energies

Article

Accurate State of Charge Estimation for Real-World Battery Systems Using a Novel Grid Search and Cross Validated Optimised LSTM Neural Network

Jichao Hong [1,2,3], Fengwei Liang [1,2,3], Xun Gong [1], Xiaoming Xu [1,2,3,*] and Quanqing Yu [4,*]

[1] State Key Laboratory of Automotive Simulation and Control, Jilin University, Changchun 130025, China
[2] School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China
[3] Shunde Innovation School, University of Science and Technology Beijing, Foshan 528399, China
[4] School of Automotive Engineering, Harbin University of Technology, Weihai 150006, China
* Correspondence: xuxiaoming3777@163.com (X.X.); qqyu@hit.edu.cn (Q.Y.)

Abstract: State of charge (SOC) is one of the most important parameters in battery management systems, and the accurate and stable estimation of battery SOC for real-world electric vehicles remains a great challenge. This paper proposes a long short-term memory network based on grid search and cross-validation optimisation to estimate the SOC of real-world battery systems. The real-world electric vehicle data are divided into parking charging, travel charging, and finish charging cases. Meanwhile, the parameters associated with the SOC estimation under each operating condition are extracted by the Pearson correlation analysis. Moreover, the hyperparameters of the long short-term memory network are optimised by grid search and cross-validation to improve the accuracy of the model estimation. Moreover, the gaussian noise algorithm is used for data augmentation to improve the accuracy and robustness of SOC estimation under the working conditions of the small dataset. The results indicate that the absolute error of SOC estimation is within 4% for the small dataset and within 2% for the large dataset. More importantly, the robustness and effectiveness of the proposed method are validated based on operational data from three different real-world electric vehicles, and the mean square error of SOC estimation does not exceed 0.006. This paper aims to provide guidance for the SOC estimation of real-world electric vehicles.

Keywords: electric vehicle; battery system; state of charge; grid search and cross-validation; long short-term memory

Citation: Hong, J.; Liang, F.; Gong, X.; Xu, X.; Yu, Q. Accurate State of Charge Estimation for Real-World Battery Systems Using a Novel Grid Search and Cross Validated Optimised LSTM Neural Network. *Energies* **2022**, *15*, 9654. https://doi.org/10.3390/en15249654

Academic Editor: Muhammad Aziz

Received: 1 December 2022
Accepted: 16 December 2022
Published: 19 December 2022

1. Introduction

With the energy crisis and environmental issues becoming increasingly prominent, new energy vehicles are receiving more and more attention. Electric vehicles (EVs) are gradually gaining ground due to their clean, efficient, and pollution-free characteristics [1–3]. However, it is important to have an accurate SOC estimation for EVs when they are driven under different operating conditions [4]. Battery SOC estimation is one of the key technologies for EVs, and its accuracy directly affects the EV energy management control strategy and EV performance. Meanwhile, it is also an important parameter in the battery management system (BMS). On the one hand, it provides the driver with important information about the driving range. On the other hand, it also provides an important basis for preventing overcharging and overdischarging of the battery [5]. However, due to the complex electrochemical properties of the battery, it exhibits a high degree of non-linearity during operation. Additionally, it is not possible to measure the battery SOC state parameters directly. Battery SOC estimation can only be made from externally measurable battery terminal voltages, charging and discharging currents, etc. In addition, the estimation process is susceptible to temperature, cycle time, discharge rate, voltage, and other factors, making it difficult to accurately estimate battery SOC in real time [6]. As discussed above, this paper presents grid search and cross-validation (GSCV)-based optimised LSTM neural network hyperparameters for real-world EV battery system SOC estimation. Therein, the grid search and cross validation are used to select the optimal hyperparameters of the LSTM neural network to improve the accuracy of SOC estimation. This optimization method is more accurate than manually adjusting the hyperparameters of the LSTM neural network. Currently, most SOC estimation methods are studied in a laboratory environment, and most have little reference for real-world battery SOC estimation due to the complex operating conditions of real-world EVs, large data sampling intervals, and low data accuracy. In this paper, the SOC estimation of the battery system was researched using real-world EV operation data, and the results show that the absolute error of the SOC estimation is less than 4%, and this paper aims to provide guidance on the theoretical implications for real-world EV battery SOC estimation.

1.1. Literature Review

There has been more research on battery SOC estimation. Based on the choice of battery SOC estimation methods, battery state estimation algorithms can be broadly classified into: direct measurement methods, SOC estimation methods based on black-box battery models, and SOC estimation methods based on state-space battery models.

The direct measurement method is based on battery voltage, current, internal resistance, impedance, and other reproducible parameter variables significantly correlated to the battery. These battery parameter variables should be relatively easily measurable in practice. Direct measurement methods include the ampere-hour (Ah) integration method, the open circuit voltage method, the internal resistance method, and special methods adapted for specific objects. The Ah method is still widely used due to limitations in BMS computing capabilities. However, the traditional Ah method is sensitive to external environmental influences and is not precise in its estimation. Introducing a capacity-integrated correction factor and building an adaptively improved Ah formula and a complete SOC estimation model can significantly improve the accuracy of SOC estimation [7]. The estimation error of the enhanced adaptive method is less than 2% under the combined working conditions of the two operating conditions. In contrast, the estimation error of the traditional approach is 5%~10%. Open circuit voltage (OCV) is also widely used for battery SOC estimation. Ren et al. used adaptive Kalman filtering for power cell SOC estimation based on OCV-SOC curves [8], and the results showed that the SOC estimation method was effective. In addition, a joint SOC and internal resistance-based estimation algorithm are used for practical studies [9]. The relationship between internal resistance and battery capacity is established by linear fitting, and the capacity converted by internal resistance is applied to SOC estimation, and the estimation results show that the method can effectively improve the accuracy of SOC estimation regardless of temperature changes and battery degradation.

Black-box model-based SOC estimation considers the battery as an unknown system, using the battery current, voltage, and temperature measured online as the model's input and the battery SOC as the model's output. It trains the input and output data through intelligent algorithms and establishes a relationship between the inputs and outputs. Black box battery models typically use neural networks, support vector machines, and deep learning to estimate battery SOC values based on input battery state parameters. Chemali et al. [10] proposed a method for estimating SOC using an LSTM network. The measured voltages, currents, and temperatures are fed directly into the created network, which learns the input time series and mapping between the target SOC. In addition, the support vector machine-based SOC state estimation method is used by mapping samples from the non-linear space of the battery to the linear space [11]. Experiments show that the joint colourless Kalman filter and support vector machine algorithm has higher accuracy for SOC estimation, while the tracking error is lower than 1%. Meanwhile, a deep neural network (DNN)-based SOC estimation method is proposed [12], using only 10 min of charging voltage and current data as input to estimate SOC. The method can estimate SOC quickly and accurately with an error range of less than 2.03%, and the SOC estimation method based on DNN can also be transferred to other different types of batteries for SOC estimation.

