



Article

# Data-Driven Algorithm Based on Energy Consumption Estimation for Electric Bus

Xinxin Zhao \* , Ming Zhang and Guangyu Xue

Department of Vehicle Engineering, School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China; mingzhang126@163.com (M.Z.); 41804218@xs.ustb.edu.cn (G.X.)

\* Correspondence: xinxinzhao@ustb.edu.cn

**Abstract:** The accurate estimation of battery state of charge (SOC) for modern electric vehicles is crucial for the range and performance of electric vehicles. This paper focuses on the historical driving data of electric buses and focuses on the extraction of driving condition feature parameters and data preprocessing. By selecting relevant parameters, a set of characteristic parameters for specific driving conditions is established, a process of constructing a battery SOC prediction model based on a Long short-term memory (LSTM) network is proposed, and different hyperparameters of the model are identified and adjusted to improve the accuracy of the prediction results. The results show that the prediction results can reach 1.9875% Root Mean Square Error (RMSE) and 1.7573% Mean Absolute Error (MAE) after choosing appropriate hyperparameters; this approach is expected to improve the performance of battery management systems and battery utilization efficiency in the field of electric vehicles.

**Keywords:** SOC; long short-term memory; data mining; hyperparameter tuning



**Citation:** Zhao, X.; Zhang, M.; Xue, G. Data-Driven Algorithm Based on Energy Consumption Estimation for Electric Bus. *World Electr. Veh. J.* **2023**, *14*, 329. <https://doi.org/10.3390/wevj14120329>

Academic Editor: Joeri Van Mierlo

Received: 27 October 2023

Revised: 22 November 2023

Accepted: 22 November 2023

Published: 29 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The battery management system (BMS) for electric vehicles is one of the key technologies used to monitor and effectively manage electric vehicle batteries, thus enabling electric vehicles to improve battery energy efficiency [1]. The main role of a BMS is to manage and maintain the battery intelligently, monitor the battery condition, and stop the battery from overcharging and overdischarging to cause an irreversible negative impact on the battery, thus achieving the effect of extending the cycle life of the battery, ensuring the safety of battery use and guiding the user to make the right choice. The estimation of battery SOC is an integral part of the BMS [2]. Accurate estimation of the battery SOC in real time can quantify the remaining usable battery power and provide a basis for drivers to plan their trips, prevent shortening of the battery life, and improve the battery energy usage, which is of great importance.

Electric buses have positive significance in terms of environmental protection, energy savings, sustainability, intelligence, economy, enhancing the image of public transportation, and adapting to urban development [3]. Electric buses use electricity as a power source, and compared to traditional fuel buses, they can significantly reduce exhaust emissions and noise pollution, which has a positive impact on improving the urban environment and air quality [4]. In addition, the economic benefits of electric buses are more pronounced as they also have lower maintenance costs. The energy efficiency of electric buses is higher, and the use of electric energy is much more efficient than fuel, which can reduce energy consumption. Electric bus operating conditions and battery operating conditions are closely related, and it is very important to predict the SOC of power batteries for electric buses, which not only helps them rationalize charging time and improve operational efficiency but also detects battery problems in time and avoids potential operational interruptions.

At present, there are three main methods for electric vehicle power battery SOC prediction: the first is the traditional estimation method, which uses the relationship curve

between voltage and current and SOC value during charging and discharging to calculate the battery SOC [5–7]; the second is the adaptive filtering method, which estimates the SOC by building a battery model to obtain the observation equation [8–10]; and the third is the intelligent estimation method, which uses machine learning, neural networks, and other methods for estimation [11]. The above three methods have their own advantages and disadvantages. Although the traditional estimation method is simple in principle and easy to implement, the accuracy is low and easy to deviate [12], so it is difficult to be directly applied to the actual estimation and is generally used in combination with other methods. The adaptive filtering method is more robust but requires the establishment of an accurate battery model [13]. The intelligent estimation method does not need to understand the internal parameter changes of the battery, but it requires a large amount of data as the basis to establish a training model; the more accurate the data and the larger the training volume, the better the training effect, but the training process will be more difficult to implement [14].

Electric buses require complex energy consumption estimation methods, which consider various parameters. Models like the Trip Energy Consumption (TEC) model [15] incorporate factors such as traction, battery thermal management, air conditioning, ambient temperature, vehicle weight, distance, and travel time. A physical model considering speed, acceleration, and passenger count has been developed [16], alongside a modified Kalman filter method for energy estimation [17]. Corinaldesi et al. proposed a linear optimization model to guide charging strategies. Additionally, machine learning models have been employed for estimating the instantaneous and cumulative energy consumption of electric buses [18].

Long short-term memory (LSTM) is an intelligent estimation method that extracts the information from the data to achieve its purpose. It is a transformation neural network, and its statistical prediction model is more powerful in terms of nonlinear mapping ability and self-learning ability [19]. The LSTM model is not only limited to solving single-input single-output problems. Because of its powerful learning capability, the number of inputs and outputs can be freely chosen; single-input multiple-output, multiple-input single-output, and multiple-input multiple-output mappings can also be fitted; and different model structures can be chosen according to different research requirements. The SOC of an electric bus is affected by multiple parameters so the overall model is a multiple-input single-output model; therefore, the LSTM model is well suited for SOC prediction. In this paper, based on the operation data of electric buses, a data-driven energy consumption prediction method based on LSTM is proposed. Compared with other methods, the model structure is not complicated, the prediction accuracy is high, and the training process is relatively simple and convenient, which is of significance in guiding the overall scheduling of electric buses.

The remainder of this paper is organized as follows: Section 2 is the historical bus driving data mining, Section 3 is the driving condition feature parameter extraction based on the data mining results, Section 4 introduces the LSTM model and makes the battery SOC prediction model based on the LSTM, the results are analyzed in Section 5, and finally, conclusions and future research directions are provided in Section 6.

