


Review

A Review of SOH Prediction of Li-Ion Batteries Based on Data-Driven Algorithms

Ming Zhang ¹, Dongfang Yang ², Jiaxuan Du ³, Hanlei Sun ^{1,*}, Liwei Li ⁴, Licheng Wang ⁵ and Kai Wang ^{1,*} 

¹ School of Electrical Engineering, Weihai Innovation Research Institute, Qingdao University, Qingdao 266000, China

² Xi'an Traffic Engineering Institute, Xi'an 710300, China

³ Electrical Engineering and Automation, Northeast Electric Power University, Ji'lin 132012, China

⁴ School of Control Science and Engineering, Shandong University, Jinan 250100, China

⁵ School of Information Engineering, Zhejiang University of Technology, Hangzhou 310000, China

* Correspondence: 2020025742@qdu.edu.cn (H.S.); wangkai@qdu.edu.cn (K.W.)

Abstract: As an important energy storage device, lithium-ion batteries (LIBs) have been widely used in various fields due to their remarkable advantages. The high level of precision in estimating the battery's state of health greatly enhances the safety and dependability of the application process. In contrast to traditional model-based prediction methods that are complex and have limited accuracy, data-driven prediction methods, which are considered mainstream, rely on direct data analysis and offer higher accuracy. Therefore, this paper reviews how to use the latest data-driven algorithms to predict the SOH of LIBs, and proposes a general prediction process, including the acquisition of datasets for the charging and discharging process of LIBs, the processing of data and features, and the selection of algorithms. The advantages and limitations of various processing methods and cutting-edge data-driven algorithms are summarized and compared, and methods with potential applications are proposed. Effort was also made to point out their application methods and application scenarios, providing guidance for researchers in this area.

Keywords: LIB; SOH; data-driven algorithms; data processing



Citation: Zhang, M.; Yang, D.; Du, J.; Sun, H.; Li, L.; Wang, L.; Wang, K. A Review of SOH Prediction of Li-Ion Batteries Based on Data-Driven Algorithms. *Energies* **2023**, *16*, 3167. <https://doi.org/10.3390/en16073167>

Academic Editors: Carlos Miguel Costa, Vincenzo Spallina and Qingsong Wang

Received: 31 October 2022

Revised: 26 March 2023

Accepted: 28 March 2023

Published: 31 March 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

With the substantial increase in productivity and the consumption of power in the modern era, the demand for energy is increasing. This has promoted the further development of the energy and power industry [1]. As one of the basic energy storage devices with the widest application coverage, LIBs have been becoming an important storage device in the energy industry. However, one of the most serious challenges for LIBs applications, such as in power grids, electric vehicles, mobile phones, etc., is degradation, which has a crucial impact on LIBs cycling life. However, battery degradation is a complex process combining internal reactions with external environments. The internal reactions include electrochemical reactions and side reactions. Lithium iron phosphate (LiFePO_4) and commonly used negative electrode materials serve as positive electrode materials in the primary reactions of lithium-ion batteries (LIBs). The main chemical reaction of LIBs involves graphite (C), with side reactions including salt decomposition, electrolyte oxidation, and active material dissolution. The degradation mechanism of LIBs is illustrated in Figure 1 [2] and can be classified into two main modes [3,4]: (1) reduction in lithium-ions due to the continued consumption by side reactions, and (2) reduction in active materials leading to decreased capacity of LIBs. More specifically, the loss of active material is mainly caused by graphite peeling, binder decomposition, electrical contact loss caused by current collector corrosion, and electrode particle breakage [5]. The reduction in lithium-ions is mainly caused by three side reactions of solid electrolyte interface (SEI) film formation and decomposition reaction, electrolyte decomposition reaction, and electroplating reaction. It is worth mentioning that

these degradation mechanisms are highly related to materials. For example, the working voltage of a graphite anode is lower than that of a common electrolyte, resulting in the formation of an SEI film [6,7]. However, no SEI film is formed in lithium titanium oxide (LTO) anodes because the potential of LTO is within the electrochemical threshold of the electrolyte [8]. Another example is that the volume change in the lithium iron phosphate (LFP) positive electrode is smaller than that of the Lithium manganese oxide (LMO) positive electrode, so its structural deformation is also smaller [9]. In addition to the difference in materials, the degradation mechanism is also very different under different operating conditions and different battery designs. For example, the lithium-ion plating reaction has a high probability of occurrence during rapid charging, but rarely occurs during discharge [10]. For battery design, a smaller cathode particle size will result in less stress and therefore less particle breakage, but due to the high specific surface area, it will also cause more cathode material to dissolve [11].