The state-space battery model is used for battery SOC estimation with filters or observers. The model inputs are battery current, voltage, and temperature parameters, and the output is the battery SOC. The equivalent circuit model, parameter identification, and SOC estimation observer are jointly involved in the state-space model study [13]. An appropriate model is a prerequisite for accurate SOC estimation [14]. A novel SOC estimation method combining an equivalent circuit model and an adaptive unscented Kalman filter is carried out, which is demonstrated to have an accuracy advantage in battery system SOC estimation by comparing the voltage response curves of different n-Resistor-Capacitor (RC) models and common equivalent circuit models. In addition, SOC estimation combined with parameter identification also has high accuracy. A variable-length block-wise least-squares SOC estimation algorithm is proposed [15], which considers the local linear SOC and open-circuit voltage relationship in operation. The proposed algorithm can accurately estimate the parameters and track the changes in the parameters. After estimating the battery parameters, the SOC is calculated directly from the combination of the estimated parameters and the OCV-SOC relationship. The algorithm is validated by experiments on real data obtained from laboratory tests. In addition, observer-based SOC estimation algorithms are also widely used. A SOC estimation method based on the descending order of unknown input observers is proposed for determining SOC using the OCV-SOC characteristic curve [16]. A set of Sylvester constraint parameters guaranteed the unbiasedness of the estimation error. Simulation and experiment demonstrate that the proposed observer has high accuracy in SOC estimation.

As mentioned above, although there are many methods to conduct SOC estimation for batteries, most of these methods have been studied in laboratory environments. Real-world EVs operate under complex conditions, so SOC estimation should be conducted for real-world EV battery systems, and the following are the challenges of conducting SOC estimation for real-world EVs:

(1) Large sampling interval and low data accuracy for real-world EV data.
(2) Model-based approaches rely on complex mathematical models and have limited estimation accuracy.
(3) Most machine learning-based SOC estimation methods rely on large amounts of offline data and are still in the laboratory stage.

1.2. Contributions of this Work

This paper attempts to make several original contributions and improvements to the current research, as shown in the following:

(1) The EV operating conditions are divided into parking charging, travel charging, and finish charging. The relevant parameters are extracted under each operating condition for SOC estimation using Pearson analysis.
(2) Optimizing LSTM neural network hyperparameters based on grid search and cross validation to improve the accuracy of the proposed method and the absolute error of SOC estimation within 4% for real-world EVs.
(3) The method's accuracy is verified by using Gaussian noise to expand the data for working conditions with small data, and the robustness of the method was verified by operating data of different EVs.

1.3. Organization of the Paper

The remainder of this paper is structured as follows: Section 2 describes the data details and pre-processing content. In Section 3, the network architecture of the LSTM, the GSCV algorithm, and the Gaussian noise algorithm content are described. Section 4 carries out the analysis of the results, and Section 5 draws conclusions.

2. Data Description and Pre-Processing

2.1. Data Description

The data used in this study is the driving data of an EV from July to December with a data sampling interval of 10 s. It contains more than 15 data items, such as the total voltage, current, and charging mode of the battery system. For this paper, the data is divided into parked charging, travel charging, and end of charge, depending on the SOC of the battery system.

2.2. Data Preprocessing

For real-world battery systems, the nonlinear coupling of multiple parameters is evident, making it difficult to identify the correlation between different parameters and SOC. Data irrelevant to the battery system were removed from the original data. For different operating conditions, parameters related to the variation of the battery SOC were extracted based on Pearson correlation analysis. The data were also grouped and cleaned for different operating conditions. Finally, due to the small amount of data in the charging condition, gaussian noise enhancement data was carried out for this condition.

3. Methodology

This section presents the architecture of the LSTM and the main elements of the GSCV method. In this study, the SOC estimation parameters corresponding to the three working conditions are extracted by Pearson analysis as the input of the LSTM neural network. The model accuracy is also improved through the GSCV optimisation algorithm, which optimises the hyperparameters of the LSTM neural network. The flow of this research is shown in Figure 1.

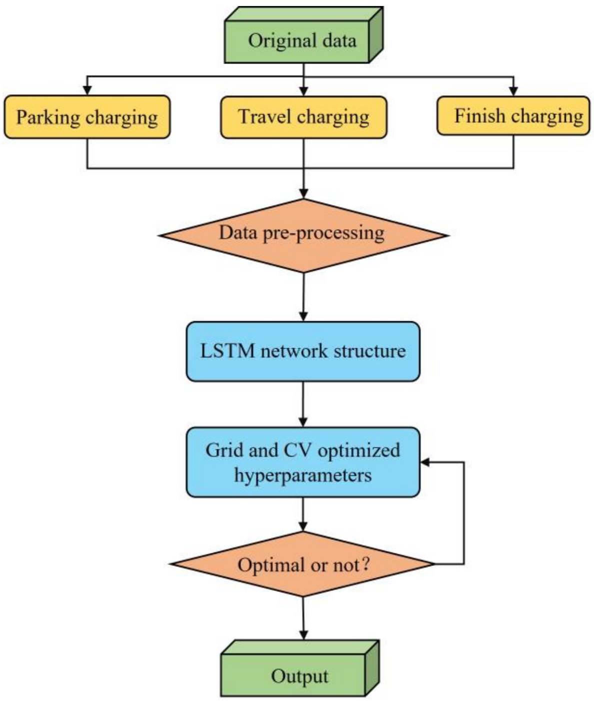

Figure 1. The proposed GSCV optimisation-based LSTM network for SOC estimation process.

3.1. Battery SOC Definition

The SOC of the battery system is used to reflect the remaining capacity of the battery, which is defined numerically as the ratio of the remaining capacity to the battery capacity [17], the definition as shown in Equation (1). Where Q_c refers to the remaining available capacity of the battery at a given moment and Q_n refers to the rated capacity of the battery. Q_n decreases as the battery ages, but the rate of change is small. Battery SOC

cannot be measured directly but can only be estimated by the battery terminal voltage, charge and discharge current, and internal resistance parameters. These parameters are also affected by various uncertainties, such as battery ageing, ambient temperature changes, and the driving status of EVs, so *SOC* estimation of the battery system based on real-world EV data is of great importance to improve the EV's range and optimise battery performance.

$$SOC = \frac{Q_c}{Q_n} \times 100\% \tag{1}$$

3.2. Pearson Related Analysis

In statistics, the Pearson correlation coefficient is adopted to measure the correlation between two variables, X and Y. The Pearson correlation coefficient between two variables is defined as the quotient of the covariance and standard deviation between the two variables [18], as shown in Equation (2).

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \tag{2}$$

The above equation defines the overall correlation coefficient, which is commonly represented by the ρ. Estimating the covariance and standard deviation of the sample gives the sample Pearson correlation coefficient, as shown in the following:

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \tag{3}$$

The correlation coefficient is a dimensionless statistical indicator that takes values in the range $-1 \le r \le 1$. A correlation coefficient less than 0 is a negative correlation, more than 0 is a positive correlation, and equal to 0 indicates no correlation. Wherein, the correlation coefficients are classified as shown in Table 1. The larger the absolute value of the correlation coefficient, the closer the correlation between the two variables.