## 2. Electric Bus Historical Driving Data Mining

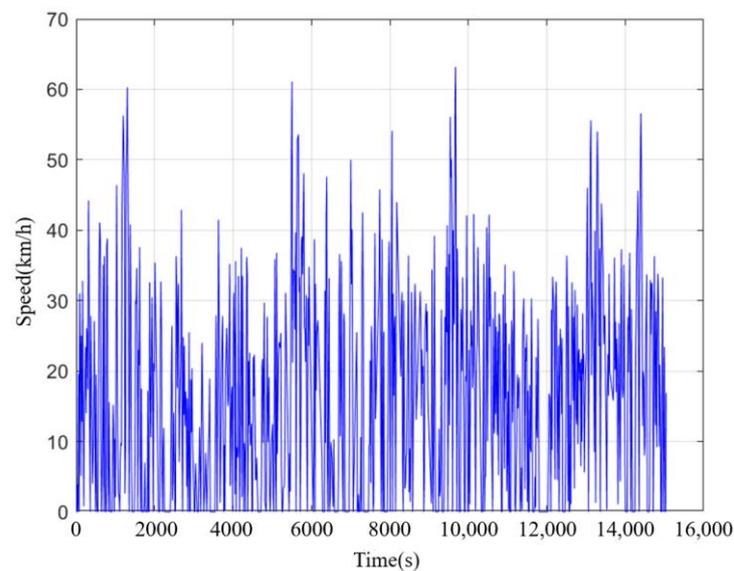
### 2.1. Electric Bus Driving Condition Analysis

The history of the bus travel points, after looking them up on Baidu Maps, revealed the car travel data for the Beijing 51 road bus history, the charging location for the Beijing world flower Wonderland Park, the line of the bus throughout a total of 29 stations, and revealed that the driving road conditions are urban road conditions in Figure 1.



**Figure 1.** Route 51 bus route map.

The scatter plot of the variation in vehicle speed with time for the battery SOC down cycle is shown in Figure 2, which shows that the bus line operation state has a strong regularity, and its operation time and operation conditions have certain rules that must be followed. Therefore, this paper starts by analyzing the characteristics of bus routes and selects the corresponding quantitative indexes and uses them as characteristic parameters to lay the foundation for the subsequent prediction research.



**Figure 2.** Variation in bus speed with time during the bus journey.

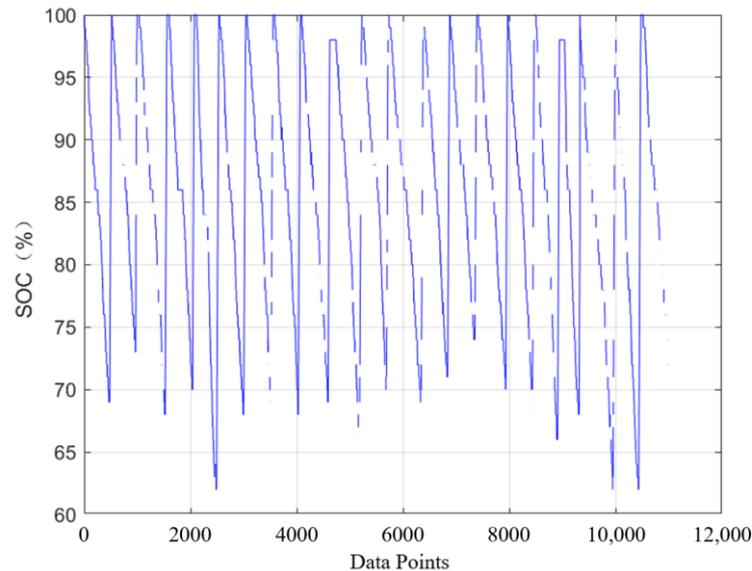
## 2.2. Data Preprocessing

Data mining is a method that uses algorithms to extract knowledge and information hidden in the original data that people cannot see directly, but are valuable to the subsequent research content, from a huge amount of actual real data with no obvious regular features [20].

During the driving process of the electric bus, the vehicle terminal reads the vehicle bus data in real time and collects data through a mobile network according to a certain period to transmit the data to a remote management platform. During the collection process, it is easily affected by vehicle start/stop, the surrounding environment, abnormal work

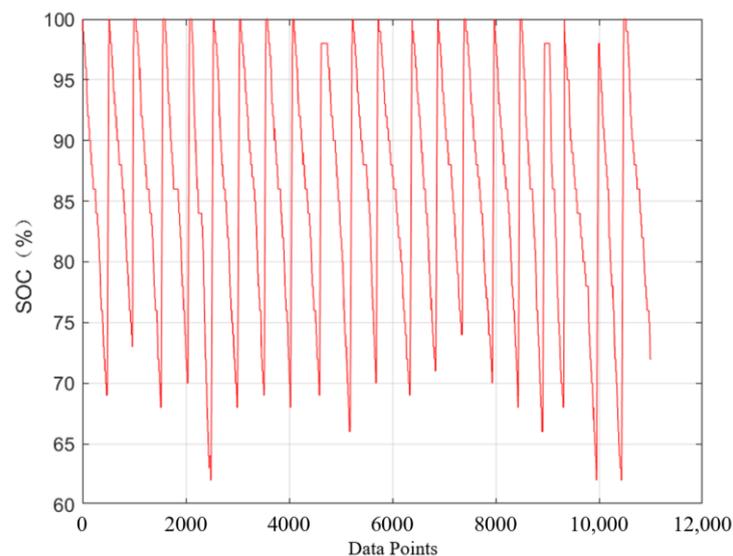
of vehicle terminal, etc., which generate abnormal data. To ensure the accuracy of SOC prediction, abnormal data must be rejected.

Eleven thousand rows of data were continuously extracted from the data set and plotted using Matlab 2018a software to create a scatter plot of the battery SOC values over time, as shown in Figure 3.



**Figure 3.** Scatter plot of SOC without removing abnormal data.

From the graph, it can be found that there are many blank anomalous data points in the SOC falling segment. And the data of the SOC rising indicate that the data of stopping the charging process are collected. For this paper, because only the energy consumption prediction problem is studied, the SOC rising section data and the anomalous data where the vehicle speed jumps to 0 should be excluded. The new data set after the exclusion is re-plotted using Matlab 2018a software for the SOC scatter plot as shown in Figure 4.



**Figure 4.** Scatter plot of SOC with abnormal data removed.

### 3. Driving Condition Feature Parameter Extraction

#### 3.1. Analysis of Vehicle Energy Consumption Characteristics and Driving Conditions

The bus driving data parameters obtained from the collection are numerous and contain 40 different parameters such as time, speed, charging status, running status, gear, drive motor speed, etc., and the accuracy is generally low. Because some of the data are not related to battery power estimation, 11 parameters in Table 1 are first selected to establish the set of driving condition characteristic parameters for bus line characteristics, and the specific parameters, units, calculation methods, and other information are shown in Table 1.