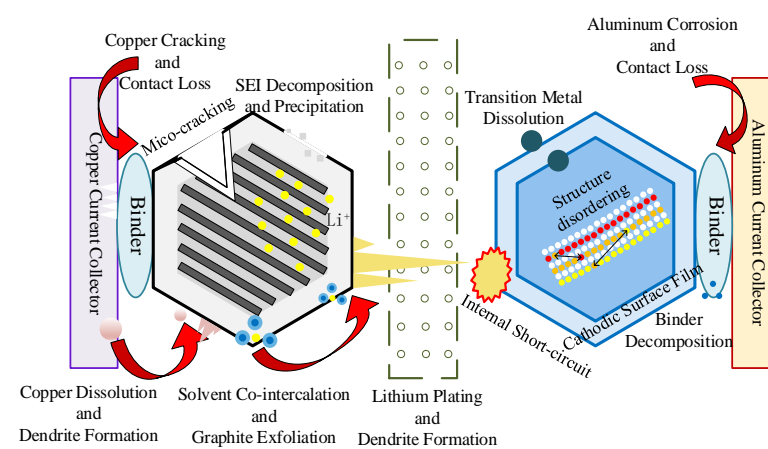


Figure 1. LIB internal activities.

At present, the prediction methods of LIB mainly include model-based driving methods and data-based driving methods [12]. The model-driven method is mainly based on the complex internal physical model of LIB established partially based on experience, and transformed into mathematical problems, using the method of learning modeling to represent the degradation process of LIB, such as the Kalman filter method [13], the particle filter method [14], and other non-linear model prediction methods.

With the advent of the era of big data, thanks to the powerful computing power and robustness of computers, data-driven machine learning (ML) and deep learning (DL) have become increasingly important tools. For example, such tools include a backpropagation neural network (BPNN) [15,16], a support vector machine (SVM) [17], a long short-term memory (LSTM) neural network [18–20], etc. Data-driven machine learning methods have the following characteristics:

1. There is no need to have too much understanding of the internal mechanism of the subject matter, and various parameters and hyperparameters that are highly correlated with the results need to be extracted;
2. General ML and DL need a large amount of data as support, as a method that does not need to establish an internal mechanism model, a small amount of data alone cannot support ML for accurate model establishment [21];
3. The quality of the data are generally considered to be the reason that hinders ML from further improving the prediction accuracy. Due to the inevitable factors of actual measurement, the data are full of noise, which will severely limit the offline training of ML models.

As an indicator of the degree of battery degradation, its SOH is not clearly defined. It is generally believed that the SOH of a LIB can be calculated from the ratio of the current maximum capacity to the initial capacity [22]. The mathematical formula is:

$$SOH = \frac{C_t}{C_0} \quad (1)$$

Among them, C_t is the current remaining capacity of the battery, and C_0 is the rated capacity.

Therefore, in the process of estimating the SOH of LIB, the following main issues need to be discussed and analyzed:

1. Obtaining an accurate and large number of LIB charge and discharge datasets;
2. Performing mathematical analysis to extract features that have a higher degree of correlation with SOH, and extracting a large amount of data related to features;
3. Data processing of battery characteristics to reduce fluctuations and noise;
4. Choosing a suitable machine learning model which takes into account accuracy and efficiency.

By analyzing data-driven ML and DL, this paper initially discusses the issue with lithium-ion battery (LIB) datasets in chapter two. Subsequently, it evaluates and compares different techniques for reducing data noise in the latter half, while dedicating significant attention to the selection of the most suitable method. The third chapter, which is the central focus of the paper, utilizes features to predict the state of health (SOH). It analyzes the algorithm principle and summarizes the latest research findings, discussing the advantages and drawbacks of common data-driven algorithms and offering improvement directions. This provides readers with the latest guidance on LIB SOH prediction.

2. Data Acquisition and Processing

2.1. LIB Dataset

Researchers can conduct a comprehensive training and verification of SOH prediction by performing an aging test on the battery, which necessitates a large amount of data, experimental testing, and data monitoring throughout the cycle, serving as a key element in machine learning (ML) support. However, this experiment requires strict experimental conditions, an appropriate environment, and a significant amount of time. Hence, based on the results of the aging test, researchers can effectively complete SOH prediction using the following datasets.