Table 1. Classification of correlation coefficients.

r	Correlation
0.8~1	Extremely strong
0.6~0.8	Strong
0.4~0.6	Moderate
0.2~0.4	Weak
0.0~0.2	Extremely weak or no

3.3. LSTM Neural Networks

LSTM networks can represent various nonlinear dynamic systems by mapping input sequences to output sequences. Recently, some studies have used LSTM neural networks as a research method, such as energy consumption prediction of power systems based on LSTM neural networks [19]. Since a single LSTM neural network cannot better satisfy the time-series and nonlinear regression studies, convolutional neural networks are combined with LSTM to form hybrid neural networks to improve the accuracy of LSTM models, and they are applied to power system transient stability identification and photovoltaic system power generation prediction, etc. [20,21]. In addition, the selection of LSTM optimal hyperparameters for urban electricity price prediction based on a differential evolutionary algorithm is also investigated [22]. Thus, LSTM neural networks have great application prospects in nonlinear regression problems and time series forecasting. The schematic diagram of the LSTM network structure is shown in Figure 2. During model training, the input gate, output gate, and forget gate can allow the LSTM network to delete information or write new information to the memory unit [23].

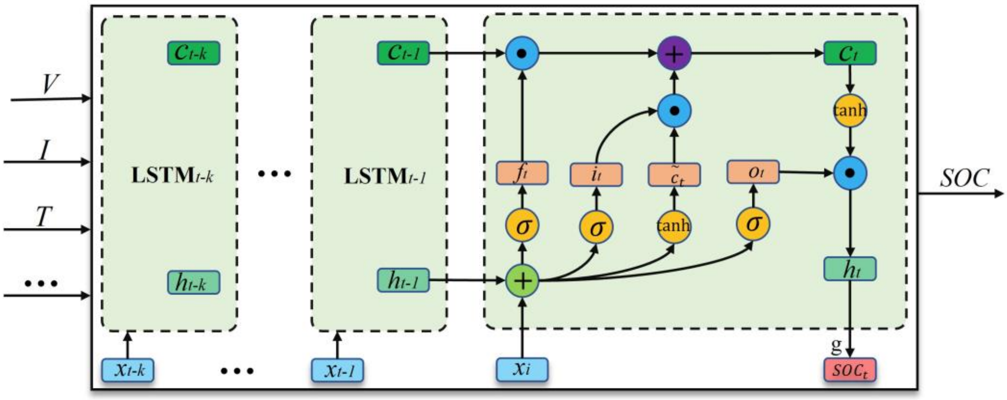

Figure 2. Schematic diagram of the LSTM predictor.

Among them, the memory unit determines what information is stored through the switch of the door, including the element-by-element multiplication operation of the sigmoid function, whose output range is all between 0 and 1. The new sequence value InP_t is connected to the unit $OutP_{t-1}$ from the previous output, and this combined input will be squeezed through the tanh layer and passed through the input gate. The input gate is the tanh layer, activated by the sigmoid function. The sigmoid function of the input gate can receive and clear any unwanted input vector elements. Values from 0 to 1 can be output from the sigmoid function, so values close to 0/1 can be trained by connecting inputs to the weights of these nodes, while some inputs can be blocked or discarded [24,25]. In addition, the LSTM memory unit has an internal variable P_t that is one time step behind, and the risk of gradient disappearance can be reduced by adding P_{t-1} to create an effective recursive layer. This recursive loop is controlled by the forgetting gate, and the compression function of the output layer tanh is ultimately controlled by the output gate, which determines which values can ultimately be output from the unit $OutPt$.

Input Gate: The tanh layer activation function compresses the input and can be represented as:

$$g_t = \tanh(b_g + InP_t IW_g + OutP_{t-1} OW_g) \tag{4}$$

where IWg and OWg represent the weights of the input and previous outputs, respectively, and bg is the input bias.

Multiply the compressed input by the output units of the input gate, which are the nodes activated by a series of sigmoid activation functions:

$$i_t = \sigma(b_i + InP_t IW_i + OutP_{t-1} OW_i) \tag{5}$$

This is the first step in LSTMs, as they decide which information will be discarded. This decision was made by forgetting the door. It accepts inputs from $InPt$ and $OutP_{t-1}$ and outputs numbers between 0 and 1, where the output value 1 indicates a fully reserved state and 0 indicates a fully discarded state [26]. For multiparameter prediction of a battery system, the discarded information can be redundant information such as outliers, noise, or unrelated parameters. The forgetting door can be calculated as follows:

$$f_t = \sigma(b_f + InP_t IW_f + OutP_{t-1} OW_f) \tag{6}$$

Therefore, the output of the forgetting gate/state loop will be (where the operation symbol "∘" represents the multiplication of each element):

$$P_t = P_{t-1} \circ f_t + g \circ i \tag{7}$$

The final output information is determined by the output gate, which can be implemented as follows:

$$O = \sigma(b_o + InP_t IW_o + OutP_{t-1} OW_o) \tag{8}$$

$$OutP_t = \tanh(P_t) \circ O \tag{9}$$

Due to the use of recursive networks to manipulate vector sequences, LSTMs can have five types of input-output mappings, as shown in Figure 3, including one-to-one, many-to-one, one-to-many, and many-to-many, including *m-m* with the same number of inputs and outputs and *m-n* with different input-output numbers.

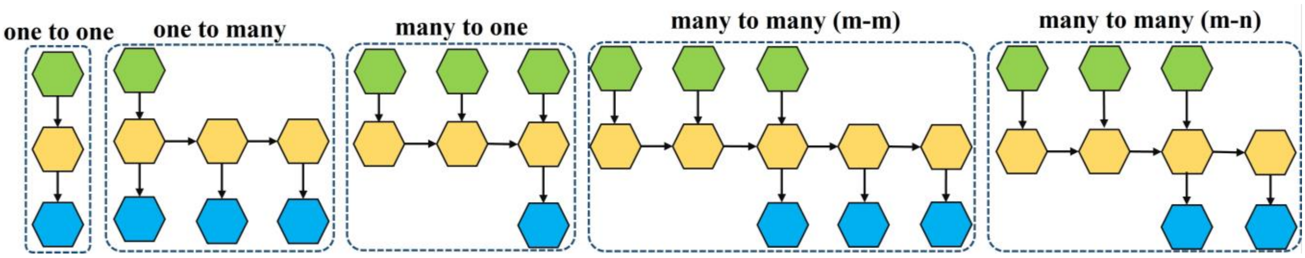

Figure 3. Five types of mapping for LSTM.

To evaluate the predictive performance of a model, mean relative error (MRE) and mean square error (MSE) are the two most commonly used error evaluation methods. The formulas for MRE and MSE in this study can be expressed as:

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

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

where n is the number of training samples or test samples, $\hat{P}_i$ and P_i represents the predicted and actual values, respectively.