**Table 1.** Definition and calculation method of working condition characteristic parameters based on bus route characteristics.

| Number | Symbol       | Unit                           | Implied Meaning                             | Formula   |
|--------|--------------|--------------------------------|---|---|
| 1      | $v_m$        | km/h                           | Average speed                               | $v_m = \frac{1}{k} \sum_{i=1}^k v_i$                      |
| 2      | $a_{m\_p}$   | m/s <sup>2</sup>               | Average acceleration                        | $a_{m\_p} = \frac{1}{k} \sum_{i=1}^k a_i, a_i > 0$        |
| 3      | $a_{m\_n}$   | m/s <sup>2</sup>               | Average deceleration                        | $a_{m\_n} = \frac{1}{k} \sum_{i=1}^k a_i, a_i < 0$        |
| 4      | $v_{std}$    | km/h                           | Standard deviation in vehicle speed         | $v_{std} = \sqrt{\frac{1}{k} \sum_{i=1}^k (v_i - v_m)^2}$ |
| 5      | $f_{va}$     | m <sup>2</sup> /s <sup>3</sup> | Velocity times acceleration variance        | $f_{va} = \frac{1}{n} \sum_{i=1}^n (va_i - va_m)^2$       |
| 6      | $va_{avg}$   | m <sup>2</sup> /s <sup>3</sup> | Velocity multiplied by average acceleration | $va_{avg} = \frac{1}{k} \sum_{i=1}^k av_i$                |
| 7      | $a_{max}$    | m/s <sup>2</sup>               | Maximum acceleration                        | $a_{max} = \{a_1, a_2 \dots a_T\}_{max}$<br>$a > 0$       |
| 8      | $a_{d\_max}$ | m/s <sup>2</sup>               | Maximum deceleration                        | $a_{d\_max} = \{a_1, a_2 \dots a_T\}_{max}$<br>$a < 0$    |
| 9      | $a_{std}$    | m/s <sup>2</sup>               | Acceleration standard deviation             | $a_{std} = \sqrt{\frac{1}{k} \sum_{i=1}^k (a_i - a_m)^2}$ |
| 10     | $r_{avg}$    | r/min                          | Average motor speed                         | $r_{avg} = \frac{1}{k} \sum_{i=1}^k r_i$                  |
| 11     | $T_{avg}$    | N·m                            | Average motor torque                        | $T_{avg} = \frac{1}{k} \sum_{i=1}^k T_i$                  |

In the driving condition, the specific energy is often used as a measure of the energy consumed by the car under different driving conditions [21], and the specific energy can generally be defined as the ratio of the energy used to drive the wheels of the car to run the car over a certain distance. According to the definition, the formula for calculating specific energy is:

$$\bar{F}_{trac} = \frac{1}{x_t} \int_{t \in T_{trac}} F(t)v(t)dt \quad (1)$$

where  $F(t)$  is expressed as the longitudinal driving force on the dynamics of the wheels, and  $x_t$  is the total vehicle miles traveled in time  $t$ .

Considering that urban roads consist of mostly gentle sections, the following expressions are available:

$$\bar{F}_{trac} = \bar{F}_{air} + \bar{F}_r + \bar{F}_a \quad (2)$$

where  $\bar{F}_{air}$  is the specific energy to overcome air resistance,  $\bar{F}_r$  is the specific energy to overcome rolling resistance, and  $\bar{F}_a$  is the specific energy to overcome acceleration resistance. The expressions of each specific energy on the right side of the equal sign are as follows:

$$\bar{F}_{air} = \frac{1}{x_t} \int_{t \in T_{trac}} \frac{1}{2} \rho_a C_d A_f v^3(t) dt \tag{3}$$

$$\bar{F}_r = \frac{1}{x_t} \int_{t \in T_{trac}} mg f_r v(t) dt \tag{4}$$

$$\bar{F}_a = \frac{1}{x_t} \int_{t \in T_{trac}} \theta ma(t) v(t) dt \tag{5}$$

From the equations in Table 1, it can be concluded that there is a relationship between specific energy and driving condition characteristic parameters. Therefore, an end-element linear regression model is established for analyzing the relationship between specific energy and the driving condition characteristic parameters of the bus line. The multivariate linear regression model between the vehicle specific energy  $\bar{F}_{trac}$  and the operating characteristic parameters  $x_1, x_2, x_3, \dots, x_{11}$  is shown in Equation (6):

$$\bar{F}_{trac} = \begin{bmatrix} \bar{F}_{trac} \\ \bar{F}_{trac} \\ \vdots \\ \bar{F}_{trac} \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,11} \\ 1 & x_{2,1} & \cdots & x_{2,11} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{11,1} & \cdots & x_{11,11} \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{11} \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{11} \end{bmatrix} \tag{6}$$

In order to quantify the relationship between specific energy and characteristic covariates, the correlation coefficient between them needs to be calculated. The correlation coefficient matrix is first established as follows:

$$P = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn} \end{bmatrix} \tag{7}$$

where the denoted meaning of  $\rho_{ij}$  ( $i, j = 1, 2, 3 \dots n$ ) is the correlation coefficient between the dependent variable  $y$  and the independent variable  $x$ , and  $\rho_{ij} = \rho_{ji}$  ( $i, j = 1, 2, 3 \dots n$ ), which is calculated by the expression shown in Equation (8):

$$\rho_{y,x} = \frac{Cov(y,x)}{\sigma_y \sigma_x} \tag{8}$$

where  $Cov(y,x)$  expresses the meaning of the covariance between the dependent variable  $y$  and the independent variable  $x$ .  $\sigma_y$  is the variance of the dependent variable  $y$  and  $\sigma_x$  is the variance of the independent variable  $x$ .  $\rho_{y,x}$  is calculated between  $-1$  and  $1$  and  $\rho_{y,x} > 0$  means that  $x$  is positively correlated with  $y$ ;  $\rho_{y,x} < 0$  means that  $x$  is negatively correlated with  $y$ . When  $\rho_{y,x} = -1$ , it means that  $x$  and  $y$  are completely negatively correlated; when  $\rho_{y,x} = 0$ , it means that  $x$  and  $y$  are not correlated; when  $\rho_{y,x} = 1$ , it means that  $x$  and  $y$  are completely positively correlated; the absolute value of  $\rho_{y,x}$  is proportional to the closeness of  $1$  and the correlation between  $x$  and  $y$ .

