Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 12, No. 1, 2024

Prediction of State of Charge for Lead-acid Batteries Based on GRU Network and Isolated Forest

Guocheng Li [1], Zhanying Li [1, *], Yinghao Zhang [1], Yang Xiao [1], Ming Chen [2]

[1] School of Information Science and Engineering, Dalian Polytechnic University, Dalian 116000, China
[2] Shenzhen DACEEN Technology Co. Ltd, Shenzhen 518000, China

[*] **Corresponding author**: Zhanying Li (Email: l_zy1979@126.com)

Abstract: Accurate prediction of the state of charge (SOC) of lead-acid batteries is the key to ensuring battery life. In this paper, a new combined SOC prediction model IF-GRU (Isolation Forest, Gated Recurrent Unit) is proposed. The model combines the Isolation Forest anomaly detection algorithm and the Gated Recurrent Network. The Isolation Forest algorithm is used to detect anomalous and missing values in the raw data. Length dependence of the GRU network can be further utilized to perform high-accuracy SOC estimation by implementing a sliding window that takes into account the data's charging and discharging details. In addition, the conventional Adam optimizer is utilized to improve the convergence speed of model training. The experimental data demonstrate that the IF-GRU model proposed in this paper has higher prediction accuracy and convergence speed with a RMSE of 1.59% compared with traditional LSTM network, GRU network, and BP network.

Keywords: Lead-acid battery; State of charge; Neural network; Isolation forest anomaly detection.

1. Introduction

As a kind of chemical power source with a history of more than 160 years, lead-acid battery is widely used in communications, electricity, emergency energy storage and other fields. As a secondary battery, lead-acid batteries influence the State of Charge (SOC) during the use process, such as current, voltage, time and other factors. So, it's crucial to accurately predict the SOC of lead-acid batteries. Considerable research has been undertaken both domestically and internationally regarding the prediction of battery SOC in light of the rapid development of artificial intelligence in recent years[1-3]. Currently, the key techniques for lead-acid battery prediction can be broadly categorized into two methods: one is the physical modeling method, and the other one is the system identification and parameter estimation modeling method. Within these methods, the physical modeling approach presents a complex prediction process, making it challenging to comprehensively account for all influencing factors. As a result, the prediction outcomes often lack the desired level of accuracy, leading to numerous errors. The physical modeling methods encompass various techniques such as the discharge experiment method, ampere-time integration method, density method, open-circuit voltage method, and internal resistance method[4]. The commonly used methods are the discharge experimental method, ampere-time integration method, and open-circuit voltage method. The discharge experiment method is mainly used in the laboratory to calculate the charging efficiency of battery packs, check the accuracy of SOC estimation, or for battery maintenance, and is applicable to all batteries. This method might be risky when applied to critical applications. The ampere-time integration method is the most commonly used SOC estimation method, which essentially treats the battery as a black box and considers that the power flowing into the battery has a proportional relationship with the power flowing out. This method does not consider the inner configuration and external electrical properties of the battery[5-7]. However, in practice, the inaccurate calculation of current often leads to errors. The open-circuit voltage method is based on a certain linear relationship between the remaining capacity of the battery and the open-circuit voltage. The remaining capacity can be obtained directly by measuring the open-circuit voltage. In practical scenarios, as the battery ages and the remaining capacity diminishes, the open-circuit voltage exhibits minimal noticeable changes. So, it is impossible to predict the residual power accurately[8,9]. The Neural network likes a black box, the data predicted with high accuracy by feeding them into the network[10,11]. The Long Short Term Memory (LSTM) network has some advantages in processing and predicting time series[12]. The Gated Recurrent Unit (GRU) network is a variant of LSTM with a simple structure but improves gradient disappearance and explosion problems of recurrent neural networks[13]. Therefore, the GRU network is widely used for prediction [14].