NASA provides a very complete set of battery datasets that are popular with scholars [23]. Sufficient data were collected through regular aging experiments and were made available online. Among them, B0005-B0007 and B0018 are the most commonly used datasets. These datasets represent four sets of batteries charged to 4.2 V in a constant current–constant voltage (CC-CV) mode of 1.5 A at a room temperature of 24 °C, and the CC mode of 2 A discharges to 2.7 V, 2.5 V, 2.2 V and 2.5 V, respectively. The cells were cycled until the end of life (EOL) criteria—a 30% drop in rated capacity—was observed, and the effect of aging on internal parameters was observed and recorded.

Another dataset that is widely used in forecasting research is the Calce dataset [24]. The CS2 dataset includes six sub-datasets. Cyclic aging conditions are also CC–CV charge and CC discharge. In this dataset, during charging, the battery is charged at a constant current of 0.5C (c-rate, used to indicate the magnitude of the current when the battery is charged and discharged) until the voltage reaches 4.2 V, and then it charges at a constant voltage of 4.2 V until the current drops below 0.05 A. During discharge, the battery discharges with varying constant currents until the voltage drops to 2.7 V. Furthermore, the battery is cycled to the EOL standard, and through the use of an ARBI battery tester, various data are measured and recorded in the experiment.

Both datasets described above follow the same charging protocol, but the charging current is adjusted, leading to variations in parameters due to changes in experimental

conditions, while the nature of the datasets remains similar. These datasets can be utilized in research for predicting SOH, and are recommended as a basis for feature extraction and data processing, which can then be applied to various efficient algorithms.

This article takes 18650LIB as an example, where the cells had a rated capacity of 2.7 Ah. The nominal voltage was 3.6 V, with upper and lower cut-off voltages of 4.2 V and 2.5 V, respectively. The materials consisted of graphite on the anode and $\text{Li}(\text{NiCoAl})\text{O}_2$ on the cathode. The cell specifications are summarized in Table 1.

Table 1. 18650 battery information introduction.

Parameter	
Rated Capacity	2.7 Ah
Material	$\text{Li}(\text{NiCoAl})\text{O}_2/\text{Carbon}$
Voltage range	2.5–4.2 V
Charging Temperature Range	10–45 °C
Discharge temperature Range	−20–60 °C

2.2. Data Analysis

The core and foundation of data-driven ML and DL algorithms is the selection of input features. The correlation between features and output, their own smoothness, and the tensor of data all affect the fitting rate and accuracy of the model to a large extent. Figure 2a shows the constant current–constant voltage charging curves of LIB under different cycle times. From Figure 2a, part of the aging law of LIB can be seen, as with the increase in the battery cycle life, the charging time of the CC phase showed a decreasing trend, whereas the duration of the CV phase showed an increasing trend. Since there is a large correlation between the battery charging curve and cycle life, it is helpful to select the characteristics related to Li-ion battery SOH through the relevant data of the charging curve.

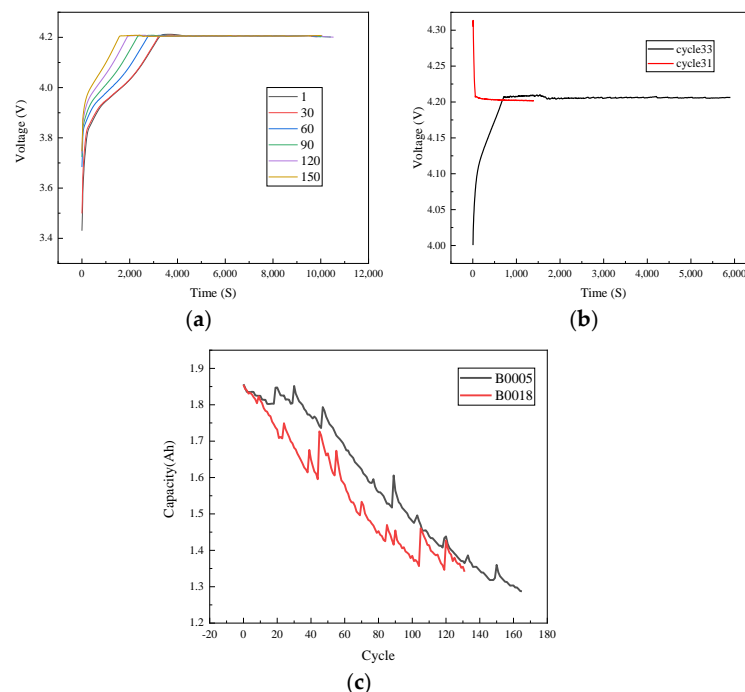


Figure 2. (a) Charging voltage curve for different cycles; (b) Charging voltage curve and discharge voltage curve; (c) Capacity decline curves for different data sets.

