep neural network (DNN) to directly estimate SoC from voltage, current, and temperature data of the battery packs, achieving a minimum error of 2.17% mean absolute error (MAE) [12]. However, the drawback of this method is the large estimation error. Hansen and Wang [13] and Anton et al. [14], developed a SoC estimation method based on support vector machine (SVM), using US06 cycling driving data for SoC estimation results with a root mean square (RMS) error less than 6%. Authors in [15] used a 3-layer feed-forward neural network, using three input data of current, voltage, and temperature and adding first and second-order derivatives of them to estimate SoC for Ni-MH batteries. The estimation results showed a relatively large estimation variance due to over-fitting or under-fitting phenomena. A common weakness of SoC estimation is that all estimation methods tend to exhibit large errors when the SOC is below 30% or above 80%, highlighting the limitations of the above approaches. This is due to the increased nonlinearity of SoC in these two ranges, driven by the electrochemical characteristics of the cell. On the other hand, the change of SoC with aging also increases the estimation error over time. Therefore, although FNN and CNN can estimate SoC quite accurately with new cells, the estimation error increase as the cells age [16]. LSTM neural networks have significantly advanced the accuracy of SoC estimation for LiB, which is crucial for the reliable operation of electric vehicles and the management of battery-powered systems. The LSTM's ability to learn and retain information over long sequences makes it particularly well-suited for modeling the complex, nonlinear dynamics of battery behavior under various conditions. With an increase in hidden nodes, the ability to represent features of aging cells

may improve. This network shows promising potential, with higher accuracy in estimating SoC [6], [17]. Recent studies have demonstrated the efficacy of LSTM models in capturing the intricate relationships between SoC and factors such as temperature, voltage, and current. For example, an LSTM-based approach has been shown to accurately estimate SoC with errors of less than 1.5% under varying ambient temperatures, representing a significant improvement over traditional methods [18]. Another study employed deep LSTM networks to estimate SoC across different battery specifications and discharge cycles, highlighting the model's versatility and robustness [19]. The authors in document [17] applied Bi-LSTM networks to estimate SoC, with MAEs of the estimation error at some temperatures: 0°C, 10°C, and 25°C are 0.498%, 0.411%, and 0.738% respectively. However, this method has not yet been improved in terms of accuracy. The key findings of the papers on the use of LSTM neural networks for estimating the SoC of LiB are quite ground-breaking and make a significant contribution to the field of battery management system (BMS). Firstly, LSTM's ability to process and learn from timeseries data has proven highly effective for SoC estimation, which is a dynamic process influenced by various factors such as temperature, aging, and load cycles. These papers demonstrate that LSTM models can accurately track the non-linear characteristics of batteries, providing more precise SOC readings than traditional methods. Some of the research has shown that incorporating environmental variables such as temperature into the LSTM model can greatly improve the accuracy of SoC predictions. This is particularly important for EVs operating in varying climatic conditions. Studies [17], [19], [20] show that LSTM networks can be trained with a relatively small amount of data to predict SoC with high accuracy, which is beneficial for reducing computational costs and improving the efficiency of the training process. Another important finding is the potential of LSTM models to generalize across different types of LiB. This means that a model trained on one battery type can be adapted to other types with minimal adjustments, making it a versatile tool for SoC estimation. Moreover, the integration of LSTM with other techniques, such as Kalman filters, has been shown to further refine the estimation process, suggesting a hybrid approach could be the way forward in advanced BMS [21], [22].

This article presents the use of LSTM networks to estimate the SoC of electric vehicle batteries during use, selecting several more relevant variables. The data we collected from 80,000 samples, including vehicle velocity, pack bus voltage, pack current, electric motor speed, electric motor temperature, electric motor control voltage, average cell voltage, and average cell temperature collected from pure electric vehicle tests under different driving modes, temperatures and road conditions for 30 days. After analyzing their correlation, we selected eight variables as inputs, with SoC as the output. After analyzing the characteristics of SoC, we established training data intervals and estimation data intervals (choosing different intervals, focusing on those with the most variation), especially in the final stages of the test cycle. Selecting data intervals should focus on ranges where SoC varies most, quickly, and notably contain intervals with small SoC values. Next, an LSTM network was set up to estimate SoC. The LSTM network is then calibrated to reduce the number of hyperparameters and the length of the data series with the aim of reducing the computational level to be able to deploy to low-cost hardware. Using three criteria to evaluate the SoC estimation error, including RMSE, MSE and MAPE for four regions SoC $>$ 80%, SoC $<$30%, 30% $\leq$ SoC $\leq$ 80%, and 0% $\leq$ SoC $\leq$ 100%, and comparing with the FNN and CNN models, it shows that the estimation error of the LSTM network is the smallest.

The structure of this article is as follows: Part 1 serves as a general introduction; Part 2 discusses the experimental data collection methods and data preparation steps; Part 3 covers the configuration of the LSTM network and the training steps, as well as the selection of training data, evaluation data, and estimation data. Part 4 presents the SoC estimation results, evaluation, and comments; Part 5 includes discussions and exchanges, and conclusions.

II. DATA PREPATION STEPS

In this study, we utilized data collected from the platform of a pure electric vehicle (EV) operated under various road conditions, temperatures, and driving modes. The data was sampled using a 10-second/sample cycle over a period of 30 days. The electric vehicle has a battery pack consisting of 95 series-connected cell modules with a nominal bus voltage of over 300V. The collected data includes odometer readings, vehicle speed, electric motor speed, electric motor torque, electric motor temperature, motor voltage, and current, while regarding the battery pack, data collected includes bus voltage, current through the battery, actual SoC, battery insulation resistance, voltage across the 95 battery cells, and 30 temperature measurement points. To determine the actual SoC, the coulombic efficiency is calculated as a function of the temperature, from the total ampere-hours discharged and ampere-hours charged int the test scenarios. Based on the coulombic efficiency, the depth of discharge (in ampere-hours) is calculated at each point in time. The discharge capacity at the time of measurement is $Q(t)$ is 1 - depth of discharge(t), and finally, the actual SoC is calculated from (1). These data are shown in Figure 1.