The subsequent sections of this paper are structured as follows: Section 2 presents an analysis of the Lead-acid Battery dataset. Section 3 introduces a combined approach of the isolated forest anomaly detection algorithm with the GRU network. In Section 4, the performance of our model is evaluated through testing. Lastly, Section 5 provides the concluding remarks of this article.

2. Model construction and algorithms

2.1. GRU neural network

The lead-acid battery SOC prediction process is actually the analysis of the battery historical discharge data time series. The longer the time series, the more battery historical discharge data analyzed, and the higher the prediction accuracy. The traditional RNN cannot handle long-time data feature information well. The LSTM network, as an improved algorithm of RNN, solves the long-term dependence problem of RNN by introducing three gates in the network.GRU is a variation of LSTM, which changes the three gate structures of LSTM into two, the reset gate and update gate. The reduction of the number of gate structures greatly reduces the training sample data required for GRU and further accelerates the

convergence speed, which overcomes the overfitting problem of LSTM and largely improves the learning effect of the network[15-20]. The structure of GRU network is shown in Fig. 1.

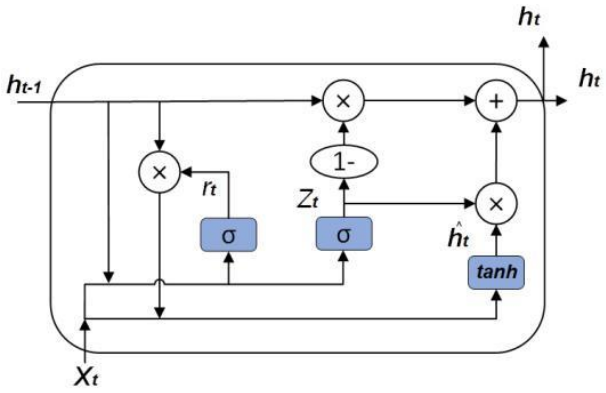

Figure 1. Structure of GRU neural network

$$\begin{cases} r_t = \sigma\left(W_r \cdot [h_{t-1}, W_r x_t] + b_r\right) \\ Z_t = \sigma\left(W_z \cdot [h_{t-1}, W_z x_t] + b_z\right) \\ \tilde{h} = \tanh\left(W_h x_t + W_h\left(r_t \otimes h_{t-1}\right) + b_h\right) \\ h_t = Z_t \otimes h_{t-1} + (1 - Z_t) \otimes \tilde{h}_t \end{cases} \quad (1)$$

In (1): t is the current moment; $t-1$ is the previous moment; r_t and Z_t is the reset gate and update gate, respectively; x_t and h_t is the current input capacity value and output capacity value of the battery, respectively; h_{t-1} is the hidden layer state information passed from the previous cell node; $\tilde{h}_t$ is the cell to be updated; W_r, b_r and W_z, b_z are the weight matrix and bias parameters required to calculate the reset gate output, update gate output and process volume, respectively; σ and $\tanh$ is the $sigmoid$ function and hyperbolic tangent function, respectively.

3. Analysis of examples

3.1. Lead-Acid battery data set

The dataset used in this paper consists of real-time online data obtained from the cloud data management system of a communication base station in Shenzhen. The lead-acid battery pack utilized in the system is of the 2V/500Ah Nandu battery type. The data collection spanned from June 2021 to August 2021, with hourly monitoring. The dataset includes three key variables: voltage, temperature, and SOC of the lead-acid battery. The voltage and temperature data exhibit minimal abnormal values, with values fluctuating within a specific range. For the SOC of lead-acid batteries, the data varies and fluctuates considerably and therefore requires anomaly detection processing. The objective is to attain a high level of accuracy in data prediction with fully use data, to detect some missing values as well as anomalous values, and to enhance the accuracy of lead-acid battery SOC prediction. In this paper, we use the isolated forest anomaly detection algorithm for anomaly detection. We also compare it with the classical local outlier anomaly detection algorithm to obtain the optimal algorithm for SOC of lead-acid batteries for anomaly detection.

