Journal of Electronic Research and Application, 2025, Volume 9, Issue 6
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Health Status Prediction Of Lithium Batteries Based On Deep Learning

Hai-Rui Zhang*, Yueling Zhao, Yan-Bo Xue

School of Electric Engineering, Liaoning University of Technology, Jinzhou 121001, Liaoning, China

**Author to whom correspondence should be addressed.*

Abstract: Aiming at the shortcomings of traditional State of Health (SOH) prediction methods in nonlinear modeling and temporal dependence handling, this paper proposes a hybrid CNN-GRU model integrated with the Dung Beetle Optimization (DBO) algorithm (denoted as DBO-CNN-GRU) for lithium battery SOH prediction. Indirect health factors strongly correlated with SOH are extracted from the NASA public dataset, and their effectiveness is verified using Pearson and Spearman correlation coefficients. A CNN-GRU model is designed: the convolutional neural network (CNN) is used to capture local features, and the gated recurrent unit (GRU) is combined to model the temporal dependence of capacity degradation. Furthermore, the DBO algorithm is introduced to optimize the model's hyperparameters, enhancing the global search capability. Experiments show that the DBO-CNN-GRU model achieves significantly better test performance on the NASA dataset than the single CNN, GRU, and LSTM models.

Keywords: Lithium battery health status; Convolutional Neural Network (CNN); Dung Beetle Optimization (DBO) algorithm

Online publication: December 16, 2025

1. Introduction

With the in-depth development of new energy vehicles, researchers have gradually realized that the State of Health (SOH) of lithium-ion batteries plays a decisive role in maintaining the stability of vehicle power output [1]. Lithium-ion batteries are highly favored in fields such as energy storage and electric vehicles due to their high specific capacity and excellent cycle performance. However, during repeated charging and discharging processes, the performance and capacity of lithium-ion batteries will continuously degrade, which may lead to problems such as thermal runaway, battery swelling, and shortened cycle life. Therefore, accurate evaluation of battery health status is crucial to ensuring the safe and stable operation of electric vehicles and energy storage systems.

In the field of battery health status assessment, the physical model method explores the internal working mechanism of batteries in depth by constructing complex physical and chemical models [2]. Common intelligent algorithms in data-driven methods include Long Short-Term Memory (LSTM), Support Vector Regression (SVR),

Gaussian Process Regression (GPR), and Gated Recurrent Unit (GRU) [3]. Nevertheless, single models often fall short in terms of accuracy and generalization performance, which has prompted researchers to gradually shift their focus to exploring algorithm fusion strategies. Zhang *et al.* proposed an improved LSTM model, which exhibited significant advantages in prediction accuracy and robustness compared with single models [4]. To consider the impact of different temperatures on lithium-ion batteries, Yuan *et al.* conducted battery discharge experiments at different temperatures, estimated the battery SOH using a Back Propagation (BP) neural network, and verified the generalization ability of this method [5]. Although algorithm fusion has improved the accuracy and generalization of predictions, it has also increased the number of model parameters and complicated the model input. When the model is used to predict the battery capacity regeneration region and handle data nonlinearity issues, problems such as vanishing gradients and exploding gradients inevitably occur [6].

In response to the above analysis, this study takes lithium-ion batteries as the research object and uses two datasets with different working conditions, namely the NASA dataset and the dataset from the Center for Advanced Life Cycle Engineering (CALCE) of the University of Maryland, to avoid experimental randomness. The study selects and fuses Convolutional Neural Network (CNN) and GRU models, which can process multi-scale information and capture nonlinear changes, thereby making up for the limitations of single models. On this basis, the Dung Beetle Optimizer (DBO) algorithm is used for hyperparameter optimization of the model. Leveraging the global search feature of DBO, the model is prevented from falling into local optimal solutions, ultimately achieving high-precision SOH prediction of lithium-ion batteries.

2. DBO-CNN-GRU model

2.1. CNN-GRU model

The core idea of CNN is to extract local features of input data through convolutional layers and pooling layers, and then perform classification or regression of overall information via fully connected layers [7]. As an improved architecture of recurrent neural networks, GRU selectively forgets the memory of historical information through the synergistic effect of update gates and reset gates. It simplifies the triple gating structure of LSTM into a dual gating design, which effectively reduces model complexity. This dynamic adjustment mechanism endows the model with long-term memory capability [8]. In time series prediction tasks, this network can not only capture the long-term dependencies of sequence data but also avoid the overfitting problem through its simplified structure, demonstrating higher computational efficiency and model generalization ability [9].