With continuous testing over 30 days, comprising nearly 80,000 data points, features related to battery aging and SoC fluctuations were observed in the current, voltage, and temperature data. Driving mode characteristics were reflected in vehicle speed, motor torque, and electric motor speed. Battery capacity showed an inverse correlation with the vehicle's distance traveled, with the rate of battery capacity depletion being proportional to the electric motor speed. For the neural network training process to be effective, the collected data needs to be analyzed for its correlation with SoC. Some parameters should exhibit a

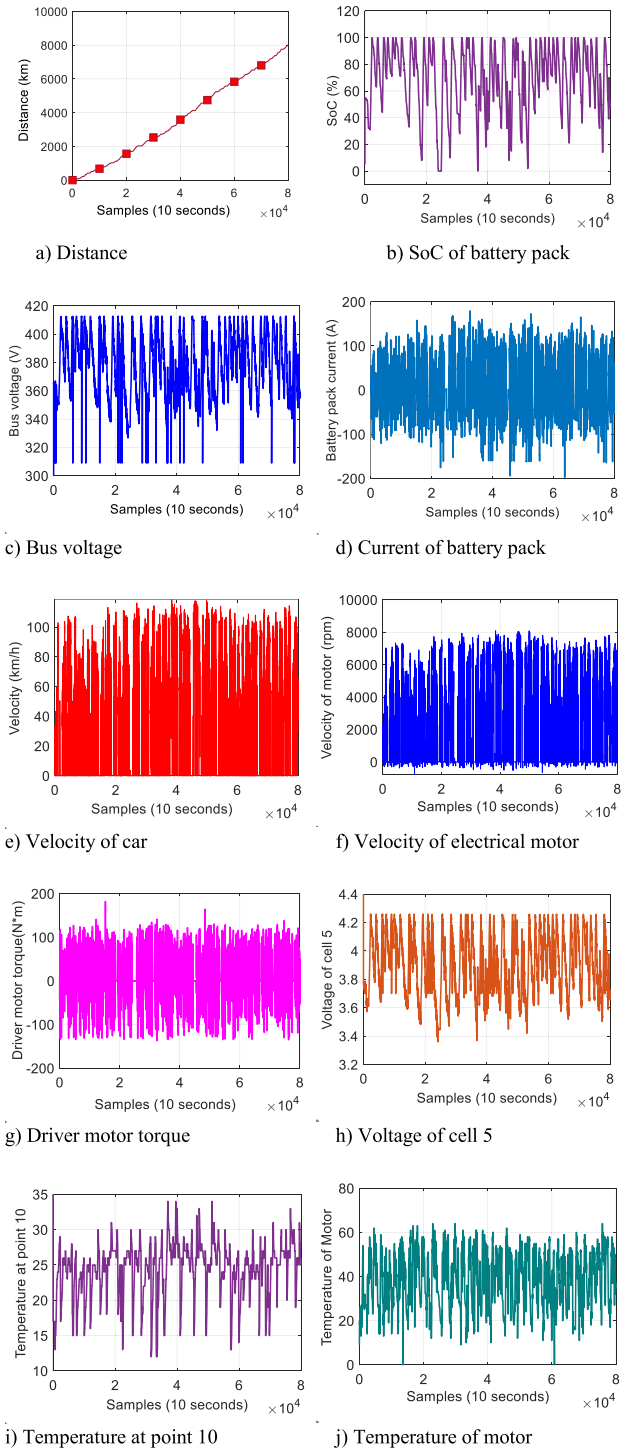

FIGURE 1. **Input-output data after normalization to the range from 0 to 1. The upper plot shows the input data consisting of 8 types, within approximately 80,000 sampled points. Similarly, the lower plot depicts the normalized output data (SoC).**

strong correlation with SoC, while others may show a weaker correlation.

In practice, there are two methods to analyze correlation coefficients: the Pearson method and the Spearman method. In this study, we employed the Spearman method because the collected data is influenced by noise and has high nonlinearity. The formula for calculating the Spearman correlation coefficient is as follows:

$$\rho_{X,Y} = 1 - \frac{6\sum_{i=1}^{n} d_i}{n(n^2-1)} \tag{2}$$

in which, $-1 \leq \rho_{X,Y} \leq 1$ is Spearman's rank correlation coefficient of two observations X, Y; d_i is difference between the two ranks of each observation, n is number of observations. A value of $+1$ indicates a perfect positive rank association, a value of 0 means no association between ranks, a value of -1 means a perfect negative association between ranks. Figure 2 describes the correlation coefficients of the variables with SoC.

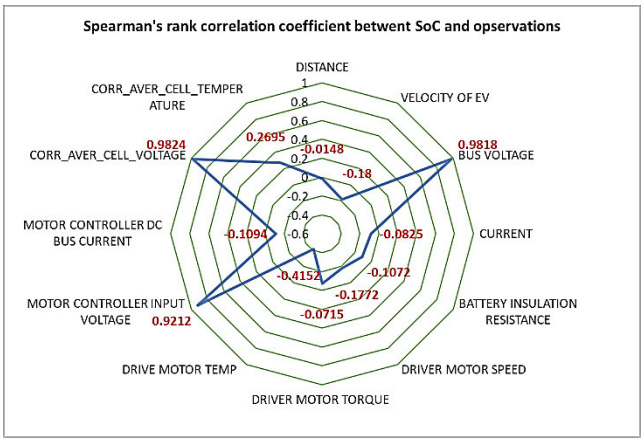

FIGURE 2. **Spearman's rank correlation coefficient between SoC and observations.**