A. Data noise reduction

After the original data are extracted, Z-score standardization is carried out uniformly [25], and the mean and standard deviation of the data are used to eliminate the

dimension of the data and make it conform to the standard normal distribution (mean is 0, standard deviation is 1) to greatly reduce the computational burden.

After the data are normalized, noise processing needs to be performed. By observing Figure 2b, it can be seen that there will be erroneous data with large errors in the cycle data, similar to cycle 31, and the voltage data of each cycle has regeneration and strong errors. In a similar manner, just as local fluctuations exist, features extracted from the original data also possess noise and fluctuations, while the global trend of capacity degradation is accompanied by corresponding regeneration and fluctuations. Such abnormal fluctuations caused by poor data quality have a great impact on the accuracy of the model. In order to solve such problems, a variety of methods have been proposed and employed, as follows.

Liu Kailong et al. used empirical mode decomposition (EMD) to deal with such data errors [26]. EMD is an effective signal processing technique that has been applied in many practical fields (e.g., ocean waves, rotating machinery) due to its powerful ability to extract low- and high-frequency components from highly dynamic signals. Using EMD through an iterative screening process, a non-stationary dataset can be decomposed into a sequence of residuals and a series of intrinsic mode functions (IMFs), which represent orthonormal basis components. Specifically, the IMF necessitates satisfying various criteria: (1) the count of zero crossings should be the same or at most vary from the count of extrema for the entire dataset; (2) at any given point, the envelope characterized as a local extrema should produce a mean of zero. Since regeneration phenomena and local fluctuations can be regarded as high-frequency signals, while the global trend of capacity degradation is a low-frequency signal, the EMD technique is employed to break down the initial capacity degradation dataset into several IMFs and a residual. Figure 3a is IMF and a residual curve after EMD decomposition.

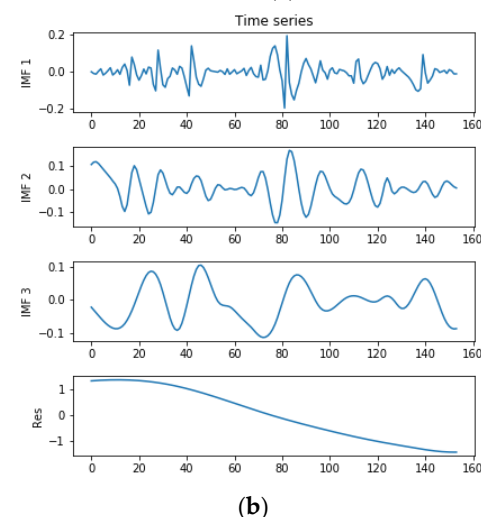
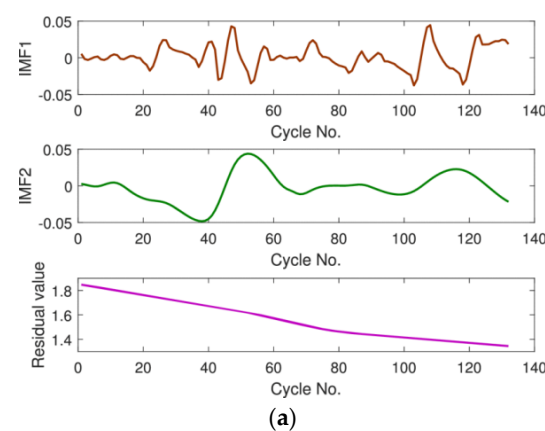


Figure 3. (a) EMD Decomposition Example 1; (b) EMD Decomposition Example 2.

Zhang Chaolong et al. believe that the measured capacity data of batteries are often polluted by noise [27]. Experimenting with noise-contaminated data fails to make accurate capacity predictions. Therefore, in order to extract noise-free data, the volume data needs to be processed. The EMD denoising method is used to solve the problem in the work, and the image after EMD processing is shown in Figure 3b.

After data preprocessing, She Chenqi et al. discovered that the characteristic curve was significantly influenced by imperfect voltage measurement, resulting in noticeable ripple [4]. To effectively extract the characteristic curve, they employed a filtering algorithm, specifically the Gaussian window (GW) filtering method, to smooth out the curve. This filtering method is an improvement on the traditional moving average (MA) filter, which treats all sampled data equally when computing the estimated value and is unable to identify sampling anomalies that may cause significant estimation deviations. The GW filter algorithm overcomes this limitation by employing a Gaussian distribution-shaped weighted average of the data to derive the estimated value, effectively addressing the problem of sampling anomalies.