### 3.2. Feature Parameter Extraction

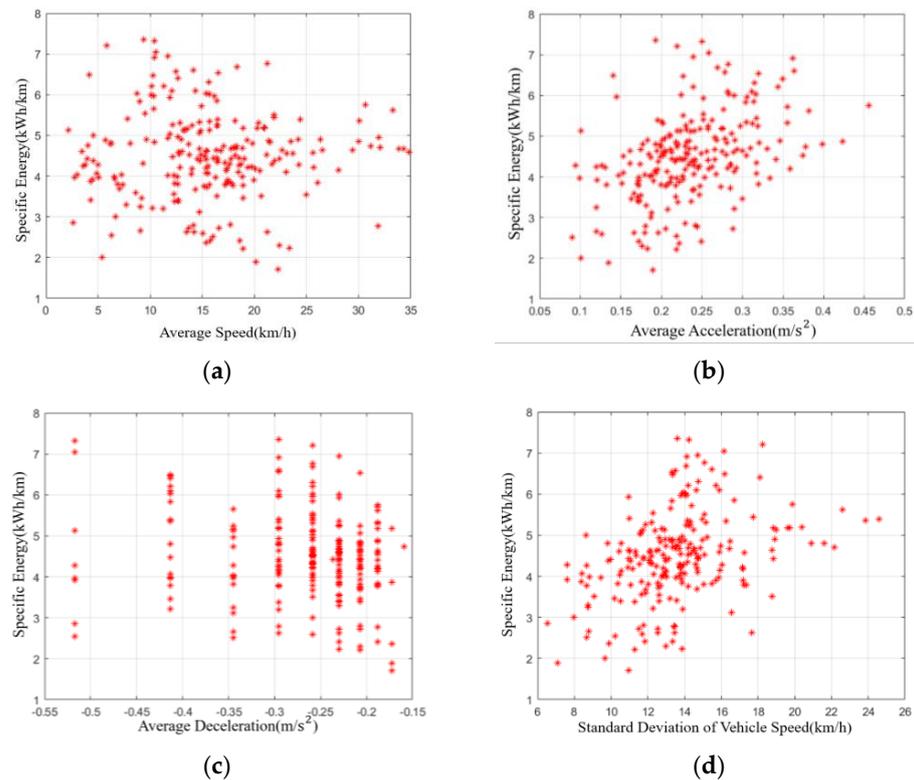
The model of the Route 51 pure electric bus is Yutong ZK6809BEVG. According to the calculation parameters mentioned in Equations (1)–(5), the relevant calculation parameters of this model bus are found on the official website and in the values of Table 2.

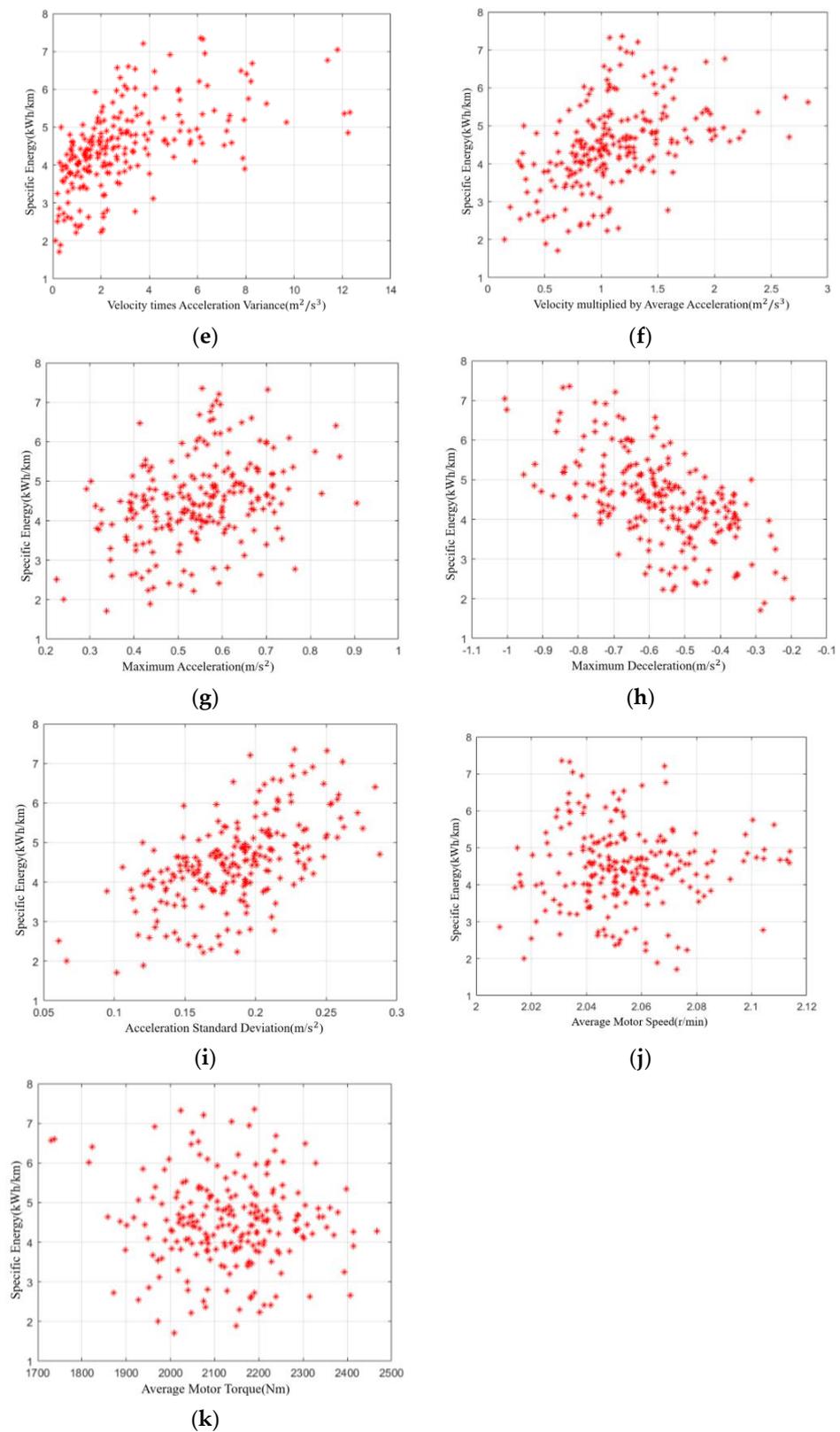
**Table 2.** The relevant parameters and values in the specific energy calculation formula.