Figure 2 shows that voltage is the most relevant parameter to SoC, followed by some motor-related parameters. Considering that the Spearman correlation coefficient is sensitive only to the direction and intensity of changes between variables, it cannot capture variables with strong instantaneous changes in speed and total current. Furthermore, because mileage is a continuously increasing parameter, it is not constrained by the correlation coefficient. Taking all factors into account, the input parameters, excluding battery insulation resistance (KΩ), motor torque (Nm), and DC bus current of motor controller (A), will be used as the final input parameters. Therefore, the input parameters used to train the network in this study are eight parameters: Velocity (km/h), Bus voltage (V), Current (A), Driver motor speed (r/min), Drive motor temp (^{0}C), Motor controller input voltage (V), average voltage of cells (V), and average cell temperature (^{0}C). The distance traveled is a parameter related to the aging process. These parameters are normalized to a range of 0 to 1 as follows:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{3}$$

where, x_{min} and x_{max} are the minimum and maximum values of the observed variable X. After normalization, the input and output data are as described in Figure 3.

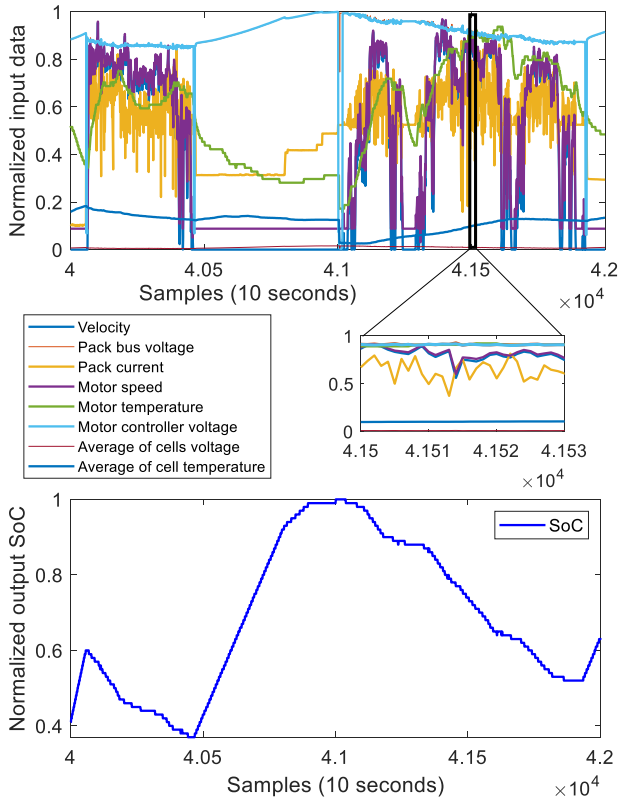

FIGURE 3. Input-output data after normalization to the range from 0 to 1. The upper plot shows the input data consisting of 8 types, within approximately 80,000 sampled points. Similarly, the lower plot depicts the normalized output data (SoC).

The selection of the dataset used for estimation: from the normalized output data, the principle is to select data segments with rapid changes in SoC, especially those with many segments where $SoC < 30\%$ and $SoC > 80\%$. These are the segments where the nonlinearity of SoC with respect to the input variables is most pronounced, and it is crucial that SoC is accurately represented in the lower regions. We selected data for training in the following regions (in the red-colored cells, in consecutive order) to ensure that aging characteristics are represented in the data. Based on the length of the input data series, the input data series is divided into 6 regions. The input data region corresponding to the output normalized SoC data with the most values between 0.3 and 0.8 will be selected as the testing set. Dividing the data in this way will help the model focus on learning regions with values below 0.3 and above 0.8, because these value regions have very high variability. The description of the data regions selected for network training and testing is shown in Figure 4.

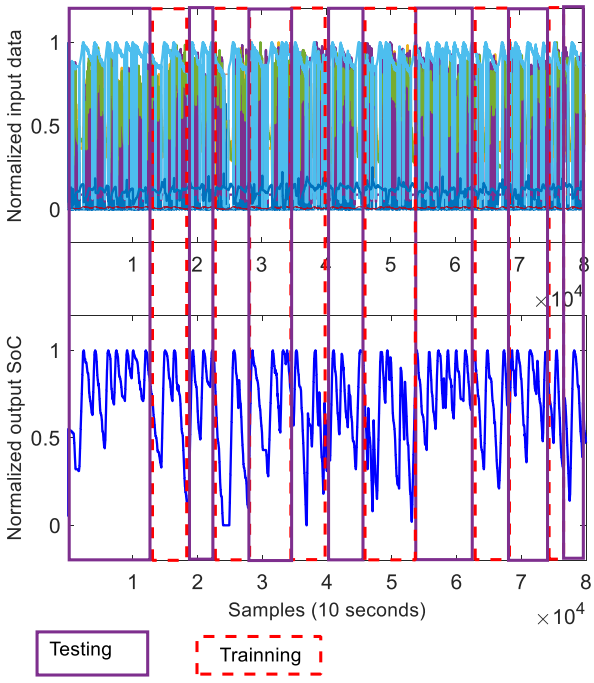

FIGURE 4. The data selected for training and testing.

III. CONFIGURATION OF LSTM AND TRAINING

A. LSTM NETWORK INITIALIZATION

In this study, we use an LSTM network with a single node structure as described in Figure 5, where x_n is the normalized input including velocity, bus voltage, bus current, drive motor speed, drive motor temperature, motor controller input voltage, average cell voltage, and average cell temperature. c_n and h_n represent the state output and hidden output of the network node n, respectively. These outputs are recursively used to estimate SoC of the battery cell. c_{n-1} and h_{n-1} are the two outputs of the previous network node $n-1$. The σ symbol represents the sigmoid activation function with a matrix output, and tanh is the hyperbolic tangent function with a vector output. Each LSTM network node has three gates: the forget gate (f_n), the input gate (i_n) and the output gate (o_n). These gates are implemented by a regular neural network and act as filters of information and functions.