Zhang Tingting et al. believed that traditional filtering methods mainly include linear filtering methods and nonlinear filtering methods, such as median filtering and Wiener filtering [28]. The traditional filtering method can only transform all the signals to the time domain or all to the frequency domain, which cannot describe the non-stationary characteristics of the signal, and cannot obtain the correlation of the signal. Wavelet transforms have good time domain and frequency domain characteristics, and has the advantages of low entropy, multi-resolution, wavelet basis selection diversity, and de-correlation. Therefore, the wavelet transform is widely used in the field of noise reduction. The wavelet noise-reduction process is as follows:

(1) Decompose the noisy signal into high-frequency and low-frequency signals using a wavelet basis function. Typically, low-frequency signals are real signals, while high-frequency signals are noise signals. They used the db5 wavelet as the wavelet basis function, which is a commonly used wavelet transform due to its simple calculation. (2) Threshold the high-frequency signal. (3) Reconstruct the signal using wavelet transform and output the denoised signal. The original signal is decomposed into three layers using db5, and heuristic thresholding and soft thresholding functions are used. The db5 wavelet was used to denoise the B5 and B6 battery data. The wavelet denoising did not destroy the original features of the data, and the denoised data retained the peak points of the original data. Compared with the original data, the denoised capacity decay data are smoother. Wavelet denoising removes the noise signal in the original data while preserving the authenticity of the original data to the greatest extent.

Zhang Zhi-liang and others also believed that WTM provides localization characteristics in both time and frequency domains [29]. Compared with traditional Fourier transform, which can only give frequency domain information, it is an effective method for analyzing non-stationary signals. method. After sampling the noise current and voltage signals, the decomposition can be done using a $2n$ order WTM. During the denoising process, the obtained wavelet coefficients are adjusted according to the threshold rule. Then, using these denoised wavelet coefficients, the denoised current and voltage signals are reconstructed with a $2n$ -order inverse WTM (IWTM). Finally, these denoised signals are used in the following computations.

The above data processing methods include EMD, GW filter, and wavelet analysis. All of them deal with the noise that will appear in the data. They are relatively good data filtering methods. For datasets of different data quality, you can try to select the appropriate method for many times., the specific data processing results can be observed according to the curve trend of the output image, and the data quality can be judged based on experience, and its quality directly affects the calculation accuracy.

B. Feature extraction

After the data noise reduction process, the obtained data are relatively smooth. At this time, it is necessary to select the features with high correlation with the SOH of LIB, and

then perform correlation analysis for screening, and select the features with high correlation coefficient for model establishment.

Xiao Bin et al. selected 6 features according to the geometric features of the charging curves under different cycles [30], as shown in Figure 4a: F1 is the number of cycles, indicating the number of cycles of battery charge and discharge. F2 is the duration of the CC phase. F3 is the duration of the CV phase. F4 is the duration of the CC phase at 3.9 V, F5 is the duration of the CC phase at 4.0 V, and F6 is the duration of the CC phase at 4.1 V. Feature selection was performed using grey relational analysis (GRA). It's a grey system-based theory that attempts to measure how high or low a relationship is based on the similarity between factors. GRA provides quantitative measurements of system evolution and is ideal for dynamic process analysis.