3.2. Battery data preprocessing

The SOC data of lead-acid batteries contain missing values, anomalous values and other irregularities. The isolated forest algorithm is a classical and unsupervised learning anomaly detection algorithm, which is introduced in this paper to detect outliers within the lead-acid battery data. iForest, derived from Isolation Forest, is similar to a random forest as it comprises a substantial number of decision trees. The trees within iForest are referred to as isolated trees and tested for each iTree, as shown in Fig.2. iForest is a rapid anomaly detection algorithm based on isolation, introduced by Zhou et al.错误!未找到引用源。-16]. It exhibits linear time complexity and offers high accuracy. Anomalies, due to their distinct characteristics, tend to be more easily isolated compared to normal instances.

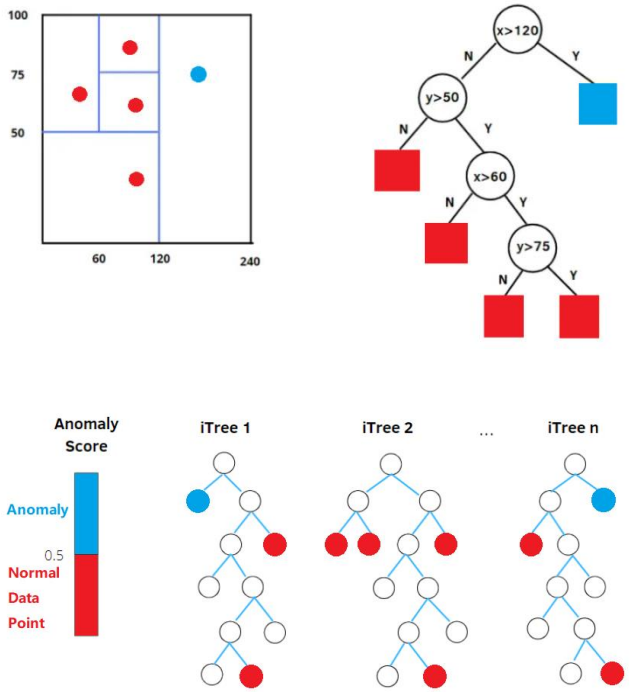

Figure 2. Schematic diagram of isolated forest segmentation

The implementation of iForest can be divided into two phases [17]: the training phase and the evaluation phase.

1) The training phase is as follows:

a. Select randomly sampled points Ψ as a sub-sample set;

b. Randomly selecting the dimensions and generating the variation of cut points p in size between minimum and maximum values;

c. Place data of specified size less than p in the left side of the current node, greater or equal than p in the right side;

d. Repeat steps a and b in the sub-node to construct a new sub-node in succession until the data is indivisible or the child nodes have reached a limited height. The iForest training is finished after obtaining N iTrees.

2) Evaluation phase: for each test point, calculate the anomaly score. The formula is as follows:

$$S(x, \varphi) = 2^{-\frac{E(h(x))}{c(\varphi)}} \quad (2)$$

$h(x)$ is the height of sample x in each tree for a given sample. $c(\varphi)$ is the given number of sample. Φ is the average of the path length, which is used to normalize the path length $h(x)$ of sample x. $E(h(x))$ is the mean value of the

sample. When the anomaly score is close to 1, the data points are easily isolated, which means they are anomalous points. When the score is less than 0.5, it is a normal data point.

As shown in Fig. 3, SOC data of lead-acid battery has been abnormal detected. A total of 33 data outliers were found, printed and marked in red. In the case of abnormal detection, because the data changes before and after are relatively large, the method of outlier elimination is adopted.