3.4. Grid Search and Cross Validation Optimation

In machine learning models, parameters that need to be manually selected are called hyperparameters. For example, the number of decision trees in the random forest, the number of hidden layers and nodes in each layer in the artificial neural network (ANN) model, and the size of the constant in the regular term need to be specified in advance. Improper selection of hyperparameters can lead to underfitting or overfitting. When selecting hyperparameters, there are two ways, one is to fine-tune by experience, and the other is to select parameters of different sizes, bring them into the model, and select the parameters with the best performance. The GSCV algorithm is a method to optimise the model's performance by traversing a given combination of parameters [27]. This parameter tuning method cycles through all the candidate parameter choices for each possibility, with the final result being the best-performing parameter. Grid search is suitable for three or five hyperparameters. By listing a relatively small range of hyperparameter values, the cartesian product of these hyperparameters is a set of hyperparameters. The grid search algorithm trains the model using each set of hyperparameters and picks the hyperparameter combination with the smallest validation error. The algorithm search process is shown in Figure 4.

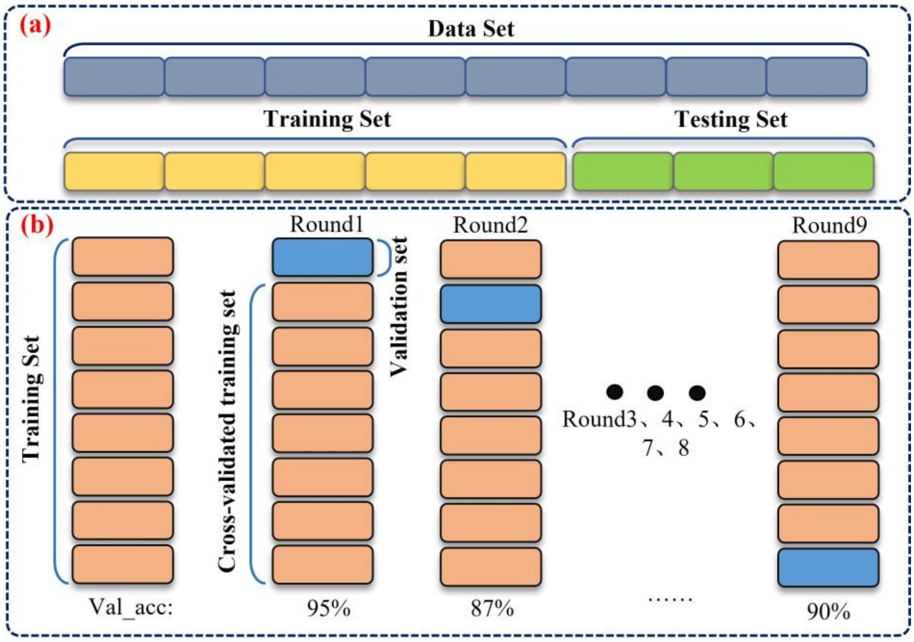

Figure 4. Data set partitioning and grid search: (**a**) dividing the data set into a training set and a test set; (**b**) grid search and cross-validation Schematic.

3.5. Gaussian Noise

Gaussian noise is a class whose probability density function follows a gaussian distribution, commonly known as Gaussian white noise [28]. The amplitude distribution of Gaussian white noise obeys the Gaussian distribution, while the power spectral density is also uniformly distributed. The second-order moments of gaussian white noise are uncorrelated, and the first-order moments are constant, which refers to the correlation of successive signals in time. Gaussian white noise includes thermal noise and scattered particle noise. In this paper, the amount of data is increased to improve the robustness and accuracy of the model through Gaussian noise with a Gaussian distribution, as shown in Equation (12).

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{12}$$

where x is the random variable, μ is the position parameter, and σ is the scale parameter.

4. Results and Discussion

4.1. Pearson and Heat Map Analysis of SOC Estimation Related Parameters

In this section, the real-world EV data is divided into: parking charging, travel charging, and finishing charging. Therein, the Pearson correlation analysis is used to extract the correlation parameters from the real-world EV battery SOC estimation. The results are shown in Table 2, where Pearson correlation coefficients r_p for the parking charging, r_T for the traveling charging condition, and r_F for the finish charging are indicated. The table shows that the standard SOC variation is related to the *total voltage* for three operating conditions. Therefore, the *total voltage* can be used as one of the input parameters for the SOC estimation under three operating conditions. Similarly, the *maximum cell voltage* and the *minimum cell voltage* can also be used as one of the input parameters. For the parking charging condition, *speed*, *charging status*, *single temperature number*, etc. are independent of SOC variation. At the same time, the correlation coefficient for *total current* is 0.5501, which is moderately positive and can therefore be used as an input parameter for the parked charging condition. In addition, the *maximum cell voltage number* correlation coefficient of −0.5286 is moderately negatively correlated and can be used as an input parameter in the parked condition. In the uncharged condition, the *charging status, single temperature* *number*, and *single voltage number* are independent of the standard SOC variation. Finally, the standard SOC variation is independent of the *speed*, *total voltage*, *total current*, and *single temperature number* in the charging completion condition. The next section will show a heat map analysis of the correlation between specific correlation coefficients.

Table 2. Pearson analysis with parameters related to the SOC for working conditions.

Parameters	r_P	r_T	r_F
Speed	NAN	−0.0300	NAN
Charging status	NAN	NAN	NAN
Total voltage	0.9673	0.9561	0.9736
Total Current	0.5501	−0.0227	NAN
Mileage	−0.1967	−0.2587	−0.1295
Maximum cell voltage	0.9579	0.9657	0.9752
Maximum cell voltage number	−0.5286	−0.1829	−0.0228
Minimum cell voltage	0.9661	0.9665	0.9754
Minimum cell voltage number	−0.268	−0.2797	−0.2254
Maximum temperature	−0.1938	0.2884	−0.2567
Maximum temperature probe number	0.0750	−0.0685	0.0805
Minimum temperature	−0.1549	0.2711	−0.2037
Minimum temperature probe number	−0.1507	0.0103	−0.2108
Insulation resistance	0.0890	−0.0501	−0.0186
DCDC status	0.0050	0.0417	−0.0356
Single temperature number	NAN	NAN	NAN
Single voltage number	NAN	NAN	NAN
Gear	NAN	−0.0396	NAN

The analysis of the parking charging condition is illustrated in Figure 5. After eliminating the parameters that are not related to the real-world battery SOC variation for heat map analysis, it can be seen that the parameters related to the parked condition SOC estimation are mainly *total voltage*, *total current*, *maximum cell voltage*, *maximum cell voltage number*, and *minimum cell voltage*. The analysis of the travel charging condition is shown in Figure 6, where it can be seen that the real-world battery SOC variation is related to the *total voltage*, *maximum cell voltage number*, and *minimum cell voltage number*. These parameters can therefore be used as input parameters for the SOC estimation. Similarly, according to Figure 7, it can be seen that the SOC estimation for the charge completion condition is related to the *total voltage*, *maximum cell voltage number*, and *minimum cell voltage number*, and can also be used as input parameters for SOC estimation.