| Parameters | Implied Meaning                 | Unit                       | Value  |
|------------|---------------------------------|----------------------------|--------|
| $A_f$      | Windward side (of an area)      | $m^2$                      | 9      |
| $\rho$     | Air density                     | $N \cdot s^2 \cdot m^{-4}$ | 1.2258 |
| $C_d$      | Atmospheric drag coefficient    | —                          | 0.6    |
| $m$        | Overall mass (of a vehicle)     | kg                         | 9000   |
| $g$        | Gravitational acceleration      | $m/s^2$                    | 9.8    |
| $\delta$   | Rotating mass conversion factor | —                          | 1.2    |
| $f$        | Rolling resistance coefficient  | —                          | 0.02   |

Twenty SOC decline cycles of the bus driving process were randomly selected and divided into groups of 20 rows of data for calculation using the formulae for the specific energy and driving condition characteristic parameters introduced in Section 3.1. The scatter diagram between the bus specific energy and driving condition characteristic parameters was drawn.

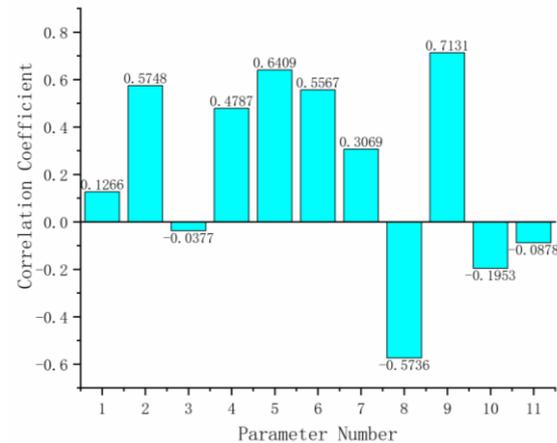
Using the correlation analysis model, the correlation between the specific energy and the characteristic covariates was further quantified, the correlation coefficients between each characteristic covariate and the specific energy were calculated, and the correlation coefficients between each covariate and the specific energy were plotted as shown in Figure 5.

**Figure 5.** Cont.



**Figure 5.** Scatter plot of specific energy and characteristic parameters of each driving condition: (a) Average Speed; (b) Average Acceleration; (c) Average Deceleration; (d) Speed Standard Deviation; (e) Velocity Times Acceleration Variance; (f) Velocity Times Acceleration Mean; (g) Maximum Acceleration; (h) Maximum Deceleration; (i) Acceleration Standard Deviation; (j) Average Speed of Drive Motor; (k) Average Torque of Drive motor.

From Figure 6, it can be seen that the correlation coefficients of the following parameters with absolute values outweighing 0.5 are highly correlated with specific energy: 2—average acceleration, 5—velocity times acceleration variance, 6—velocity times acceleration mean, 9—acceleration standard deviation is more positively correlated with specific energy; and 8—maximum deceleration is more negatively correlated with specific energy.

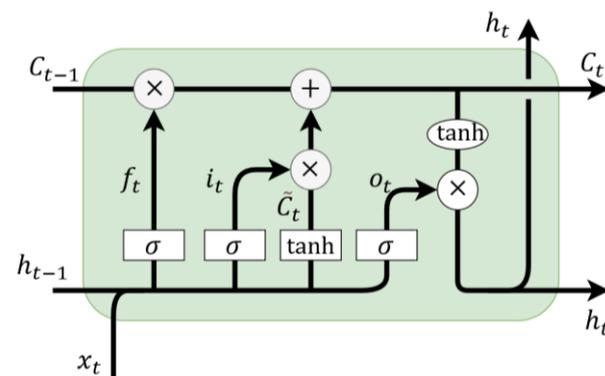


**Figure 6.** Correlation coefficient between each characteristic parameter and specific energy.

## 4. LSTM-Based Battery SOC Prediction

### 4.1. LSTM Model Structure

LSTM is a special recurrent neural network RNN model that can achieve controllable adjustment of the weight coefficients [22], and its cell structure is shown in Figure 7.



**Figure 7.** LSTM unit structure.

In Figure 7, the letters represent the following meanings:

$C_{t-1}$ —Cell unit state at moment  $t - 1$ ;

$C_t$ —Cell unit state at moment  $t$ ;

$\tilde{C}_t$ —The state of the cell unit at the current moment;

$h_{t-1}$ —The hidden state of the cell unit at the moment  $t - 1$ ;

$h_t$ —The hidden state of the cell unit at the moment  $t$ ;

$x_t$ —The input at time  $t$ .

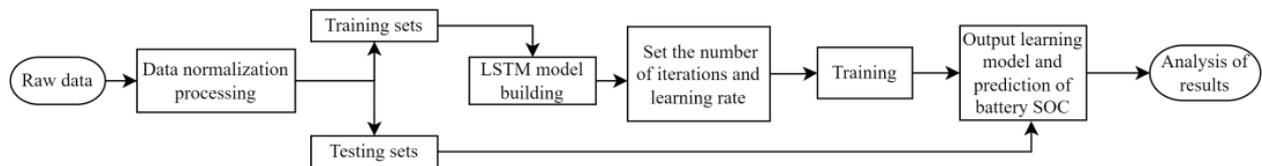
$f_t$  is the forgetting gate of the cell;  $i_t$  and  $\tilde{C}_t$  together form the input gate of the cell; and  $o_t$  and  $\tanh$  together form the output gate of the cell.  $\sigma$  and  $\tanh$  are the excitation functions.

From Figure 7, it can be seen that the LSTM structure is not a single neural network layer, but it is improved into four neural network layers with a tight connection and rich structure, which is conducive to improving the processing ability of information, controlling and protecting the cell state by introducing three gate structures: input gate, output gate,

and forgetting gate, making it possible to avoid the problem of long-term information dependence and remember useful information from very early moments without paying too much. Compared to the RNN model, this newly proposed model not only strengthens the processing ability of the network for information but also solves the problems such as abnormal gradient changes [23].