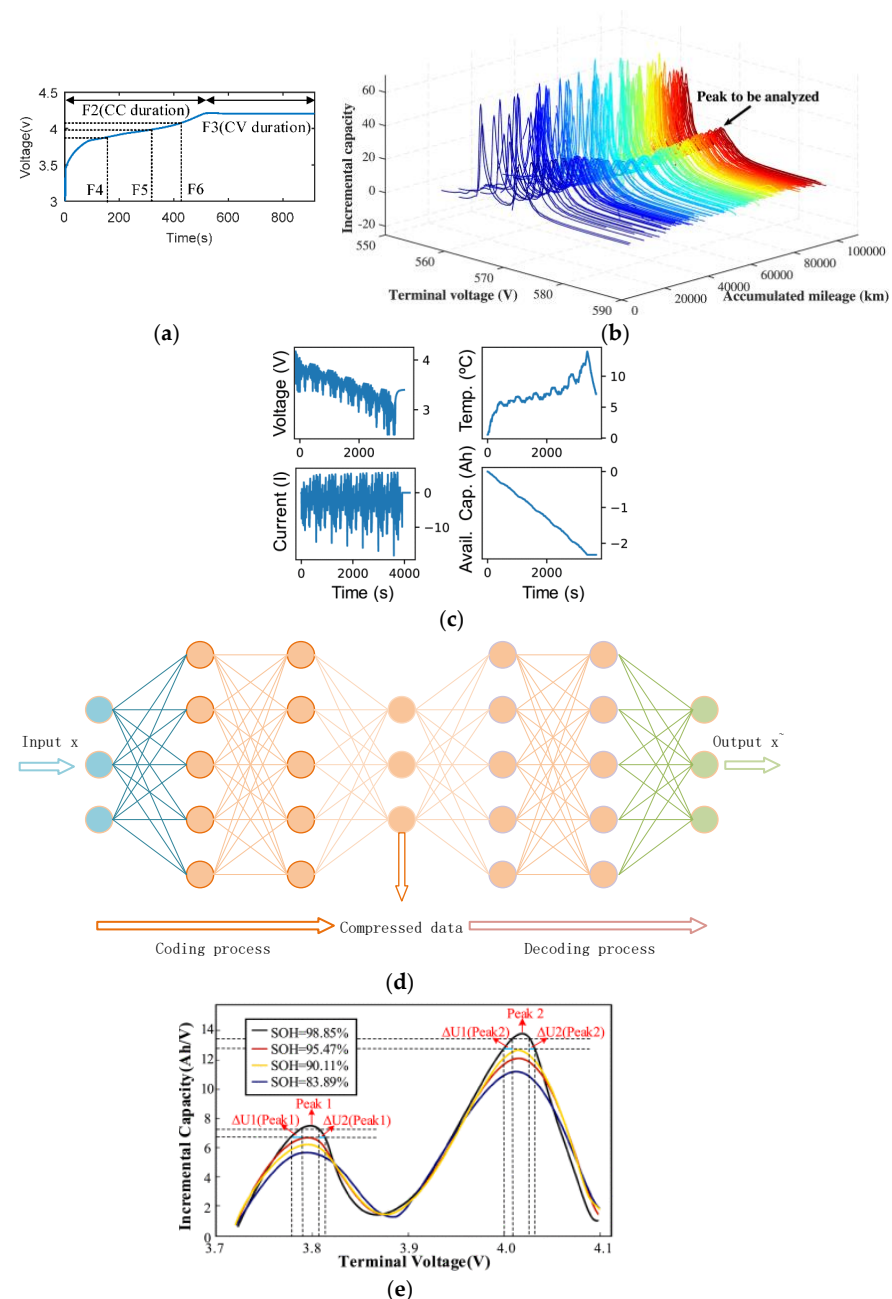


Figure 4. (a) Feature selection of charging voltage curve; (b) Comparison of IC curves under different cycles; (c) Basic charging and discharging characteristics of lithium-ion batteries; (d) Basic structure of AutoEncoder; (e) Feature Selection of IC Curves.

This intuitively obtained feature based on the charging voltage curve is quite universal. This method is relatively intuitive, but at the same time lacks the multi-dimensionality of the feature. Even if four features with a correlation coefficient of about 0.9 are selected according to the correlation, it is limited to the charging voltage. The curve itself improves the singleness and instability in the DL process, which is not beneficial to improving the accuracy and generalization of the model.

She Chengqi et al. proposed to apply the incremental capacity analysis (ICA) method to the feature extraction of LIB [4]. The IC curve derived from dQ/dV calculation is shown in Figure 4b. ICA deeply excavated and quantitatively analyzed the LIB during charging. The relationship between voltage and charge during discharge. During the charging process of a LIB, the open circuit voltage (OCV) has a relatively level section during the rising process. In this section, as the charged power increases, the internal voltage of the battery changes slowly, which is called a voltage plateau. It can be seen from the Figure 4b that as the number of cycles increases, the voltage platform will gradually shift upward. After optimizing the curve through SVR and filtering the noise with GW filter, the first peak value and its peak position of each cycle are extracted as the input of the battery feature.

This method obtains new strong correlation features through ICA based on the charging curve of the battery, and analyzes the transformation process of LIB charging to extract features, which helps to improve the accuracy of the model, but the acquisition of the IC curve and the identification of its peaks have considerable challenges. The amount of calculation increases the accuracy while slowing down the operation rate.

Dai Houde et al. incorporated additional features into the algorithm input based on ICA [31]: (1) They considered the average signal to reflect the overall magnitude of the signal by using the average value of the charging and discharging voltages. (2) The charging time, which shows a decreasing trend in CC mode with decreasing SOH, was observed from one cycle's charging curve to reflect battery degradation. (3) The voltage increment over a fixed period of time increases with decreasing SOH as the internal resistance of the battery increases. (4) The four groups of peak values shown in Figure 4e were selected. The curve shown in the figure is the IC curve. Based on the figure, it is evident that the voltage per unit capacity of the battery undergoes significant changes during the period of battery capacity decay, thus the peak value is employed. The voltage delta over the same capacitance difference in the vicinity is characterized.