The update formulas for the gates are as follows:

- Input gate: Control the update rate of new state according to (4)

$$i_n = \sigma(W_i[h_{n-1}, x_n] + b_i) \tag{4}$$

- Forget gate: Control the forgetting rate of old state according to (5)

$$f_n = \sigma(W_f[h_{n-1,}x_n] + b_f) \tag{5}$$

- Output gate: Control the update rate of state according to (6)

$$o_n = \sigma(W_o[h_{n-1}, x_n] + b_o) \tag{6}$$

where, $W_i, b_i, W_f, b_f, W_o, b_o$ are the weight and bias matrices, respectively. Each network node n has a vector $\tilde{c}_n$ determined through a hyperbolic tangent function centered at 0, with a gradient distribution designed to prevent vanishing phenomena. $\tilde{c}_n$ is used to adjust the update rate of the new state of the state cell, and it is calculated using (7),

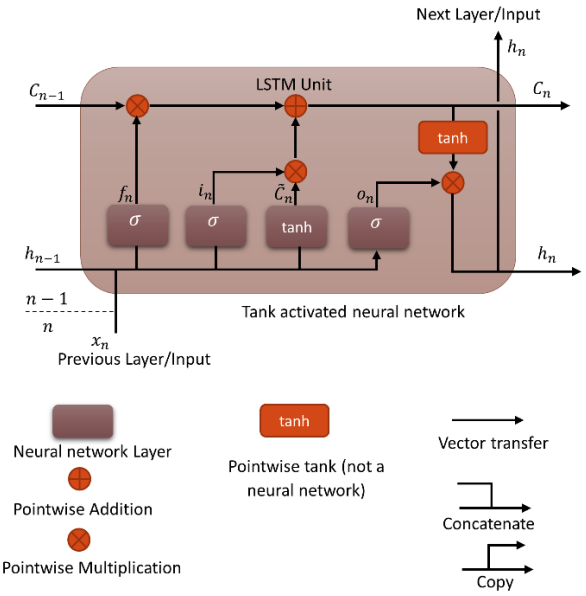

FIGURE 5. Structure of an LSTM network node.

where W_c and b_c are the weight matrix and bias vector, respectively.

$$\tilde{c}_n = tanh(W_c[h_{n-1}, x_n] + b_c) \tag{7}$$

The output of the network node's state c_n is calculated using (8) and consists of two components: the previous state component of the cell to be forgotten, $f_n * c_{n-1}$, and the new state component to be updated, $i_n * \tilde{c}_n$.

$$c_n = f_n * c_{n-1} + i_n * \tilde{c}_n \tag{8}$$

The hidden state output of network node is calculated using (9).

$$h_n = o_n * tanh(c_n) \tag{9}$$

The parameters of the LSTM network node include bias vectors b_i, b_f, b_c, b_0 and weight matrices W_i, W_f, W_c, W_0 across four layers. From the above formulas, we observe that $f_n < 0, i_n < 1, o_n < 1$. This suggests that $h_n, \tilde{c}_n$ are quite similar to those in RNN, making them capable of remembering short-term information. In contrast, c_n is functions like a transmission belt in the RNN model, selecting which important remote information will be sent in and used later as needed.

B. CHOOSING A LEARNING ALGORITHM

The commonly used algorithms for training LSTM networks include the following:

- 'sgdm' — Use the stochastic gradient descent with momentum (SGDM) optimizer, in which the value of momentum can be chosen
- 'rmsprop' — Use the RMSProp optimizer, which can adjust the decay rate of the squared gradient moving average

TABLE 1. Training parameters.

Hyperparameters	Values	Descriptions
Batch size	256	Amount of data fed into the model at each learning step
Epoch	100	Number of model iterations observing the entire dataset
Learning rate	0.0001	Model weight update rate
Loss function	MSE	Error function for training
Optimizer	Modified Adam	Optimal algorithm for training

- 'adam' — Use the Adam optimizer, which can also adjust the decay rates of both the gradient and squared gradient moving averages

In this study, we use the adam optimization algorithm to optimize the weights at each step, enabling faster convergence compared to methods using the slope of the random gradient. This approach is particularly suitable for models with many parameters, as the learning rate adjusts based on the distortion of the moving average of the gradient and the square gradient. The training parameters are listed in Table. 1.

C. TURNING THE HYPERPARAMETERS OF THE MODEL

We use a standard LSTM model with the configuration described in Figure 5 with the hyperparameters listed in Table. 2. To evaluate the SoC estimation error, we use three error evaluation criteria: MSE, RMSE, and MAPE as defined in (10)-(12). The MSE criterion is commonly used and is calculated based on the mean square of the SoC estimation error. The closer the MSE value is to 0, the better the model's performance. RMSE measures the dispersion of the SoC estimation error and indicates the difference between the SoC values estimated by the model and the actual value. RMSE is useful for understanding the absolute fit of a model. The MAPE criterion helps to evaluate the error relative to the correct data, so the MAPE criterion should be used when evaluating the estimation error with different data sets. The configuration of the LSTM model with hyperparameters listed in Table. 2 is shown in Figure 6a.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2 \tag{10}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2} \tag{11}$$

$$MAPE\ (\%) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| 100 \tag{12}$$

Firstly, the parameter in LSTM that is adjusted is the length of the input data sequence. If the sequence is too long, it may introduce unnecessary noise. Reducing the length of the sequence allows the model to focus more on important signals, potentially improving its ability to capture context, as the output data may depend only on a small segment of previous data. By keeping hidden size at 128 and the number or layers at 4, and varying the sequence length with five values (256, 128, 64, 32 and 16), we get the results describing

TABLE 2. Hyperparameters of the standard LSTM network model.