The CNN-GRU model achieves the complementary advantages of the two networks through phased feature processing. The raw feature data of lithium-ion batteries is input into the CNN module, which utilizes local receptive fields to explore the spatial correlation between features. The max-pooling layer compresses the feature dimension through downsampling, enhances the model's translation invariance and noise robustness, and strips off redundant information, thereby ensuring the accuracy and effectiveness of subsequent analysis [10]. Specifically, the update gate is responsible for controlling the retention ratio of historical state information; the reset gate, on the other hand, determines the degree to which the current input modifies the historical information. This enables the model to more accurately capture the dynamic laws of feature evolution during battery aging. **Figure 1** shows the structure diagram of the CNN-GRU neural network.

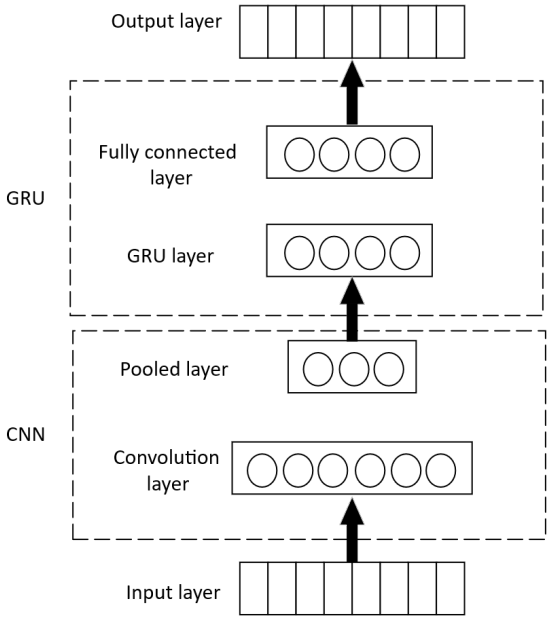

Figure 1. Result diagram of the CNN-GRU model.

2.2. DBO algorithm

DBO is a meta-heuristic optimization algorithm inspired by the swarm intelligence behavior of dung beetles. It achieves efficient parameter optimization by simulating the characteristics of dung beetles' ecological behaviors [11]. DBO divides dung beetle behaviors into four stages as shown below.

2.2.1. Ball rolling stage (global exploration)

Dung beetles determine the direction of ball rolling through celestial navigation (e.g., polarized light). In the algorithm, this behavior is simulated using dynamic direction factors and deflection angles. Position update formula:

$$x_i^{t+1} = x_i^t + \alpha \cdot \kappa \cdot x_i^{t-1} + \beta \cdot \triangle x \cdot \tan(\theta) \quad (1)$$

In the formula, x_i^t represents the current individual position; α is the direction random factor (taking a value of 0 or 1); k is the ball rolling attenuation factor (decreasing as the number of iterations increases); β is the disturbance coefficient; and θ is the deflection angle (controlling the diversity of search directions), which helps avoid falling into local optima.

2.2.2. Breeding stage (local exploitation)

The regional boundary definition is shown below:

$$L_b^* = \max(x^* - R, L_b), U_b^* = \min(x^* + R, U_b) \quad (2)$$

x^* represents the current optimal solution; R is the dynamic radius (decreasing with iterations); L_b and U_b are the original boundaries of the solution space. Offspring individuals are randomly generated within these boundaries to enhance local search capability.

2.2.3. Young dung beetle stage (dynamic search)

Hatched young dung beetles explore the surrounding environment through random walks. The position update formula is:

$$x_i^{t+1} = x_i^t + C \cdot (x_i^t - L_b^*) + D \cdot (x_i^t - U_b^*) \quad (3)$$

Among them, C and D are random numbers following the Lévy distribution, which simulate jump search behaviors to balance exploitation and exploration.

2.2.4. Stealing stage (competition mechanism)

Some dung beetles snatch dung balls from other individuals. In the algorithm, the optimal solution is updated through random competition:

$$x_i^{t+1} = x^* + S \cdot g \cdot (|x_i^t - x_i^t| + |x_k^t - x_i^t|) \quad (4)$$

Among them, S represents the step size factor, g is the Gaussian disturbance term, x_i and x_k are randomly selected individuals. This mechanism avoids premature convergence and improves population diversity.