This method selects more peak points as features on the basis of ICA, improves the fault tolerance rate of ICA, and combines more features with strong correlation. The combination of multiple features is arguably a better feature selection method, which is helpful for improving accuracy.

MA HPBNNan et al. selected CNN as the neural network algorithm for constructing the prediction model [32]. The input chosen was the four-dimensional lithium-ion battery data of voltage, current, temperature, and time during each cycle as shown in Figure 4c. To avoid the input data volume being too large, a sampling rate of 1 Hz was determined to be used directly. This method expands a relatively novel data processing method, which uses large-scale data input and then uses neural network algorithms to automatically extract features, instead of Manually screen data and select features.

By fully entrusting feature selection to the computer, this method partially eliminates the limitations and biases of human selection. However, it also leads to difficulties in monitoring and modifying the input features, sacrificing a certain degree of controllability and leaving little room for improvement.

Ren Lei et al. expanded on the basis of M. A. HBPNNan et al. and selected the data of the charging process and the discharging process at the same time, and obtained the measured voltage or current by observing how it changed to a certain value [33]. For example, they took the value of the maximum measured voltage during each complete charge and recorded the time and value of the charging (discharging) current (voltage) reaching the preset value described in the dataset documentation. After extracting the features of each LIB, the method applied an auto-encoder (AE) to increase the

dimensionality of the original data. The AE is an unsupervised neural network that compressed the input into a hidden space representation and then reconstructed this representation as the output to perform the task. The neural network structure of the AE is shown in Figure 4d. It consists of an input layer, hidden layer, and output layer. The AE mapped the input to the output 21-dimensional vectors of each original adjacent 14 charge–discharge cycles stacked vertically into a temporal feature map of size 14×21 . A 21-dimensional LIB data vector was input to the encoder, resulting in a 50-dimensional output. After the encoding process of the auto-encoder, the feature map of the original dataset could be enlarged to a 14×50 time-domain feature map as the input for subsequent calculations. The vector formed in this way included the voltage measured during charging and the discharge temperature for each charge and discharge cycle.

This method involves expanding the data dimension and utilizing the AE to enhance it. While the process of data information extraction for the subsequent neural network remains unobservable, it improves the feature's interpretability and consequently enhances the feature and model quality to some extent. However, increasing the dimensionality also incurs a higher computational cost for the model and may lead to decreased fitting accuracy.

Kristen A. Severson and others also captured the information in the charge and discharge voltage curves of LIB [34]. They first extracted common features such as voltage, current, temperature, and internal resistance, and excluded some features with low correlation coefficients. Then, they obtained the capacity-voltage function and observed and calculated the dynamic changes of the discharge voltage during the cycle. The quantitative changes of the charge and discharge voltage were obtained by applying a fixed cycle period. The key factor in the quantitative analysis of the change in charge and discharge voltage is the variance, as well as other metrics such as the minimum value, average value, and the difference between the discharge voltage of a given cycle and that of the 10th cycle. The average negative correlation coefficient of -0.92 was obtained through correlation analysis. Through various comparison experiments, the variance of the discharge voltage difference was calculated. A comprehensive experimental analysis was carried out mainly in combination with various other characteristics.

The proposed discharge voltage difference scalars may not have a clear physical meaning, but the analysis of LIB SOH decay process from a dynamic changes perspective strongly impacts the data-driven algorithm. The author's comparative experiments provide ample evidence for this. The variance of the single-selected discharge voltage difference and its combination with various other features have been compared and analyzed to prove the reliability of this feature. Under the premise of guaranteed accuracy, the amount of calculation is also very appropriate, and there will be no ICA and AE increase. This method incurs a significant computational burden, and while its simplicity may result in some loss of accuracy, it does offer a very high computational speed.

Feature selection has a strong connection with the selection of data-driven algorithms in the following. Common LIB features include time, voltage, current, temperature, etc. Further features can be analyzed and calculated. The variance of the discharge voltage difference also has a strong correlation with the SOH. For a network with strong data mining capabilities such as CNN, the data can be arranged reasonably, or the AE can be used to increase or reduce the dimension and directly input it as a feature. This feature is the key step of SOH prediction of LIB, which has strong flexible combination characteristics. The reader can analyze the correlation of each feature and adjust it according to the characteristics of the data-driven algorithm.

3. Data-Driven Algorithms

After implementing feature extraction and data processing, it is necessary to select an appropriate algorithm to train the SOH prediction model. This is the core step of SOH prediction, and various algorithms need to be screened.