4.2. SOC Estimation by LSTM Network Based on GSCV Optimization

This section is based on the LSTM network with GSCV optimization for real-world EV battery SOC estimation. As shown in Figure 8 and Table 3, (a), (c) and (e) indicate the results of SOC estimation under three different operating conditions. Additionally, (b), (d) and (f) illustrated the absolute error for SOC estimation results. It can be seen that under three different working conditions, the absolute error of SOC estimation is controlled within 4%. This is because the sampling accuracy and interval are relatively low under real-world EV working conditions, leading to a decline in estimation accuracy. However, for Figure 8c,d, the absolute error can be well controlled within 2% because of the large amount of data. Meanwhile, the MSE is kept within a small range. As shown in Table 4, the errors of various SOC estimation studies are compared. The results of this study are acceptable due to the low data accuracy and large sampling interval of the real-world EV data. Therefore, it can be seen that this method greatly affects the large amount of data generated by the operation of real-world EVs and can be applied to the SOC estimation of real-world EVs. However, due to the limited finish charging data, this study adopts the Gaussian noise algorithm to enhance the data, where μ = 0 and σ = 0.12. Every piece of fifth data in the original data is grouped, and the noise data is added at the end of each group. Nevertheless, some operating points still exist where the SOC estimation cannot be estimated well after data enhancement by the Gaussian noise algorithm. Therefore, the proposed method in this paper can achieve accurate SOC estimation for real-world EVs.

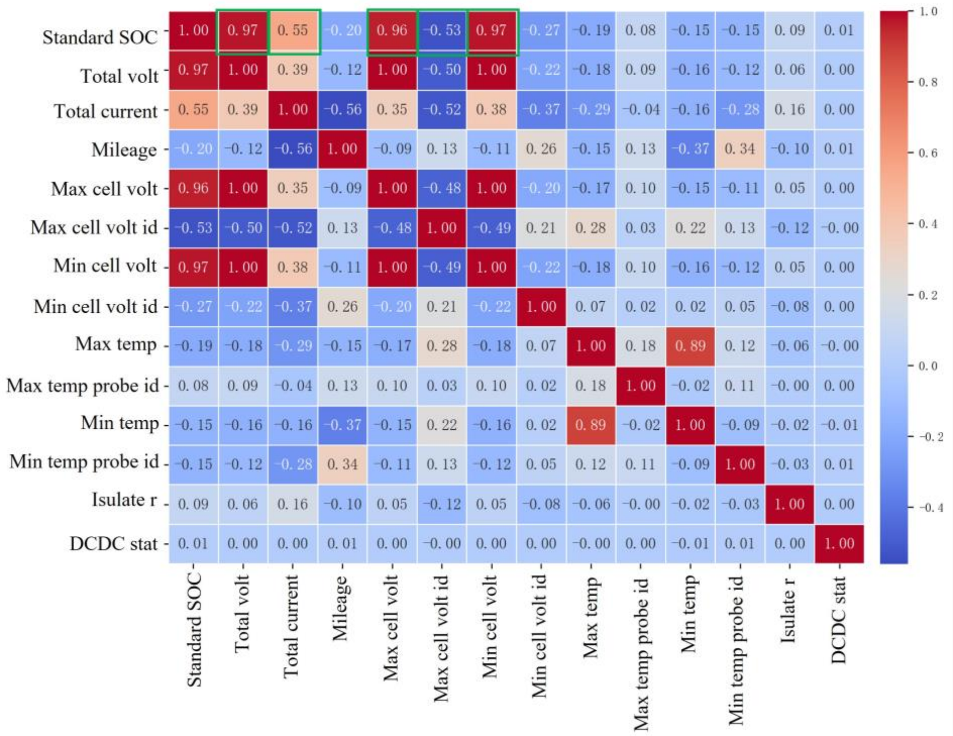

Figure 5. Heat map analysis of parking and charging correlation coefficients.

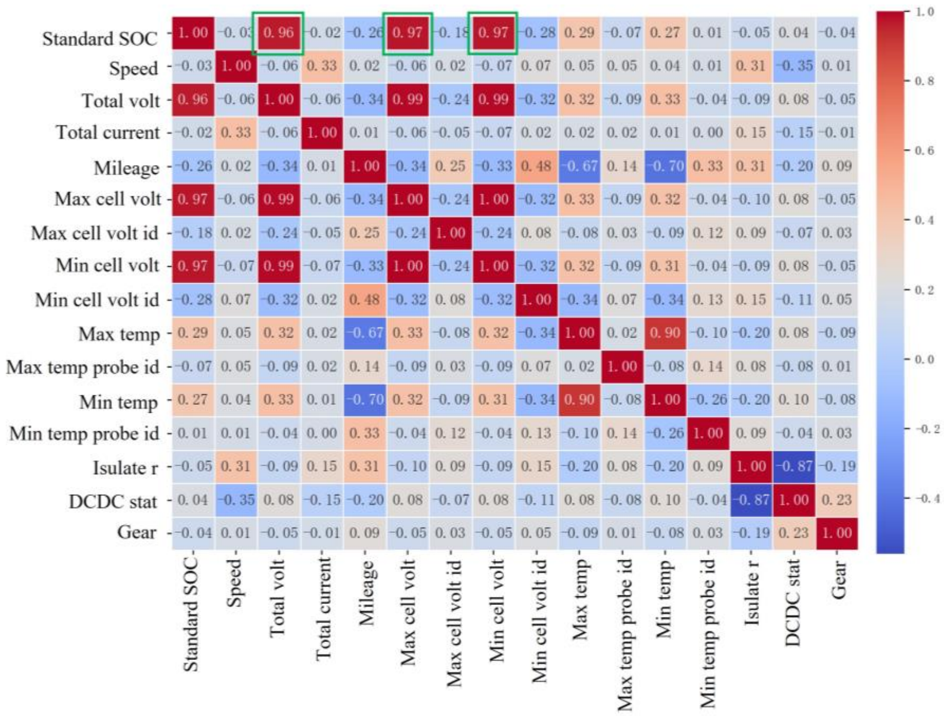

Figure 6. Heat map analysis of the correlation coefficient of the uncharged state.