Hyperparameters	Values	Descriptions
Seq_length	256	Length of each input sequence
Num_layer	4	Number of LSTM layers in the model
Hidden state	128	Dimension of hidden state at each LSTM layer
Input_size	8	Number of input data channels
Output_size	1	Number of output data channels

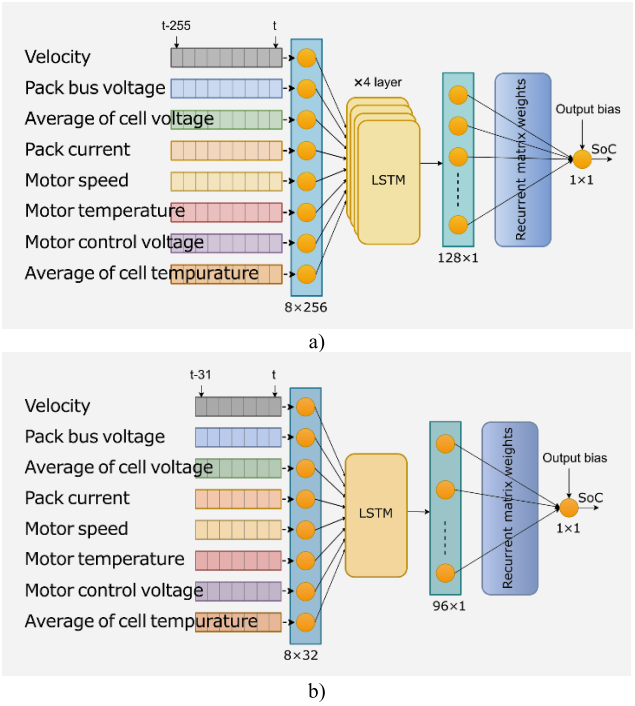

FIGURE 6. a) The configuration of standard LSTM model with hyperparameters listed in Table. 2 and b) configuration of the final LSTM model used for SoC estimation.

the values of FLOPs, RMSE, MAP and MAPE, as shown Table. 3.

From the results in Table. 3, we observe that reducing the sequence length significantly decreases the FLOPs index, with a slight improvement in error performance. However, when the sequence length is reduced to 16, although the computational resources are reduced, the error increases substantially due to the loss of critical information related to the output sequence. Therefore, we select a sequence length of 32 to continue to investigate the effects of changes in hidden size and the number of layers. By reducing the hidden size to 128, 96, and 64, and the number of layers to 4, 3, 2 and 1, respectively, we obtain the results on the number of training parameters, the values of FLOPS and, the error evaluation metrics (RMSE, MAE, and MAPE), as shown in Table. 4.

The results in Table. 4 show that reducing the hidden size significantly decreases the amount of computational resources and memory usage. Additionally, reducing the hidden size lowers the possibility of overfitting the model.

TABLE 3. Comparison results of FLOPs, RMSE, MAP and MAPE when changing the length of data series for 05 cases.

Name	SEQ_LENGTH	FLOPs	RMSE	MAE	MAPE
V1	256	118M	0.011	0.0086	1.4783
V2	128	59M	0.0095	0.007	1.2353
V3	64	30M	0.0097	0.0075	1.3565
V4	**32**	**15M**	**0.0109**	**0.008**	**1.4955**
V5	16	7M	0.012	0.0085	1.5686

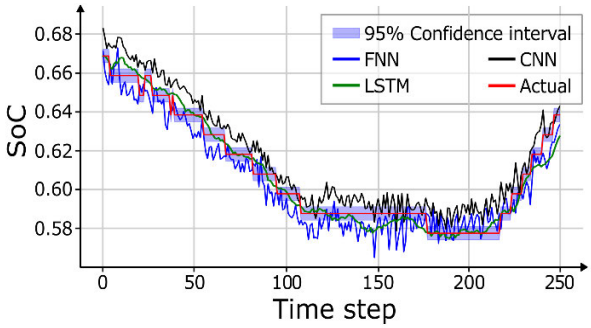

FIGURE 7. Comparison of SoC estimation results in the range of 30% ≤ SoC ≤ 80% for the FNN, CNN, and LSTM models.

However, when the hidden size is reduced to 64, the model is unbale to capture complex relationships in the data series. Therefore, we select the V11 configuration with a hidden size of 96 and one layer. While deeper LSTM models can learn more complex relationships, they often face challenges such as vanishing or exploding gradients, which make training difficult. Reducing the number of layers also improves the computational speed and resource efficiency. Thus, the V11 LSTM model provides a balance between the number of FLOPs and the output error. The configuration of the V11 LSTM model is shown in Figure 6b.

IV. SoC ESTIMATION RESULTS

For comparison, we used CNN and FNN for evaluation. Both FNN and CNN are widely recognized for their strong performance in time series forecasting tasks, as they can model the nonlinear characteristics of data that are correlated over long periods. Therefore, the results of the proposed LSTM model were compared with those obtained from applying the FNN and CNN models to the same dataset. The SoC estimation results were evaluated across four SoC test regions: SoC $<$ 30%, SoC $>$ 80%, 30% $\leq$ SoC $\leq$ 80%, and 0% $\leq$ SoC $\leq$ 100%. Figures 7, 8 and 9 compare the SoC estimation results for the three regions using the FNN, CNN and LSTM models, respectively. Table 5 presents the error metrics (RMSE, MAE and MAPE) when estimating SoC using the LSTM model in this study, as well as the FNN and CNN models. Figure 10 provides a comparison of error metrics for the LSTM, FNN, and CNN estimation models.

The criteria for evaluating the SoC estimation error across the 0% to 100% range are described in Table. 5. The proposed LSTM model achieved RMSE, MAE, and MAPE values of

TABLE 4. Comparison of FLOPs, RMSE, MAP, and MAPE results for five cases with varying hidden sizes of data series.