#### 4.2. Battery SOC Prediction Process

The general flow of this project for battery SOC prediction using the LSTM network is shown in Figure 8.



**Figure 8.** Battery SOC prediction model based on LSTM network.

In the original data, different types of feature parametric data have different value ranges, resulting in large differences in the weights assigned during the training process, resulting in the neural network model not converging [24]. To solve this problem, the individual variables must be standardized to eliminate the effect of inconsistency in order to enable the neural network to achieve better convergence and improve the training speed [25].

$$x^* = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (9)$$

where  $x_{max}$  is the maximum original data value;  $x_{min}$  is the minimum original data value.

Battery SOC prediction is a nonlinear regression process where the goal is to make the predicted value approximately the same as the actual value. The purpose of model training is to minimize the loss function. The mean squared error loss function is set as the loss function of the model in the prediction model established in this topic, and its calculation formula is shown in Equation (10):

$$loss = \frac{1}{2k} \sum_{i=1}^k (SOC_i - \hat{SOC}_i)^2 \quad (10)$$

where  $k$  is the amount of sample data;  $SOC_i$  is the true value of SOC; and  $\hat{SOC}_i$  is the estimated value of SOC obtained after model training.

In this paper, the Adam optimization algorithm is used, and this optimization method not only saves training time but also avoids the occurrence of the loss function falling into local optimum.

The steps of Adam's optimization algorithm are as follows.

Step 1: Calculating gradients:

$$g_t = \nabla_{\theta} f_t(\theta_{t-1}) \quad (11)$$

Step 2: Updating the first-order moment estimates:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (12)$$

In Equation (12),  $\beta_1$  is the exponential decay rate, and it generally takes  $\beta_1 = 0.9$ .

Step 3: Updating second-order moment estimation:

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (13)$$

In Equation (13),  $\beta_2$  is the exponential decay rate, which is generally taken as  $\beta_2 = 0.999$ .

Step 4: Correction of bias in first-order moment estimates:

$$\hat{m}_t = m_t / (1 - \beta_1^t) \quad (14)$$

Step 5: Correction of bias in second-order moment estimation:

$$\hat{v}_t = m_t / (1 - \beta_2^t) \quad (15)$$

Step 6: Updated parameters:

$$\theta_t = \theta_{t-1} - \alpha \cdot \hat{m}_t / (\sqrt{\hat{v}_t} - \epsilon) \quad (16)$$

In Equation (16),  $\alpha$  is the learning rate and  $\epsilon = 10^{-8}$ . It is seen that the first-order moments and second-order moments participate collaboratively in adjusting the parameter  $\theta$ .

### 4.3. Hyperparameter Selection

In the process of network model construction, parameters such as the number of hidden layers and the number of neurons in hidden layers are related to the network structure; parameters such as the dimensionality of input and output quantities and the time step are related to the data; and parameters such as the learning rate, training data batch size, and loss function are related to training learning. Changing the size of these parameters affects the accuracy of training and thus the final prediction results, so reasonable hyperparameters need to be selected.

#### (1) Number of hidden neurons

The number of neurons in the hidden layer is usually linearly related to the predictive ability of the network model and the complexity of the network training; the more hidden neurons, the better the learning ability of the model, but at the same time, it leads to a more complex model and a longer training time. The input layer of the LSTM model to be built in this topic is related to the dimensionality of the input data, while the output layer has only battery SOC values, hence the number of nodes in the output layer is one.

#### (2) Batch size

The batch size of training determines how much data are input to the network model for training each time. The larger the batch size, the slower the adjustment of the data weighting coefficients; and although the training time can be reduced, the training results will be relatively poor. Therefore, an appropriate batch size is needed to improve efficiency during training.

#### (3) Training cycle

The training period is the number of iterations of the data, and choosing a suitable training period is beneficial to obtain prediction results with high accuracy. If the training period is too long, it will lead to a prolonged training time of the model, the over-learning phenomenon, and the over-fitting phenomenon; this must be avoided in order to excessively reduce the error of the training set, not affect the generality of the test set, and avoid the deterioration of the results. According to the multiple training runs, it was found that the loss function changes very little after 6000 iterations; in order to improve efficiency, this paper uses 6000 as the model training period and increases the number of hidden layers to 8000 instead.

#### (4) Learning Rate

By controlling the size of the learning rate, we can control how fast or slow the gradient update is. Too large a learning rate will cause large fluctuations in the gradient update, thus making it impossible to find the global optimal solution and making the final prediction results produce large errors. The model built in this paper uses the Adam optimization

algorithm, which can adjust the learning rate size by itself, so it only needs to choose an initial learning rate, and the learning rate  $l_r = 0.01$  is chosen in this paper.

#### 4.4. Model Building

From the original data analysis, it can be concluded that vehicle speed is highly correlated with SOC and is easy to measure; battery voltage and current are correlated with battery status; drive motor speed is correlated with vehicle operation and battery power consumption; and SOC at the previous moment must have an impact on current SOC. Therefore, the vehicle speed, total voltage, total current, drive motor speed, and SOC value of the previous moment are selected as the input of the model, and the number of nodes in the input layer is five. The battery SOC value is taken as the output, and the number of nodes in the output layer is one.

Using PyCharm2021, we build a multiple-input single-output battery SOC prediction model under the Python 3.8 platform. The specific steps are as follows:

Step 1: The call model mainly uses library functions such as Pytorch and pandas.

Step 2: A long- and short-term memory network class is defined based on the input parameters, an LSTM network is constructed, and the hyperparameters of the model are initialized.

Step 3: The training set and test set data are imported, and the input quantity contains vehicle speed, voltage, motor speed, current, and the SOC value of the previous moment; the output quantity is SOC, and data normalization is performed.

Step 4: The loss function is defined, the loop is set, and training is ended when the loss function is less than  $10^{-3}$ , otherwise training continues for the set period.

Step 5: The data are converted into tensor-type data, the hyperparameters of the model are set, and training is started.

Step 6: After the training is completed, the test set data are put into the model for prediction, and the predicted values are plotted against the actual values.

Step 7: The model is trained and predicted several times by adjusting hyperparameters such as the number of hidden layers and the number of neuron nodes to compare the accuracy of the prediction results.