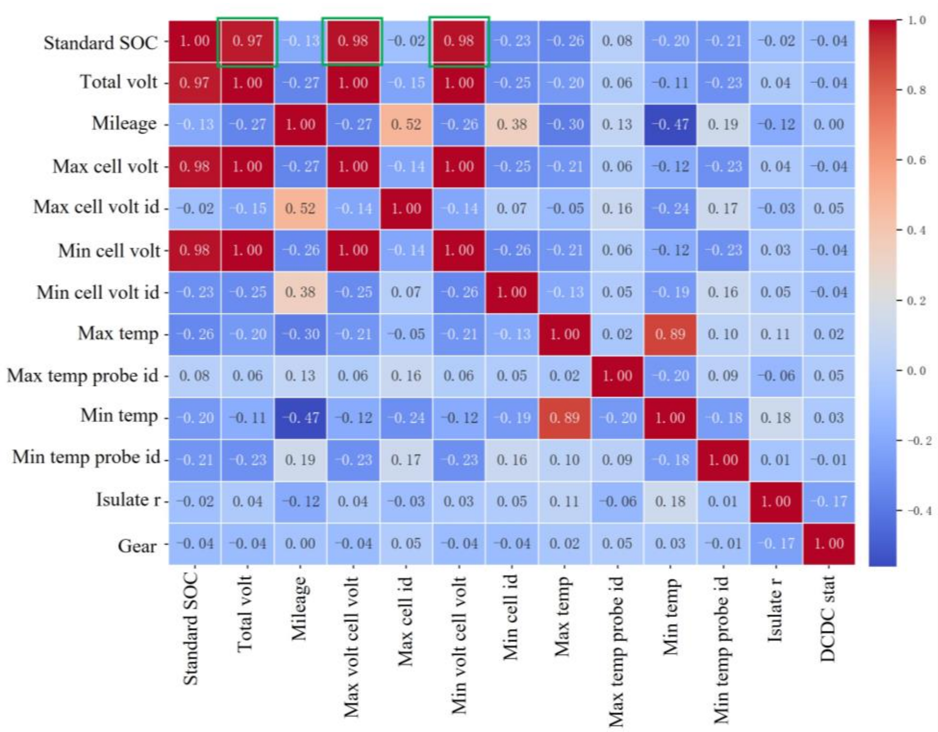

Figure 7. Charge completion correlation coefficient heat map analysis.

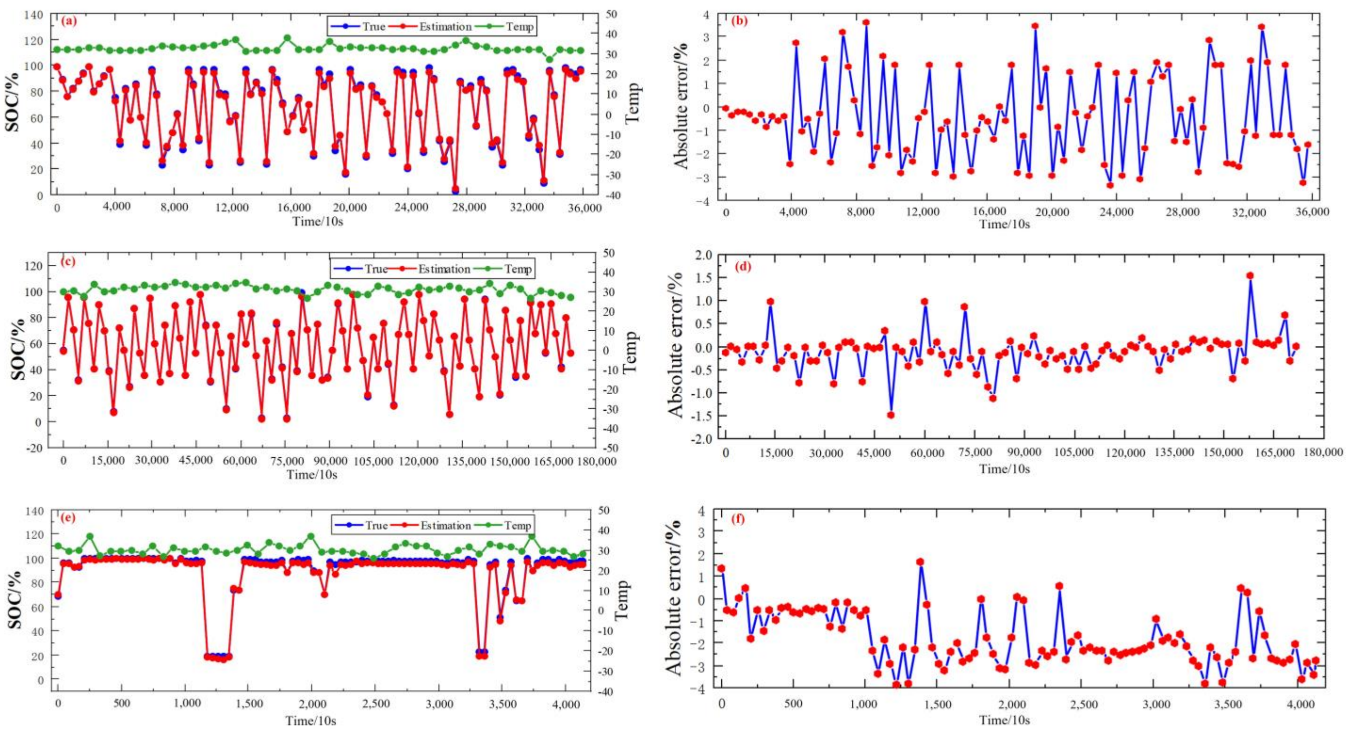

Figure 8. SOC estimation results and absolute errors under three different working conditions. (**a**) SOC estimation results under parking charging; (**b**) SOC estimation absolute error under parking charging; (**c**) SOC estimation results under travel charging; (**d**) SOC estimation absolute error under travel charging; (**e**) SOC estimation results under finish charging; (**f**) SOC estimation absolute error under finish charging.

Table 3. SOC estimation result error in three different operating conditions.

Condition/Error	Maximum Absolute Error	Minimum Absolute Error	MSE
Parking charging	3.61	−3.35	0.0042
Travel charging	1.54	−1.49	0.0031
Finish charging	1.64	−3.85	0.0067

Table 4. Comparison of errors for various studies.

Method	Battery Chemistry	Estimated Target	Research Environment	Precision
PSO-LSTM [1]	Li-ion battery	SOC	Laboratory experiment	MAE < 0.2% RMSE < 0.3%
LSTM-RNN [4]	Li-ion battery	SOC	Laboratory experiment	RMSE < 1.5%
CNN-LSTM [5]	Li-ion battery	SOC	Laboratory experiment	RMSE < 1% MAE < 1%
LSTM [29]	Li-sulfur battery	SOC	Laboratory experiment	RMSE < 6%
NARX-LSTM [30]	Li-ion battery	SOC	Laboratory experiment	RMSE < 1% MAE < 1%

At the same time, in order to verify the robustness and effectiveness of this method, this study selects the operation data of three other real-world EVs and divides them into three different operating conditions. Figures 9 and 10 are the results and MSE of SOC estimation under different validation conditions, respectively. (a-1), (a-2), and (a-3) indicated the SOC estimation results for vehicle A. (b-1), (b-2), and (b-3) show the SOC estimation results for vehicle B. Moreover, (c-1), (c-2), and (c-3) show the SOC estimation results for vehicle C. It can be seen that the SOC estimation results for different EVs verify the effectiveness and accuracy of the proposed method. Wherein the MSE does not exceed 0.006 in any case. Therefore, this method has excellent performance for the SOC estimation of real-world EVs.