Name	HIDDEN	Layer	Parameters	FLOPs	RMSE	MAE	MAPE
V6	128	4	465K	15M	0.0109	0.008	1.4955
V7	96	4	262K	8M	0.0104	0.0075	1.355
V8	64	4	117K	4M	0.0112	0.0082	1.4633
V9	96	3	188K	6M	0.0106	0.0076	1.402
V10	96	2	144K	4M	0.0101	0.0073	1.3751
V11	**96**	**1**	**40K**	**1M**	**0.0106**	**0.0077**	**1.4116**

TABLE 5. Comparison of error metrics for SoC estimation using the proposed LSTM model and FNN, CNN models across various SoC regions.

Name	SoC < 30%			30% ≤ Soc ≤ 80%			SoC > 80%			0% ≤ Soc ≤ 100%		
	RMSE	MSE	MAPE	RMSE	MSE	MAPE	RMSE	MSE	MAPE	RMSE	MSE	MAPE
LSTM	**0.0169**	**0.0122**	**5.7849**	**0.0082**	**0.0059**	**0.6459**	**0.0114**	**0.0082**	**1.5825**	**0.0106**	**0.0077**	**1.4116**
CNN	0.0415	0.0104	15.9699	0.0111	0.0059	0.9744	0.0148	0.0089	2.1980	0.0144	0.0104	2.0146
FNN	0.0425	0.0352	16.1799	0.0141	0.0115	1.2565	0.0175	0.0125	2.4004	0.016	0.0116	2.2281

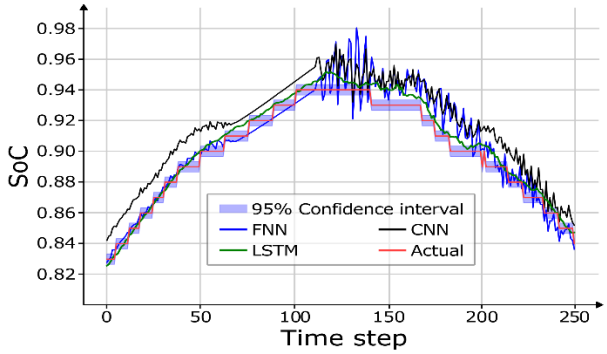

FIGURE 8. Comparison of SoC estimation results in the SoC > 80% region for the FNN, CNN, and LSTM models.

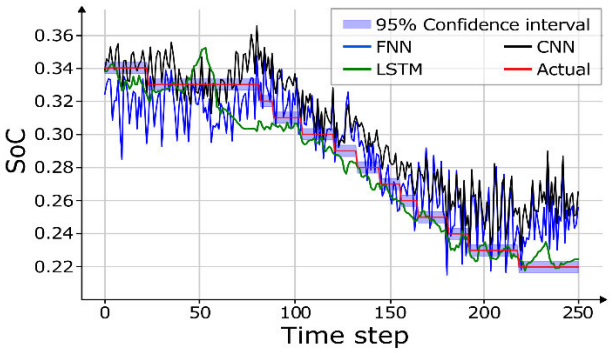

FIGURE 9. Comparison of SoC estimation results in the SoC < 30% region for the FNN, CNN, and LSTM models.

0.0106, 0.0077, and 1.4116%, respectively. Low RMSE and MAE indicate minimal prediction error and high accuracy in forecasting actual values. The low MAPE further indicates that the LSTM model provides high accuracy in predictions compared to actual values, with an average error below 2.3%.

The SoC estimation error is smallest in the range 30% ≤ SoC ≤ 80%, while the largest error occurs in the SoC < 30% range, which aligns with the electrochemical behavior of the battery. Overall, the LSTM model consistently shows smaller errors in nearly all regions. However, it is noted that the LSTM model exhibits a higher MSE than CNN in the SoC < 30% range (0.0122 for LSTM compared to 0.0104 for CNN), despite demonstrating better overall predictive capability.

The LSTM model, with an RMSE of 0.0106, demonstrates high prediction accuracy and very small average prediction errors. Since RMSE is particularly sensitive to large errors, a low RMSE value indicates that the LSTM model is not only accurate but also less impacted by noise during estimation. The MAE value of 0.0077 further highlights the model's precision, as it suggests very small absolute errors, meaning the LSTM model's predictions closely match actual values.

Unlike RMSE, which can be influenced by large errors, MAE provides a straightforward view of the average error. The small MAPE value of 1.4116% also indicates that the LSTM model has high accuracy, with an average prediction error of less than 2%. MAPE expresses prediction error as a percentage of the actual value, making it useful for comparing different models and problems. A low MAPE indicates that the LSTM model's predictions are very close to actual values, highlighting its high accuracy.

Thus, the LSTM model, specifically designed to handle time series data and capable of retaining long-term information, is well-suited for time series forecasting or cyclic data problems. Its gating mechanisms help mitigate the vanishing gradient problem, enabling better learning over long periods. However, in practice, LSTM model may require more computational resources and longer training times compared to simpler models, especially with large or complex datasets. LSTM often demands extensive parameter tuning and can be challenging to optimize hyperparameters. For this reason, in this study, we employed an LSTM model with a single layer and a minimal number of parameters, making it easier to implement on processors. Despite a configuration with a hidden size of 96 and just one layer, the LSTM model