2.3. DBO-CNN-GRU battery SOH prediction model

CNN is used for feature extraction of the input data. It captures local features in the data through convolutional operations, resulting in a set of feature sequences. Then, the feature sequences extracted by CNN are input into the GRU network. The GRU network models the sequence data via reset gates and update gates, captures the long-term dependencies in the data, and obtains an updated hidden state sequence. Next, the attention mechanism is introduced to calculate the attention weights between the GRU hidden states and input features. These weights are used for weighted fusion of the features extracted by the CNN and GRU networks, generating a context vector containing key information. This enables the model to automatically learn and focus on the features that are most critical for the prediction task. Finally, by simulating the survival behaviors of dung beetles, the model continuously adjusts parameters to find the optimal solution, thereby improving the overall model performance and prediction accuracy.

3. Experimental data and evaluation methods

3.1. Extraction of health factors

The lithium-ion battery data used in this experiment is derived from the B0005, B0006, and B0007 batteries in the NASA dataset, and the CS2_36 battery in the University of Maryland (CALCE) dataset. Taking the B0005 battery in the NASA dataset as an example, the current-voltage curves and feature factors (HF1~HF4) under different charge-discharge cycles were extracted. The duration of the constant current charging stage was selected as the first indirect health indicator (HI1). The time interval for the voltage to rise from 3.9V to 4.2V during the constant current charging stage was set as the second indirect health indicator (HI2). The constant voltage drop discharge time (with voltage decreasing from 3.8V to 3.4V) was taken as the third indirect health indicator (HI3). The time required for the surface temperature to rise to its peak value was defined as the fourth health indicator (HI4).

3.2. Correlation analysis

Through the comparison of the change trajectories of feature factors and capacity degradation trajectories, Pearson

and Spearman correlation coefficients were used to quantify their correlation. The Pearson correlation coefficient quantifies and analyzes the degree of linear correlation between two variables by examining the covariance and standard deviation of raw data. It requires the data to follow a normal distribution and the relationship between variables to be approximately linear, but it is sensitive to outliers. The Spearman correlation coefficient, on the other hand, is based on the rank ordering of data. It evaluates the monotonic correlation between variables (i.e., the trend of variables changing in the same direction or opposite directions) and is not limited to linear relationships, thus exhibiting stronger robustness.

$$Pearson = \frac{E(XY)-E(X)E(Y)}{\sqrt{E(X^2)-E^2(X)}\sqrt{E(Y^2)-E^2(Y)}} \tag{5}$$

$$Spearman = \frac{\sum_i (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_i (x_i-\bar{x})^2}\sqrt{\sum_i (y_i-\bar{y})^2}} \tag{6}$$

In Equations (5) and (6), X represents the feature factor, Y denotes the population of capacity samples, while x_i and y_i represent the samples. As seen from **Table 1** and **Table 2**, the Pearson correlation coefficients between the health factors extracted in this study and capacity all exceed 0.9, indicating that the extracted health factors exhibit a high degree of linear correlation with capacity.

Table 1. Correlation coefficients of B0005 battery

Health factor	Pearson correlation coefficient	Spearman correlation coefficient
Constant current charging time (HI1)	0.997	0.993
Constant voltage rise charging time (HI2)	0.997	0.995
Constant voltage drop discharging time (HI3)	0.992	0.99
Time to reach maximum temperature during charging (HI4)	0.992	0.989

Table 2. Correlation coefficients of B0006 battery

Health factor	Pearson correlation coefficient	Spearman correlation coefficient
Constant current charging time (HI1)	0.994	0.995
Constant voltage rise charging time (HI2)	0.995	0.997
Constant voltage drop discharging time (HI3)	0.991	0.998
Time to reach maximum temperature during charging (HI4)	0.987	0.968

3.3. Evaluation metrics

To quantify the model's prediction accuracy from multiple dimensions, three key metrics are used for comprehensive evaluation: Mean Absolute Error (MAE) measures the average deviation between predicted values and true values, Mean Absolute Percentage Error (MAPE) reflects the level of relative error, and Root Mean Square Error (RMSE) pays more attention to the impact of larger errors on the overall performance. Their formulas are shown in Equations (7)–(9):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{7}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |(y_i - \hat{y}_i)| \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{9}$$

4. Experimental verification and analysis

Two sets of experiments are designed in this study. The first set of experiments is conducted on the NASA dataset, where the proposed model in this study is compared with combined models and single models to verify the effectiveness of the proposed model's combination, and the specific simulation results are shown in **Figure 2**. The second set of experiments is carried out on the University of Maryland CS2_36 dataset to verify whether the proposed model has generalization ability, and the experimental results are presented in **Table 3**.