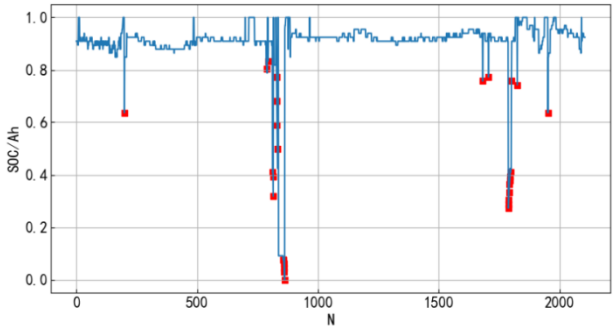

Figure 3. Data anomaly detection diagram

3.3. Evaluation metrics

Numerous evaluation methods exist to assess the effectiveness of a model. In this paper, to ensure both accuracy and efficiency, the following statistical metrics have been selected: Root Mean Square Error (RMSE) [21], Mean Absolute Error (MAE)[23-25], Mean Absolute Percentage Error (MAPE), Mean Relative Percentage Error (MRPE), and coefficient of determination (R^2). Regarding RMSE and MAE, a value closer to zero indicates better performance. For R^2, a value closer to 1 indicates a stronger correlation. MAPE is a relative measure, where a smaller value indicates better performance, but it should be compared within different models. The formulas for each evaluation indicator are as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2} \tag{3}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|y_t - \hat{y}_t\right| \tag{4}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|y_t - \bar{y}_t\right|}{y_t} \tag{5}$$

$$R^2 = 1 - \frac{\sum_{t=0}^{n-1}\left(y_t - \hat{y}_t\right)^2}{\sum_{t=0}^{n-1}\left(y_t - \bar{y}_t\right)^2} \tag{6}$$

In the above equations, n is the number of samples, y_t is the true value of lead-acid batteries, $\hat{y}_t$ is the predicted value of lead-acid batteries, and $\bar{y}_t$ is the mean value of the samples.

3.4. Sliding window

The sliding window method is used when dealing with time series prediction. Similar to the principle of CNN in image processing, this approach employs sliding windows to capture sequential segments within sequence data. Due to the influence of sampling accuracy, the battery data sequence may have a disparity in length during the actual operation of the battery. At the same time, In the battery case, the output of the system has limited temporal dependence on the input sequence, which means it is weakly correlated with the older input information[26-28]. Therefore, in this paper, a sliding window-based approach is used to implement the preprocessing of the input vectors. The experimental data presented in this paper are time series data, and the prediction strategy is single-step prediction, which means the time data of the first N moments are used to predict the N+1th moment. The size of the sliding window is denoted as N. To determine the optimal value of N, the size of the sliding window is changed for the experiment. The experimental results are shown in Table 1.

Table 1. Sliding window experiment

N	RMSE	MAE	MAPE	R^2
3	0.0177	0.0115	3.3408	0.6686
4	0.0187	0.0127	3.3582	0.6303
5	0.0167	0.0090	3.2487	0.7070
6	0.0168	0.0093	**3.2320**	0.7053
7	0.0180	0.0114	3.2996	0.6611
8	**0.0162**	**0.0081**	3.2654	**0.7251**
9	0.0204	0.0147	3.3750	0.5637

Therefore, we achieved the minimum value of convergence when we set the N to 8. At the same time, it has the best prediction with the smallest error taking into account RMSE and MAPE.

3.5. Parameter selection

Table 2. Experimental diagram of the parameter settings

A total of 50 GRU units		RMSE	MAE	MAPE	R2
Epoch=50	batch_size=1	0.0179	0.0114	3.3492	0.6654
Epoch=100	batch_size=1	0.0176	0.0097	3.3478	0.6766
Epoch=1000	batch_size=1	0.0302	0.0211	3.6202	0.0433
Epoch=50	**batch_size=8**	**0.0161**	**0.0083**	**3.2713**	**0.7186**
Epoch=100	batch_size=8	0.0162	0.0084	3.3076	0.7236
Epoch=1000	batch_size=8	0.0232	0.0137	3.6751	0.4375
Epoch=50	batch_size=16	0.0171	0.0100	3.1800	0.6938
Epoch=100	batch_size=16	0.0184	0.0123	3.3401	0.6436
Epoch=1000	batch_size=16	0.0256	0.0177	3.5867	0.3142
Epoch=50	batch_size=32	0.0181	0.0113	3.0892	0.6550
Epoch=100	batch_size=32	0.0162	0.0086	3.2365	0.7152
Epoch=1000	batch_size=32	0.0171	0.0088	3.3746	0.6932

In this paper, the GRU network is used, which has a GRU layer that maps the learned states to the desired SOC output.