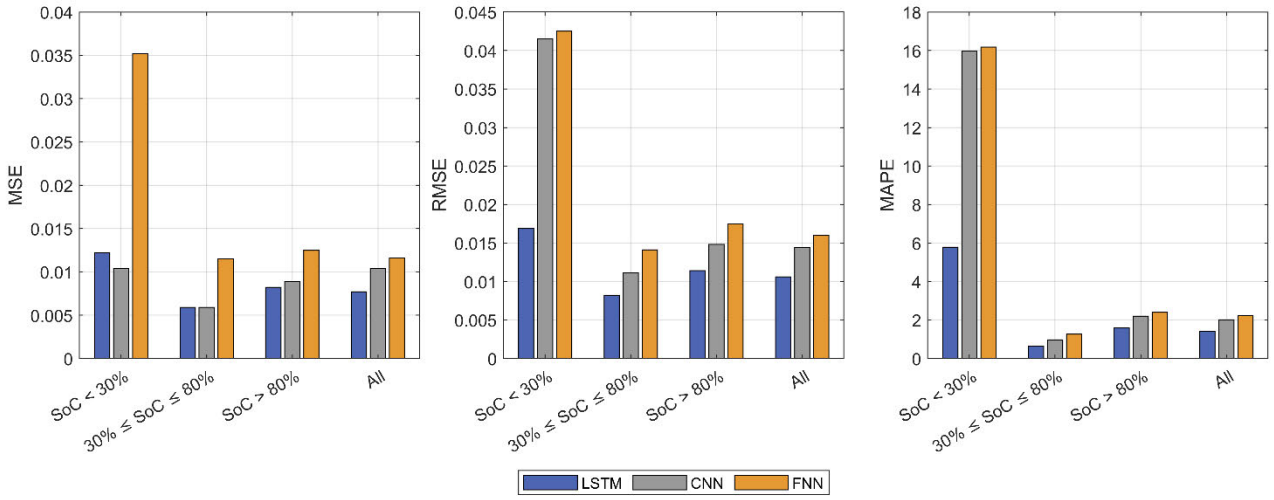

FIGURE 10. Comparison of error assessment criteria for the LSTM, FNN, and CNN estimation models.

achieved estimation accuracy metrics consistent with those reported in the literature [23].

V. CONCLUSION

This paper presented a method for estimating the SoC of battery packs in electric vehicles using LSTM networks and a modified Adam training algorithm. The proposed method utilized features extracted from one month of operational data, including vehicle speed, bus voltage, pack current, motor speed, motor temperature, motor control voltage, and average cell voltage and temperature. The collected data were normalized and fed into an LSTM network with a hidden size of 96 and a single layer. The modified Adam algorithm was employed to enhance the learning process and improve the accuracy of SoC estimation. The performance of the LSTM model was evaluated using three key error metrics: MSE, RMSE, and MAPE. The results were assessed across four SoC ranges: SoC $<$ 30%, 30% $\leq$ SoC $\leq$ 80%, SoC $>$ 80%, and 0% $\leq$ SoC $\leq$ 100%. Comparisons with CNN and FNN models demonstrated that the LSTM consistently achieved lower error values across most SoC ranges, highlighting its superior predictive capability. Future work will focus on further optimizing the LSTM model's architecture and exploring advanced training algorithms to improve its performance in regions where it underperformed. Additionally, the method could be extended to handle a wider variety of operational conditions and to incorporate real-time processing capabilities, enhancing the practicality of the approach.

APPENDIX

The tables of weight matrices of the input gate, output gate, and forget gate of LSTM are available at this link below: https://drive.google.com/file/d/1xbJRztA9pn3sMcVhVcOW4t8L1RJKHoQ4/view?usp = sharing

REFERENCES

[1] N. Sulaiman, M. A. Hannan, A. Mohamed, E. H. Majlan, and W. R. W. Daud, "A review on energy management system for fuel cell hybrid electric vehicle: Issues and challenges," *Renew. Sustain. Energy Rev.*, vol. 52, pp. 802–814, Dec. 2015.

[2] B. Zhang, N. Niu, H. Li, Z. Wang, and W. He, "Could fast battery charging effectively mitigate range anxiety in electric vehicle usage?Evidence from large-scale data on travel and charging in Beijing," *Transp. Res. D, Transp. Environ.*, vol. 95, Jun. 2021, Art. no. 102840.

[3] T.-N. Kröger, P. Harte, S. Klein, T. Beuse, M. Börner, M. Winter, S. Nowak, and S. Wiemers-Meyer, "Direct investigation of the interparticle-based state-of-charge distribution of polycrystalline NMC532 in lithium ion batteries by classification-single-particle-ICP-OES," *J. Power Sour.*, vol. 527, Apr. 2022, Art. no. 231204.

[4] M. Mokhtari, G. B. Gharehpetian, and S. M. Agah, "Distributed energy resources," in *Distributed Generation Systems: Design, Operation and Grid Integration*. Oxford, U.K.: Butterworth-Heinemann, 2017, pp. 1–19.

[5] S.-L. Wang, C. Fernandez, C.-Y. Zou, C.-M. Yu, X.-X. Li, S.-J. Pei, and W. Xie, "Open circuit voltage and state of charge relationship functional optimization for the working state monitoring of the aerial lithium-ion battery pack," *J. Cleaner Prod.*, vol. 198, pp. 1090–1104, Oct. 2018.

[6] K. S. Ng, C.-S. Moo, Y.-P. Chen, and Y.-C. Hsieh, "Enhanced Coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries," *Appl. Energy*, vol. 86, no. 9, pp. 1506–1511, Sep. 2009.

[7] M. Li, "Li-ion dynamics and state of charge estimation," *Renew. Energy*, vol. 100, pp. 44–52, Jan. 2017.

[8] R. Xiong, L. Li, Q. Yu, Q. Jin, and R. Yang, "A set membership theory based parameter and state of charge co-estimation method for all-climate batteries," *J. Cleaner Prod.*, vol. 249, Mar. 2020, Art. no. 119380.

[9] C. N. Van and T. N. Vinh, "State of charge estimation for lithium-ion batteries connected in series using two sigma point Kalman filters," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 2, p. 1334, Apr. 2022.

[10] J. Lee, O. Nam, and B. H. Cho, "Li-ion battery SOC estimation method based on the reduced order extended Kalman filtering," *J. Power Sources*, vol. 174, no. 1, pp. 9–15, Nov. 2007.