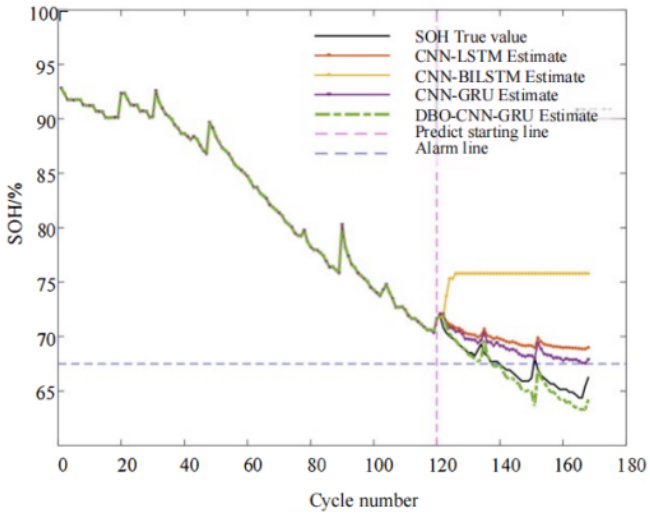

Figure 2. Comparison of multi-model prediction curves for lithium-ion battery SOH prediction model

Table 3. Prediction results on the CALCE dataset

SOH Methods	MAE	RMSE	
DBO-CNN-GRU	0.0148	0.0185	0.9920
CNN	0.0222	0.0287	0.9808
CNN-GRU	0.0185	0.0272	0.9828
LSTM	0.0258	0.0302	0.9789
BILSTM	0.025	0.0342	0.9729

5. Conclusion

This study proposes the use of the DBO algorithm to automatically search for globally optimal parameters. By simulating dung beetles' ball rolling, breeding, dynamic search, and competition mechanisms, the algorithm balances global exploration and local exploitation capabilities, thereby improving SOH prediction accuracy (with

MAE as low as 1.48% and exceeding 99%). On top of that, comparative experiments verify the superiority of the DBO-CNN-GRU model: it exhibits strong generalization capability across datasets (with a cross-dataset MAE increase of only 0.6%); its MAE and RMSE are 33%–54% lower than those of traditional models such as CNN, LSTM, and BiLSTM; and it can effectively capture the nonlinear relationships between capacity degradation and multimodal data. Therefore, the DBO-CNN-GRU method is of great significance for improving the accuracy of SOH prediction and provides methodological support for battery maintenance.

Disclosure statement

The author declares no conflict of interest.

References

[1] Zhu Z, 2023, Research on SOH Estimation and RUL Prediction Methods of Electric Vehicle Lithium-Ion Batteries Based on Deep Learning, thesis, Qingdao University of Science and Technology.

[2] Lu H, 2019, Theoretical Study on Polymer Electrode Materials for Lithium-Ion Batteries, thesis, Beijing University of Technology.

[3] Geng M, Fan M, Wei B, 2025, SOH Estimation of Energy Storage Batteries Based on Fragmented Data. Global Energy Interconnection, 8(1): 57–66.

[4] Zhang C, Chen Y, Liu M, et al., 2025, A State of Health Estimation Method for Lithium-Ion Batteries using ICA-T Features and CNN-LA-BiLSTM. Energy Storage Science and Technology, 14(3): 1–17.

[5] Yuan J, Liu D, 2025, Research on Rapid Sorting Model for Health Status of Retired Power Lithium-Ion Batteries Based on GA-BP Neural Network. Technology Innovation and Application, 15(3): 29–32.

[6] Wang Q, Sheng Q, Liu W, 2025, Research on the Improvement and Application of State Estimation Algorithm for Lithium Iron Phosphate Batteries in Low-Temperature Environments. Journal of Physics: Conference Series, 2941(1): 12023–12029.

[7] Patrizi G, Canzanella F, Ciani L, 2024, Towards a State of Health Definition of Lithium Batteries through Electrochemical Impedance Spectroscopy. Electronics, 13(8): 22–35.

[8] Braun A, Behmann R, Chabrol D, 2024, Single-Cell Operando SOC and SOH Diagnosis in a 24V Lithium Iron Phosphate Battery with a Voltage-Controlled Model. Journal of Energy Storage, 2024: 10986–10997.

[9] Runze W, Junfu L, Xinyu W, 2024, Deep Learning Model for State of Health Estimation of Lithium Batteries based on Relaxation Voltage. Journal of Energy Storage, 2024: 10189–10199.

[10] Wu X, 2024, Joint Estimation of State of Charge and State of Health of Lithium-Ion Batteries Based on Filters, thesis, Guangxi University.

[11] Du L, Li Z, Li Z, et al., 2025, DBO-CNN-BiLSTM-AM Short Term Photovoltaic Power Generation Prediction based on Feature Optimization. Renewable Energy Resources,43(11): 1–11