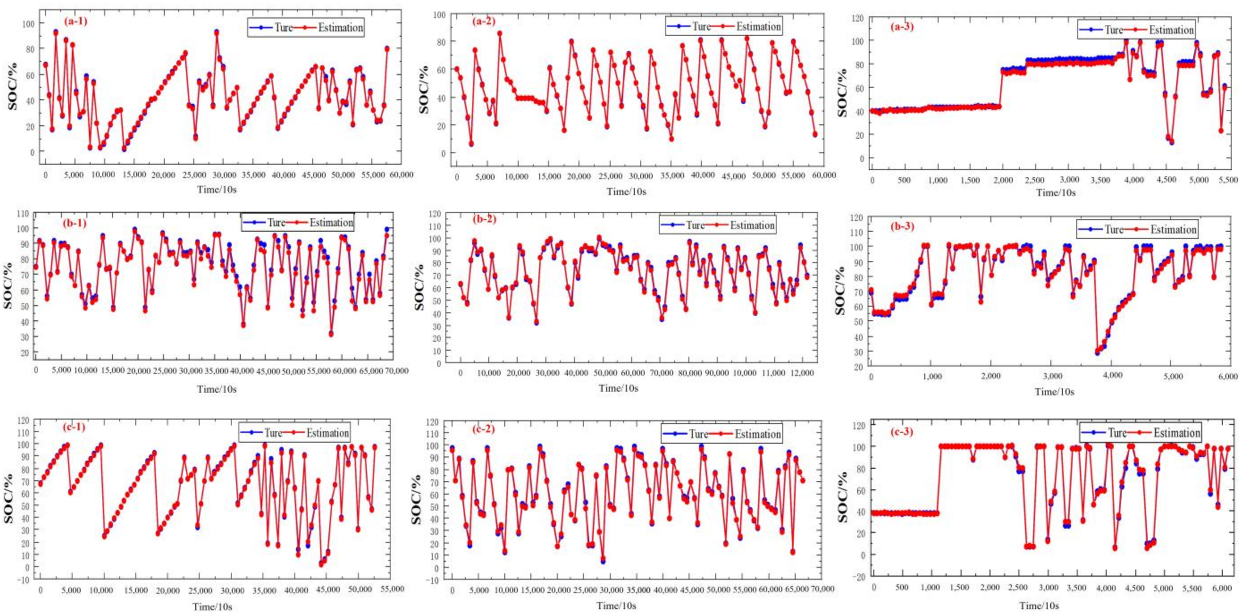

Figure 9. SOC estimation results under different validation cases. (**a-1**, **a-2**, **a-3**) SOC estimation results of vehicle a under three different working conditions; (**b-1**, **b-2**, **b-3**) SOC estimation results of vehicle b under three different working conditions; (**c-1**, **c-2**, **c-3**) SOC estimation results of vehicle c under three different working conditions.

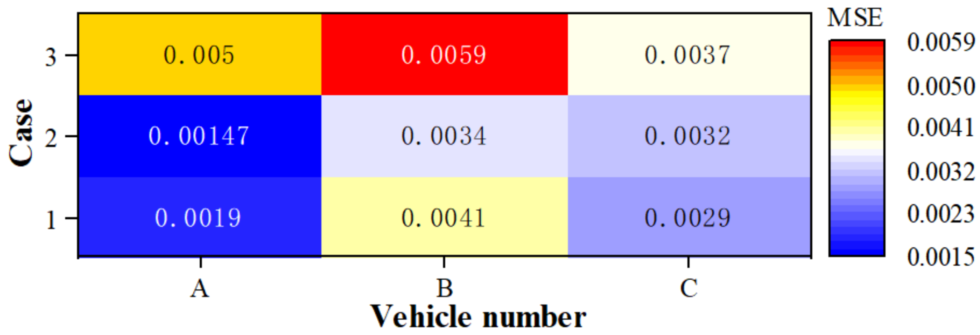

Figure 10. MSEs of SOC estimation results under different validation cases.

5. Conclusions

Battery SOC is one of the main parameters for the battery management system, but its inability to be measured directly poses a great challenge for the real-world electric vehicle. Currently, most SOC estimation methods are still in the laboratory stage. Therefore, it is critical to investigate the SOC estimation methods for real-world electric vehicles. This paper indicates a novel grid search and cross-validation optimised long and short-term memory network hyperparameter approach to conduct battery SOC estimation based on real-world electric vehicle data. The results showed that optimised hyperparameters of the long and short-term memory networks could achieve a high accuracy of SOC estimation with an absolute error of SOC estimation of less than 4%. Wherein the absolute error accuracy of SOC estimation is less than 2% under the conditions of the large dataset.

In this study, the model's applicability could not be verified due to the limited data. However, it is possible to achieve a high accuracy state of charge estimation for features such as large intervals and low sampling frequency of the sampled data from real-world electric vehicles, so it can be inferred that the method is equally applicable to data with similar features.

Author Contributions: Methodology, J.H. and F.L.; Validation, J.H., F.L. and X.G.; Formal analysis, J.H., X.G. and Q.Y.; Investigation, J.H. and X.X.; Writing—original draft, F.L.; Writing—review & editing, X.G., X.X. and Q.Y.; Supervision, X.X. and Q.Y. All authors have read and agreed to the published version of the manuscript.

Funding: The project is supported partly by the National Natural Science Foundation of China (No. 52107220), Postdoctoral Research Fund Project of China (No. 2021M690353), Foundation of the State Key Laboratory of Automotive Simulation and Control (No. 20210211), Scientific and Technological Innovation Foundation of Foshan (No. BK21BE012), the Postdoctoral Research Foundation of the Shunde Graduate Innovation School of University of Science and Technology Beijing (No. 2021BH007), Fundamental Research Funds for the Central Universities (No. FRF-BD-20-08A, NO.FRF-IDRY-21-013).