### 5. Analysis of Simulation Results

In order to make an intuitive evaluation of the prediction results, the mean absolute error and root mean square error of the predicted and actual values of SOC are selected as evaluation indicators in this paper, and the calculation formula is as follows:

$$MAE = \frac{1}{k} \sum_{i=1}^k |SOC_i - \hat{SOC}_i| \quad (17)$$

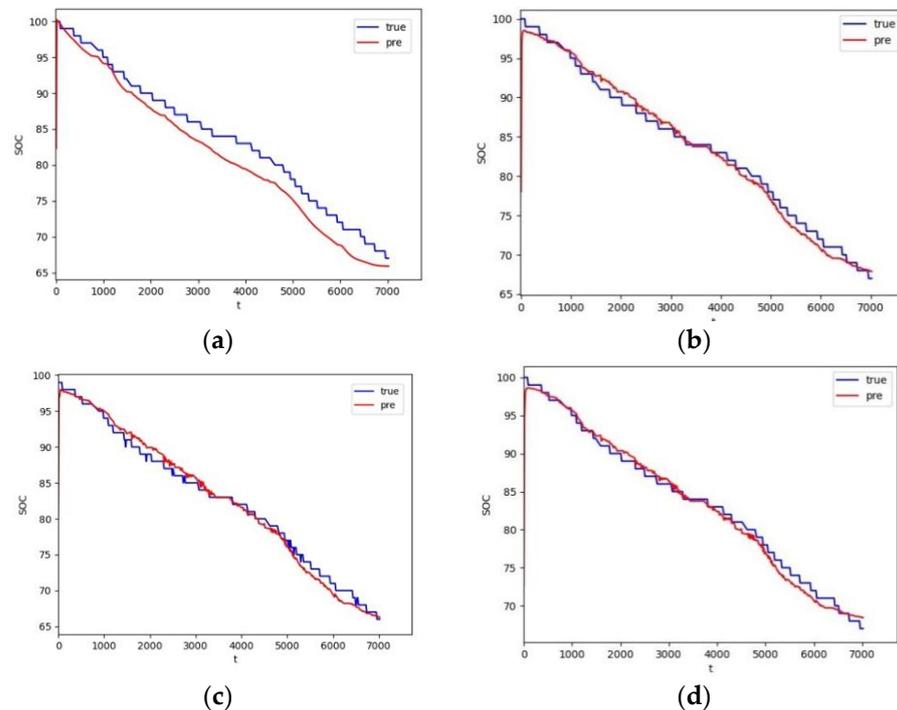
$$RMSE = \sqrt{\frac{1}{k} \sum_{i=1}^k (SOC_i - \hat{SOC}_i)^2} \quad (18)$$

Different prediction results are obtained by adjusting the hyperparameters of the model. The adjusted hyperparameters are shown in Table 3:

**Table 3.** Selection of model hyperparameters.

| Training Number | Number of Hidden Layers | Number of Neurons | Batch Size | Training Period | Learning Rate |
|-----------------|-------------------------|-------------------|------------|-----------------|---------------|
| 1               | 1                       | 32                | 5          | 6000            | 0.01          |
| 2               | 1                       | 48                | 1          | 6000            | 0.01          |
| 3               | 2                       | 48/48             | 5          | 8000            | 0.01          |
| 4               | 2                       | 48/48             | 1          | 8000            | 0.01          |

Corresponding to the parameter settings in Table 3 above, the predicted results for each number are plotted in Figure 9.



**Figure 9.** Graph of SOC prediction results with different hyperparameters selected: (a) No. 1 hyperparameter prediction results; (b) No. 2 hyperparameter prediction results; (c) No. 3 hyperparameter prediction results; (d) No. 4 hyperparameter prediction results.

In Figure 9, the blue line represents the actual battery SOC value and the red line represents the predicted SOC value. From the Figure 9, it can be found that the prediction model has a large deviation in the results in the early stage because the SOC decline cycle start data do not strictly start with the battery in a full-charge state, resulting in less accurate prediction results. Because the output SOC is a step-down type of data, the improper selection of hyperparameters may lead to a good fit of the training data but an overfitting during testing.

The errors of the prediction results are shown in Table 4 for different choices of hyperparameters, using Equations (17) and (18).

**Table 4.** Errors in prediction results for different hyperparameters.

| Training Number | RMSE (%) | MAE (%) |
|-----------------|----------|---------|
| 1               | 9.1374   | 7.6426  |
| 2               | 3.1023   | 2.5664  |
| 3               | 2.4835   | 2.0753  |
| 4               | 1.9875   | 1.7573  |

From Table 4 above, it can be seen that in model No. 1, the number of fewer neurons in one hidden layer is not completely learned for the training data, and the loss function is larger and cannot find the optimal solution, so the prediction results are poorer. In models 2 to 4, the loss function decreases by appropriately increasing the number of network layers and the number of neurons; and the prediction results are more accurate as the batch size decreases, but the efficiency decreases.

The results show that the prediction of battery SOC can be achieved using the constructed LSTM prediction model, and according to adjusting the hyperparameters of the

model, the prediction results of the model can be made more accurate and suitable for solving the SOC estimation problem.

## 6. Conclusions

In this paper, the collected electric bus driving data are processed to remove abnormal and useless data for the study based on the working condition characteristics. Only the SOC declining cycle data are used to find the correlation between the specific energy of the vehicle and other parameters and extract the feature parameters. Based on the LSTM network, the SOC prediction model is built and the prediction results with good accuracy are obtained by adjusting the model hyperparameters, which proves the practicality of LSTM for battery SOC estimation. Due to the limitation of the dataset in this paper, the training set is not rich enough in data. In the next research step, the training set data can be enriched to improve the accuracy of prediction results and add the single-input single-output LSTM network model, which takes the battery SOC value as the input and the driving mileage as the output of the model, to predict the driving mileage of the vehicle and act as an indication for the driver's travel. Subsequently, the charging and discharging frequency of multiple electric buses can be scheduled based on the present SOC prediction model to guide the charging, so that the buses can operate more efficiently and save operating costs.