Input time series of the given deep GRU network are defined as $[Input_{t0}, Input_{t1}, \cdots, Input_{tn}]$ and $[V_t, T_t, \cdots, SOC_t]$, represent the voltage, temperature and SOC for each time step, respectively. The implemented model utilizes the Keras library, with the Adam optimizer selected to update the network's weights and biases. To identify the optimal model, various experiments were conducted by adjusting the epochs, batch size, and GRU units (with the activation function set as tanh).The experimental results are shown in TABLE 2. The following are the fundamental parameters configured for the model:

Number of GRU units: 50
Batch size: 8
Epoch:50
Optimizer: Adam.

3.6. Combined prediction model of GRU network and isolated forest anomaly detection

In this paper, we proposed a combined model IF-GRU for predicting the SOC of lead-acid batteries, and the prediction flow chart is shown in Fig.4. The initial step involves data processing, which includes handling multiple variables such as voltage, temperature, and SOC data, along with performing anomaly detection and applying a sliding window technique to the input data. The main objective of the prediction model proposed in this paper is to identify deviant points within the SOC data using isolated forest anomaly detection. By achieving accurate data prediction, the model aims to effectively utilize valuable data and enhance the prediction of SOC for lead-acid batteries.

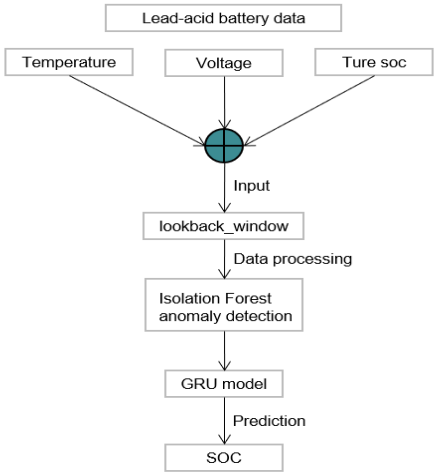

Figure 4. The prediction flow chart

4. Experiment and analysis

In this experiment, the training and validation datasets are split in a ratio of 7:3, and the dataset is from June 2021 to August 2021. The models compared in this study include the traditional LSTM, BP, GRUand LOF-LSTM network, which combines the local outlier detection method with LSTM network. The results of SOC prediction for lead-acid batteries are shown in Figs. 5, 6, 7, 8, 9and Table 3. The prediction plots comparing the predicted and real values show that the IF-GRU model has better prediction effect and the fastest convergence speed. And the RMSE of IF-GRU model reaches 1.99%.

Table 3. Evaluation metrics for each model

Model	RMSE	MAE	MAPE	R^2
LSTM	0.0451	0.0165	7.6871	0.7413
BP	0.0457	0.0226	7.8279	0.7346
GRU	0.0433	0.0170	7.7926	0.7620
LOF-LSTM	0.0369	0.0157	4.5803	0.5700
IF-GRU	**0.0199**	**0.0142**	**3.3000**	**0.7871**