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

References

1. Ren, X.; Liu, S.; Yu, X.; Dong, X. A method for state-of-charge estimation of lithium-ion batteries based on PSO-LSTM. *Energy* **2021**, *234*, 121236. [CrossRef]
2. Hong, J.; Wang, Z.; Chen, W.; Wang, L.-Y.; Qu, C. Online joint-prediction of multi-forward-step battery SOC using LSTM neural networks and multiple linear regression for real-world electric vehicles. *J. Energy Storage* **2020**, *30*, 101459. [CrossRef]
3. Hong, J.; Wang, Z.; Zhang, T.; Yin, H.; Zhang, H.; Huo, W.; Zhang, Y.; Li, Y. Research on integration simulation and balance control of a novel load isolated pure electric driving system. *Energy* **2019**, *189*, 116220. [CrossRef]
4. Chen, J.; Zhang, Y.; Wu, J.; Cheng, W.; Zhu, Q. SOC estimation for lithium-ion battery using the LSTM-RNN with extended input and constrained output. *Energy* **2023**, *262*, 125375. [CrossRef]
5. Yang, F.; Zhang, S.; Li, W.; Miao, Q. State-of-charge estimation of lithium-ion batteries using LSTM and UKF. *Energy* **2020**, *201*, 117664. [CrossRef]

6. Chen, G.; Jiang, S.; Xie, M.; Yang, F. A hybrid DNN-KF model for real-time SOC estimation of lithium-ion batteries under different ambient temperatures. In Proceedings of the 2022 Global Reliability and Prognostics and Health Management (PHM-Yantai), Yantai, China, 13–16 October 2022; pp. 1–5. [CrossRef]
7. Xiong, X.; Wang, S.; Fernandez, C.; Yu, C.; Zou, C.; Jiang, C. A novel practical state of charge estimation method: An adaptive improved ampere-hour method based on composite correction factor. *Int. J. Energy Res.* **2020**, *44*, 11385–11404. [CrossRef]
8. Ren, Z.; Du, C.; Wu, Z.; Shao, J.; Deng, W. A comparative study of the influence of different open circuit voltage tests on model-based state of charge estimation for lithium-ion batteries. *Int. J. Energy Res.* **2021**, *45*, 13692–13711. [CrossRef]
9. Bao, Y.; Dong, W.; Wang, D. Online Internal Resistance Measurement Application in Lithium Ion Battery Capacity and State of Charge Estimation. *Energies* **2018**, *11*, 1073. [CrossRef]
10. Cui, Z.; Wang, L.; Li, Q.; Wang, K. A comprehensive review on the state of charge estimation for lithium-ion battery based on neural network. *Int. J. Energy Res.* **2021**, *46*, 5423–5440. [CrossRef]
11. Xie, F.; Wang, S.; Xie, Y.; Fernandezb, C.; Li, X.; Zou, C. A novel battery state of charge estimation based on the joint unscented kalman filter and support vector machine algorithms. *Int. J. Electrochem. Sci.* **2020**, *15*, 7935–7953. [CrossRef]
12. Tian, J.; Xiong, R.; Shen, W.; Lu, J. State-of-charge estimation of LiFePO4 batteries in electric vehicles: A deep-learning enabled approach. *Appl. Energy* **2021**, *291*, 116812. [CrossRef]
13. Wang, Q.; Jiang, J.; Gao, T.; Ren, S. State of Charge Estimation of Li-Ion Battery Based on Adaptive Sliding Mode Observer. *Sensors* **2022**, *22*, 7678. [CrossRef] [PubMed]
14. Lin, X.; Tang, Y.; Ren, J.; Wei, Y. State of charge estimation with the adaptive unscented Kalman filter based on an accurate equivalent circuit model. *J. Energy Storage* **2021**, *41*, 102840. [CrossRef]
15. Kwak, M.; Lkhagvasuren, B.; Park, J.; You, J.-H. Parameter Identification and SOC Estimation of a Battery Under the Hysteresis Effect. *IEEE Trans. Ind. Electron.* **2019**, *67*, 9758–9767. [CrossRef]
16. Snoussi, J.; Ben Elghali, S.; Zerrougui, M.; Bensoam, M.; Benbouzid, M.; Mimouni, M.F. Unknown input observer design for lithium-ion batteries SOC estimation based on a differential-algebraic model. *J. Energy Storage* **2020**, *32*, 101973. [CrossRef]
17. Liu, Z.; Chen, S.; Jing, B.; Yang, C.; Ji, J.; Zhao, Z. Fractional variable-order calculus based state of charge estimation of Li-ion battery using dual fractional order Kalman filter. *J. Energy Storage* **2022**, *52*, 104685. [CrossRef]
18. Li, R.; Wang, H.; Dai, H.; Hong, J.; Tong, G.; Chen, X. Accurate state of charge prediction for real-world battery systems using a novel dual-dropout-based neural network. *Energy* **2022**, *250*, 123853. [CrossRef]
19. Peng, L.; Wang, L.; Xia, D.; Gao, Q. Effective energy consumption forecasting using empirical wavelet transform and long short-term memory. *Energy* **2021**, *238*, 121756. [CrossRef]
20. Azhar, I.F.; Putranto, L.M.; Irnawan, R. Development of PMU-Based Transient Stability Detection Methods Using CNN-LSTM Considering Time Series Data Measurement. *Energies* **2022**, *15*, 8241. [CrossRef]
21. Lim, S.-C.; Huh, J.-H.; Hong, S.-H.; Park, C.-Y.; Kim, J.-C. Solar Power Forecasting Using CNN-LSTM Hybrid Model. *Energies* **2022**, *15*, 8233. [CrossRef]
22. Peng, L.; Liu, S.; Liu, R.; Wang, L. Effective long short-term memory with differential evolution algorithm for electricity price prediction. *Energy* **2018**, *162*, 1301–1314. [CrossRef]
23. Hong, J.; Wang, Z.; Yao, Y. Fault prognosis of battery system based on accurate voltage abnormity prognosis using long short-term memory neural networks. *Appl. Energy* **2019**, *251*, 113381. [CrossRef]
24. Fahad, S.; Su, F.; Khan, S.U.; Naeem, M.R.; Wei, K. Implementing a novel deep learning technique for rainfall forecasting via climatic variables: An approach via hierarchical clustering analysis. *Sci. Total Environ.* **2023**, *854*, 158760. [CrossRef] [PubMed]
25. Hong, J.; Wang, Z.; Chen, W.; Yao, Y. Synchronous multi-parameter prediction of battery systems on electric vehicles using long short-term memory networks. *Appl. Energy* **2019**, *254*, 113648. [CrossRef]
26. Song, X.; Yang, F.; Wang, D.; Tsui, K.-L. Combined CNN-LSTM Network for State-of-Charge Estimation of Lithium-Ion Batteries. *IEEE Access* **2019**, *7*, 88894–88902. [CrossRef]
27. Erten, G.E.; Keser, S.B.; Yavuz, M. Grid Search Optimised Artificial Neural Network for Open Stope Stability Prediction. *Int. J. Min. Reclam. Environ.* **2021**, *35*, 600–617. [CrossRef]
28. Shi, Q.; Liu, M.; Hang, L. A novel distribution system state estimator based on robust cubature particle filter used for non-gaussian noise and bad data scenarios. *IET Gener. Transm. Distrib.* **2021**, *16*, 1385–1399. [CrossRef]
29. Wang, Z.; Fotouhi, A.; Auger, D.J. State of Charge Estimation in Lithium-Sulfur Cells Using LSTM Recurrent Neural Networks. In Proceedings of the 2020 European Control Conference (ECC), St. Petersburg, Russia, 12–15 May 2020. [CrossRef]
30. Wei, M.; Ye, M.; Li, J.B.; Wang, Q.; Xu, X. State of Charge Estimation of Lithium-Ion Batteries Using LSTM and NARX Neural Networks. *IEEE Access* **2020**, *8*, 189236–189245. [CrossRef]