**Author Contributions:** Conceptualization, X.Z.; Methodology, X.Z. and M.Z.; Software, G.X.; Validation, M.Z.; Formal analysis, X.Z.; Investigation, G.X.; Resources, M.Z.; Data curation, X.Z. and M.Z.; Writing—original draft preparation, X.Z. and M.Z.; Writing—review and editing, X.Z.; Visualization, M.Z.; Supervision, X.Z.; Project administration, X.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Natural Science Foundation of China (Grant No. 52275082) and the Fundamental Research Funds for the Central Universities QNXM20220029 and FRF-TP-20-037A2.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the Data Management Policy of National New Energy Vehicle Big Data Platform.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Uzair, M.; Abbas, G.; Hosain, S. Characteristics of Battery Management Systems of Electric Vehicles with Consideration of the Active and Passive Cell Balancing Process. *World Electr. Veh. J.* **2021**, *12*, 120. [[CrossRef](#)]
2. Kim, M.J.; Chae, S.H.; Moon, Y.K. Adaptive battery state-of-charge estimation method for electric vehicle battery management system. In Proceedings of the 2020 International SoC Design Conference (ISOCC), Yeosu, Republic of Korea, 21–24 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 288–289. [[CrossRef](#)]
3. Manzolli, J.A.; Trovao, J.P.; Antunes, C.H. A review of electric bus vehicles research topics—Methods and trends. *Renew. Sustain. Energy Rev.* **2022**, *159*, 112211. [[CrossRef](#)]
4. Li, Z.; Khajepour, A.; Song, J. A comprehensive review of the key technologies for pure electric vehicles. *Energy* **2019**, *182*, 824–839. [[CrossRef](#)]
5. Espedal, I.B.; Jinasena, A.; Burheim, O.S.; Lamb, J.J. Current Trends for State-of-Charge (SoC) Estimation in Lithium-Ion Battery Electric Vehicles. *Energies* **2021**, *14*, 3284. [[CrossRef](#)]
6. Li, R.; Xu, S.; Li, S.; Zhou, Y.; Zhou, K.; Liu, X.; Yao, J. State of charge prediction algorithm of lithium-ion battery based on PSO-SVR cross validation. *IEEE Access* **2020**, *8*, 10234–10242. [[CrossRef](#)]
7. Wang, Y.; Tian, J.; Sun, Z.; Wang, L.; Xu, R.; Li, M.; Chen, Z. A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems. *Renew. Sustain. Energy Rev.* **2020**, *131*, 110015. [[CrossRef](#)]
8. Xu, W.; Xu, J.; Yan, X. Lithium-ion battery state of charge and parameters joint estimation using cubature Kalman filter and particle filter. *J. Power Electron.* **2020**, *20*, 292–307. [[CrossRef](#)]
9. Xiong, R.; Zhang, Y.; He, H.; Zhou, X.; Pecht Michael, G. A Double-Scale, Particle-Filtering, Energy State Prediction Algorithm for Lithium-Ion Batteries. *IEEE Trans. Ind. Electron.* **2018**, *65*, 1526–1538. [[CrossRef](#)]
10. Xu, Y.; Chen, X.; Zhang, H.; Yang, F.; Tong, L.; Yang, Y.; Wang, Y. Online identification of battery model parameters and joint state of charge and state of health estimation using dual particle filter algorithms. *Int. J. Energy Res.* **2022**, *46*, 19615–19652. [[CrossRef](#)]
11. Luo, K.; Chen, X.; Zheng, H.; Shi, Z. A review of deep learning approach to predicting the state of health and state of charge of lithium-ion batteries. *J. Energy Chem.* **2022**, *74*, 159–173. [[CrossRef](#)]

12. Zhou, W.; Zheng, Y.; Pan, Z.; Lu, Q. Review on the battery model and SOC estimation method. *Processes* **2021**, *9*, 1685. [[CrossRef](#)]
13. Yu, Z.; Na, H.; Qi, L.; Li, R. SOC prediction of Volterra adaptive filter based on chaotic time series. *AIP Adv.* **2022**, *12*, 115113. [[CrossRef](#)]
14. Jiaqiang, E.; Zhang, B.; Zeng, Y.; Wen, M.; Wei, K.; Huang, Z.; Deng, Y. Effects analysis on active equalization control of lithium-ion batteries based on intelligent estimation of the state-of-charge. *Energy* **2022**, *238*, 121822. [[CrossRef](#)]
15. Ji, J.; Bie, Y.; Zeng, Z.; Wang, L. Trip energy consumption estimation for electric buses. *Commun. Transp. Res.* **2022**, *2*, 100069. [[CrossRef](#)]
16. Liu, Y.; Liang, H. A data-driven approach for electric bus energy consumption estimation. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 17027–17038. [[CrossRef](#)]
17. Corinaldesi, C.; Lettner, G.; Schwabeneder, D.; Ajanovic, A.; Auer, H. Impact of Different Charging Strategies for Electric Vehicles in an Austrian Office Site. *Energies* **2020**, *13*, 5858. [[CrossRef](#)]
18. Chen, Y.; Zhang, Y.; Sun, R. Data-driven estimation of energy consumption for electric bus under real-world driving conditions. *Transp. Res. Part D Transp. Environ.* **2021**, *98*, 102969. [[CrossRef](#)]
19. Staudemeyer, R.C.; Morris, E.R. Understanding LSTM—A tutorial into long short-term memory recurrent neural networks. *arXiv* **2019**, arXiv:1909.09586. [[CrossRef](#)]
20. Agarwal, S. Data mining: Data mining concepts and techniques. In Proceedings of the 2013 International Conference on Machine Intelligence and Research Advancement, Katra, India, 21–23 December 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 203–207. [[CrossRef](#)]
21. Yang, D.; Li, J.; Liu, C.; Xing, W.; Zhu, J. Design strategy and comprehensive performance assessment towards Zn anode for alkaline re-chargeable batteries. *J. Energy Chem.* **2023**, *82*, 122–138. [[CrossRef](#)]
22. Asok, R.; Michael, H.; Shashi, P.; Yiwei, F.U. Neural Probabilistic Forecasting of Symbolic Sequences with Long Short-Term Memory. *J. Dyn. Syst. Meas. Control.* **2018**, *140*, 084502. [[CrossRef](#)]
23. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* **2019**, *31*, 1235–1270. [[CrossRef](#)] [[PubMed](#)]
24. Lian, D.; Guijin, T. Low-light image enhancement algorithm using a residual network with semantic information. *J. China Univ. Posts Telecommun.* **2022**, *29*, 52–62+84.
25. Bai, Y.; Xiang, S.; Cheng, F.; Zhao, J. A dynamic-inner LSTM prediction method for key alarm variables forecasting in chemical process. *China J. Chem. Eng.* **2023**, *55*, 266–276. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.