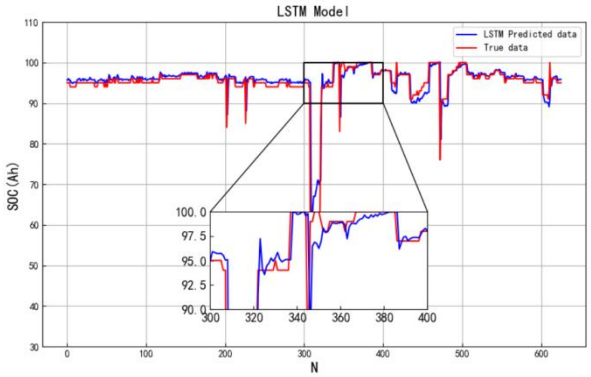

Figure 5. LSTM model prediction chart

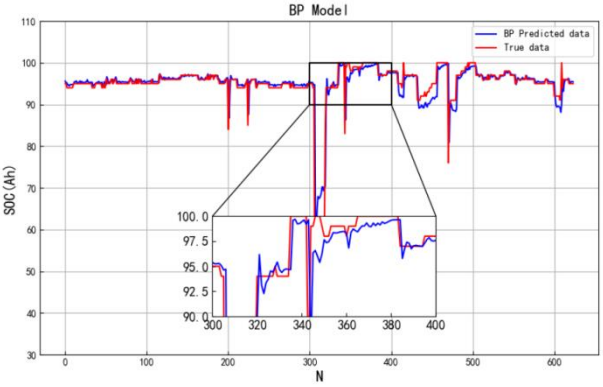

Figure 6. BP model prediction chart

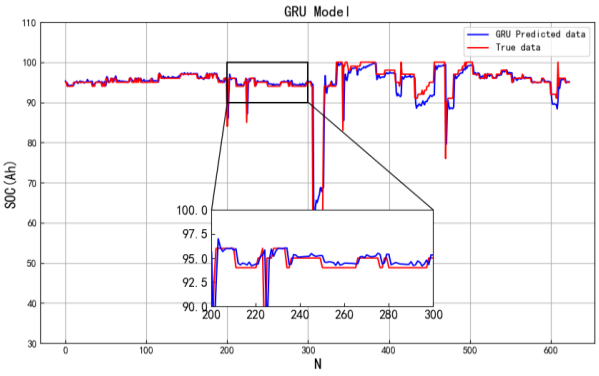

Figure 7. GRU model prediction chart

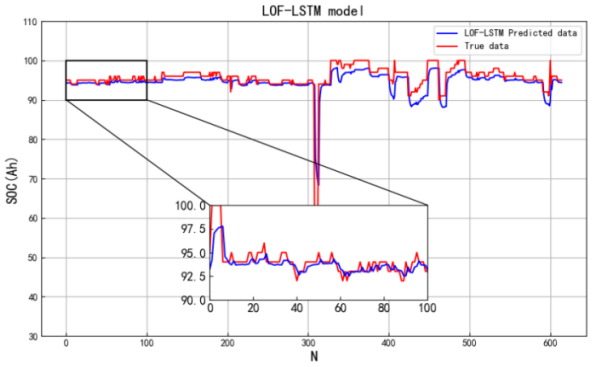

Figure 8. LOF-LSTM model prediction chart

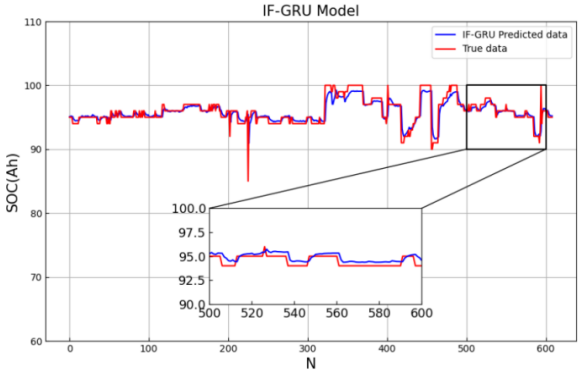

Figure 9. IF-GRU model prediction chart

5. Conclusion

Firstly, in this paper, the isolated forest anomaly detection algorithm is used to detect and pre-process anomalies in the actual data, which effectively solves the problem of mutations in the data and improves the shortcomings of the traditional algorithm with low prediction accuracy and weak immunity. Secondly, the sliding window model is employed to mitigate the impact of abnormal data on the prediction accuracy of the model. Experimental studies are conducted to examine the influence of various window sizes on the results. Finally, the GRU model serves as the foundational experimental model, leading to the development of the novel SOC combined prediction model named IF-GRU, as proposed in this study. Through extensive research based on real data, the main contributions of this research can be summarized in the following three aspects:

- To achieve accurate SOC prediction for lead-acid batteries, this study introduces a prediction model that combines a GRU neural network with the isolated forest anomaly detection algorithm and sliding window method.
- The isolated forest anomaly detection algorithm is used to preprocess the data for the lead-acid battery SOC prediction model. In contrast, the anomaly detection algorithm can swiftly identify anomalous values within a specific range and adapt the anomaly ratio accordingly.
- The accuracy of prediction is improved by changing the size of the sliding window.

We have developed a novel approach described in this research paper that demonstrates significantly reduced errors and improved accuracy in predicting abnormal mutation data. Our model also exhibits enhanced reliability, providing more robust results. To ensure practical applicability, we have utilized real-world data and placed a strong emphasis on practical implementation throughout our study. Moving forward, our research will concentrate on further optimizing our proposed model by building upon the existing framework. We will focus on implementing multi-step prediction as well as achieving real-time online prediction with more advanced models for engineering applications and working on decreasing equipment expenses and enhance estimation accuracy in future work.

References

[1] C. Zou, L. Zhang, X. Hu, Z. Wang, T. Wik, and M. Pecht, "A review of fractional-order techniques applied to lithium-ion batteries, lead-acid batteries, and supercapacitors," J. Power Sources, vol. 390, pp. 286–296, 2018.

[2] S. Dhundhara, Y. P. Verma, and A. Williams, "Techno-economic analysis of the lithium-ion and lead-acid battery in microgrid systems," Energy Convers. Manag., vol. 177, pp. 122–142, 2018.

[3] R. Xiong, J. Cao, Q. Yu, H. He, and F. Sun, "Critical review on the battery state of charge estimation methods for electric vehicles," IEEE Access, vol. 6, pp. 1832–1843, 2018.

[4] X. Hu, L. Xu, X. Lin, and M. Pecht, "Battery lifetime prognostics," Joule, vol. 4, no. 2, pp. 310–346, 2020.

[5] X.-D. Chen, H.-Y. Yang, J.-S. Wun, C.-H. Wang, and L.-L. Li, "Life prediction of lithium-ion battery based on a hybrid model," Energy Explor. Exploit., vol. 38, no. 5, pp. 1854–1878, 2020.

[6] M. Talha, F. Asghar, and S. H. Kim, "A neural network-based robust online SOC and SOH estimation for sealed lead–acid batteries in renewable systems," Arab. J. Sci. Eng., vol. 44, no. 3, pp. 1869–1881, 2019.

[7] R. Zhang et al., "State of the art of lithium-ion battery SOC estimation for electrical vehicles," Energies, vol. 11, no. 7, p. 1820, 2018.

[8] J. Hong, Z. Wang, W. Chen, L.-Y. Wang, and C. Qu, "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, vol. 30, no. 101459, p. 101459, 2020.

[9] D. Zhou, "State of health monitoring and remaining useful life prediction of lithium-ion batteries based on temporal convolutional network," IEEE Access, vol. 8, pp. 53307–53320, 2020.

[10] P. Křivík, "Methods of SoC determination of lead acid battery," Journal of Energy Storage, vol. 15, pp. 191–195, 2018.

[11] R. Zhang et al., "State of the art of lithium-ion battery SOC estimation for electrical vehicles," Energies, vol. 11, no. 7, p. 1820, 2018.

[12] Y. Xu et al., "Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation," J. Hydrol. (Amst.), vol. 608, no. 127553, p. 127553, 2022.