[11] C. N. Van and D. T. Quang, "State of charge estimation for lithium-ion battery cell considering the influence of aging parameters and operating temperature," *Int. J. Electrochem. Sci.*, vol. 18, no. 12, Dec. 2023, Art. no. 100379.

[12] E. Chemali, P. J. Kollmeyer, M. Preindl, and A. Emadi, "State-of-charge estimation of Li-ion batteries using deep neural networks: A machine learning approach," *J. Power Sources*, vol. 400, pp. 242–255, Oct. 2018.

[13] T. Hansen and C.-J. Wang, "Support vector-based battery state of charge estimator," *J. Power Sources*, vol. 141, no. 2, pp. 351–358, 2005.

[14] J. C. Á. Antón, P. J. G. Nieto, C. B. Viejo, and J. A. V. Vilán, "Support vector machines used to estimate the battery state of charge," *IEEE Trans. Power Electron.*, vol. 28, no. 12, pp. 5919–5926, Dec. 2013.
[15] B. Cheng, Y. Zhou, J. Zhang, J. Wang, and B. Cao, "Ni–MH batteries state-of-charge prediction based on immune evolutionary network," *Energy Convers. Manage.*, vol. 50, no. 12, pp. 3078–3086, Dec. 2009.
[16] S. Bockrath, V. Lorentz, and M. Pruckner, "State of health estimation of lithium-ion batteries with a temporal convolutional neural network using partial load profiles," *Appl. Energy*, vol. 329, Jan. 2023, Art. no. 120307.
[17] Z. Zhang, M. Xu, L. Ma, and B. Yu, "A state-of-charge estimation method based on bidirectional LSTM networks for lithium-ion batteries," in *Proc. 16th Int. Conf. Control, Autom., Robot. Vis. (ICARCV)*, Dec. 2020, pp. 211–216.
[18] X. Fan, W. Zhang, C. Zhang, A. Chen, and F. An, "SOC estimation of Li-ion battery using convolutional neural network with U-Net architecture," *Energy*, vol. 256, Oct. 2022, Art. no. 124612.
[19] Z. Sherkatghanad, A. Ghazanfari, and V. Makarenkov, "A self-attention-based CNN-Bi-LSTM model for accurate state-of-charge estimation of lithium-ion batteries," *J. Energy Storage*, vol. 88, May 2024, Art. no. 111524.
[20] D.-W. Chung, J.-H. Ko, and K.-Y. Yoon, "State-of-Charge estimation of lithium-ion batteries using LSTM deep learning method," *J. Electr. Eng. Technol.*, vol. 17, no. 3, pp. 1931–1945, May 2022.
[21] Y. Tian, R. Lai, X. Li, L. Xiang, and J. Tian, "A combined method for state-of-charge estimation for lithium-ion batteries using a long short-term memory network and an adaptive cubature Kalman filter," *Appl. Energy*, vol. 265, May 2020, Art. no. 114789.
[22] F. Yang, S. Zhang, W. Li, and Q. Miao, "State-of-charge estimation of lithium-ion batteries using LSTM and UKF," *Energy*, vol. 201, Jun. 2020, Art. no. 117664.
[23] S. E. Fallah, J. Kharbach, Z. Hammouch, A. Rezzouk, and M. O. Jamil, "State of charge estimation of an electric vehicle's battery using deep neural networks: Simulation and experimental results," *J. Energy Storage*, vol. 62, Jun. 2023, Art. no. 106904.

CHI NGUYEN VAN received the B.S. degree in instrument and control engineering from Thai Nguyen University, Vietnam, in 2000, and the M.S. and Ph.D. degrees in automation and control engineering from Hanoi University of Technology and Science, Hanoi, Vietnam, in 2006 and 2012, respectively. He is currently working as the Head of the Faculty of Electronic Engineering, Thai Nguyen University of Technology (TNUT), Vietnam, and a Researcher with the Dynamic and Control Laboratory. His research interests include nonlinear control, adaptive control, optimal control, and system identification for energy storage system using lithium-ion batteries and dynamic process systems. He serves as a technical reviewer for several prestigious journals and international conferences.

MINH DUC NGO received the B.S. and M.S. degrees in control engineering and automation from Thai Nguyen University of Technology (TNUT), Vietnam, in 2005 and 2009, respectively, the Ph.D. degree from the Department of Electrical Engineering, Chonnam National University, in 2015, and the Ph.D. degree in 2019. He is currently working as a Lecturer with the Department of Electrical Engineering and the Director of the International Training Cooperation Center, TNUT. His research interests include machine learning, microgrids, and automation systems.

CUONG DUONG DUC received the B.S. degree in data science from the VNU University of Science, Hanoi, Vietnam, in 2024. Since 2024, he has been a Lecturer with the Department of Control and Measurement. He has co-authored more than ten articles. His research interests include sensors, data science, and artificial intelligence.

LE QUANG THAO received the B.S., M.S., and Ph.D. degrees in physics from the VNU-University of Science, Hanoi, Vietnam, in 2005, 2009, and 2015, respectively.

Since 2007, he has been a Lecturer with the Faculty of Physics. He has authored or co-authored five course books, more than 35 articles, more than 30 inventions, and holds three patents. His research interests include sensors, embedded systems, and artificial intelligence. He serves as an Associate Editor for Information Dynamics and Applications.

SEON-JU AHN (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Seoul National University, Seoul, South Korea, in 2002, 2004, and 2009, respectively. He was a Postdoctoral Researcher at Myongji University, Seoul. He worked at the FREEDM Systems Center, North Carolina State University, Raleigh, NC, USA. He is currently working as a Professor with Chonnam National University, Gwangju, South Korea. His current research interests include power quality, distributed energy resources, energy storage, microgrids, smart grids, and real-time simulation.