[13] W. Zhang, H. Li, L. Tang, X. Gu, L. Wang, and L. Wang, "Displacement prediction of Jiuxianping landslide using gated recurrent unit (GRU) networks," Acta Geotech., 2022.

[14] A. Gasparin, S. Lukovic, and C. Alippi, "Deep learning for time series forecasting: The electric load case," CAAI Trans. Intell. Technol., vol. 7, no. 1, pp. 1–25, 2022.

[15] A. R. Yuliani et al., "Remaining useful life prediction of lithium-ion battery based on LSTM and GRU," in The 2021 International Conference on Computer, Control, Informatics and Its Applications, 2021.

[16] Y. Lingzhi, Z. Zongguang, and F. Chaodong, "Life prediction of lithium battery based on EEMD-GSGRU," J]. Energy

Storage Science and Technology, vol. 2020, no. 5, pp. 1566–1573.

[17] W. Yaxiong, C. Zhenhang, and Z. Wei, "Lithium-ion battery state-of-charge estimation for small target sample sets using the improved GRU-based transfer learning," Part B, vol. 244, 2022.

[18] P. Karczmarek and A. Kiersztyn, "Witold Pedrycz, et al, K-Means-based isolation forest, Knowledge-Based Systems," vol. 195, 2020.

[19] S. Hariri, M. Carrasco Kind, and R. J. Brunner, "Extended isolation forest," IEEE Transactions on Knowledge and Data Engineering, vol. 33, pp. 1479–1489, 2019.

[20] J. Jiang, T. Li, C. Chang, C. Yang, and L. Liao, "Fault diagnosis method for lithium-ion batteries in electric vehicles based on isolated forest algorithm," J. Energy Storage, vol. 50, no. 104177, p. 104177, 2022.

[21] W. Zhao, Y. Zhang, Y. Zhu, and P. Xu, "Anomaly detection of aircraft lead-acid battery," Qual. Reliab. Eng. Int., vol. 37, no. 3, pp. 1186–1197, 2021.

[22] D. Yang, Y. Wang, R. Pan, R. Chen, and Z. Chen, "A neural network based state-of-health estimation of lithium-ion battery in electric vehicles," Energy Procedia, vol. 105, pp. 2059–2064, 2017.

[23] J. Hong, Z. Wang, W. Chen, L.-Y. Wang, and C. Qu, "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, vol. 30, no. 101459, p. 101459, 2020.

[24] F. Yang, W. Li, C. Li, and Q. Miao, "State-of-charge estimation of lithium-ion batteries based on gated recurrent neural network," Energy (Oxf.), vol. 175, pp. 66–75, 2019.

[25] X. Song, F. Yang, D. Wang, and K.-L. Tsui, "Combined CNN-LSTM network for state-of-charge estimation of lithium-ion batteries," IEEE Access, vol. 7, pp. 88894–88902, 2019.

[26] Z. Shaofeng, Z. Qingyong, and Y. Yesen, "Lithium-ion battery model based on sliding window and long short term memory neural network," J]. Energy Storage Science and Technology, vol. 2022, no. 1, pp. 228–239.

[27] C. Chengcheng, Z. Qian, and H. Mahsa, "Forecast of rainfall distribution based on fixed sliding window long short-term memory," J]. Engineering Applications of Computational Fluid Mechanics, vol. 16, no. 1, pp. 248–261, 2022.

[28] L. Dong et al., "Prediction of streamflow based on dynamic sliding window LSTM," Water (Basel), vol. 12, no. 11, p. 3032, 2020.

[29] Y. Liu et al., "Bidirectional GRU networks-based next POI category prediction for healthcare," Int. J. Intell. Syst., vol. 37, no. 7, pp. 4020–4040, 2022.

[30] B. Yang et al., "Motion prediction for beating heart surgery with GRU," Biomed. Signal Process. Control, vol. 83, no. 104641, p. 104641, 2023.