ML and DL, as a broad category, can be classified into supervised learning and unsupervised learning. In unsupervised learning, such as the autoencoder mentioned in the previous section, only feature data are provided, and the autoencoder can learn independently, analyzing the data structure and outputting a higher-dimensional representation of strongly correlated eigenvectors. The unsupervised algorithm is also composed of a variety of clustering algorithms, which are not the focus of this paper and will not be described in detail. The supervised algorithm has a clear prediction answer, and the algorithm that learns the features and obtains accurate prediction results through the algorithm is divided into two major categories: regression problems and classification problems. This paper focuses on the SOH estimation of LIB, which involves a regression problem where the aim is to predict continuous values as output. In the field of regression, there are various common algorithms that dominate, such as linear regression (LR), support vector regression (SVR), convolutional neural network (CNN), etc. These algorithms have their own advantages for regression prediction under different principles and basic frameworks. In the following chapter, we will provide a detailed analysis of the principles, frameworks, and advantages and disadvantages of these common algorithms.

3.1. Linear Regression

As one of the core subcategories of machine learning, the LR model is the cornerstone of the regression algorithm. The classic LR assumes that the model is current for the parameters, the parameters can only appear in the form of a power, and the error needs to obey the normality with zero mean. In terms of distribution, the eigenvalues need to be variable and independent of each other. It can be seen that LR as a basic algorithm has strong limitations, and it cannot be applied in many situations where nonlinear or feature distribution are scattered. The basic LR assumption function is:

$$y_i = \theta^T x_i \quad (2)$$

Among them, y_i represents the output battery SOH predicted value vector, x_i represents the input n-dimensional feature vector, and θ^T represents the n-dimensional model parameter vector. The fitting goal of LR is to minimize the loss function, which is also the goal of most data-driven algorithms, that is, the error between the estimated value and the actual value, the formula is:

$$J(\theta) = \frac{1}{2\alpha} \sum_{i=1}^n (y_i - \theta^T x_i)^2 \quad (3)$$

Among them, $J(\theta)$ represents the loss function, and α is represented as a constant, so that the squared effect is canceled when the loss function is derived and has no effect on the optimization result.

Kristen A. Severson et al. enhanced the basic LR method by introducing the elastic net (EN) regularization technique. EN is a LR model trained using both L1 and L2 norms as regularization terms. The resulting fitted model combines the sparsity of Lasso regression with the regularization ability of ridge regression. Unlike other methods, EN does not generate cross paths, leading to a faster convergence and higher efficiency and accuracy. Its loss function is:

$$J(\theta) = \frac{1}{2\alpha} \sum_{i=1}^n (y_i - \theta^T x_i)^2 + \varepsilon \lambda \sum_{j=1}^m \theta_j^2 + \frac{1-\varepsilon}{2} \lambda \sum_{j=1}^m |\theta_j| \quad (4)$$

They argue that in this paradigm, the input features are linear or non-linear variations of the original data, generated and used in the elastic framework of the regularization look-ahead framework, and the choice of the regularized linear model enables us to propose domain-specific features of varying complexity, while maintaining high interpretability. Linear models also have lower computational costs; models can be trained offline, and on-

line prediction only requires a single dot product after data preprocessing. After accounting for multiple ways of combining features, they obtained an average error of 7.5%.

As algorithms continue to gain popularity, there is a constant influx of complex and novel models that prioritize model accuracy. However, Kristen A. Severson and colleagues introduced the idea of prioritizing efficiency and convenience [34]. The elastic network (EN) they proposed simplifies the calculation process for offline processing, and while its prediction error of 7.5% may be higher than more complex models, it meets the practical requirements of normal engineering applications. This highlights the need for ML and DL to consider the balance between efficiency and accuracy, which is a critical and sometimes conflicting issue when implementing SOH prediction methods in engineering practice.

LR is a classical machine learning method, which can be used when features and results have a strong linear relationship, with ultra-high fitting efficiency and good accuracy. However, its limitations make it difficult to use this method in the prediction of SOH of lithium-ion batteries, and few researchers use LR as the core algorithm. If linear features can be obtained through mathematical operations, then LR is a good choice.

3.2. Support Vector Regression

The second chapter mentions that SVR is applied to the zero-denominator identification of IC curve. SVR is a regression prediction algorithm based on support vector machine (SVM). The original intention of SVM is as a classification algorithm. In the classification task, the distance between the sample points closest to the hyperplane is the largest. As shown in Figure 5, a simple binary classification task is completed.

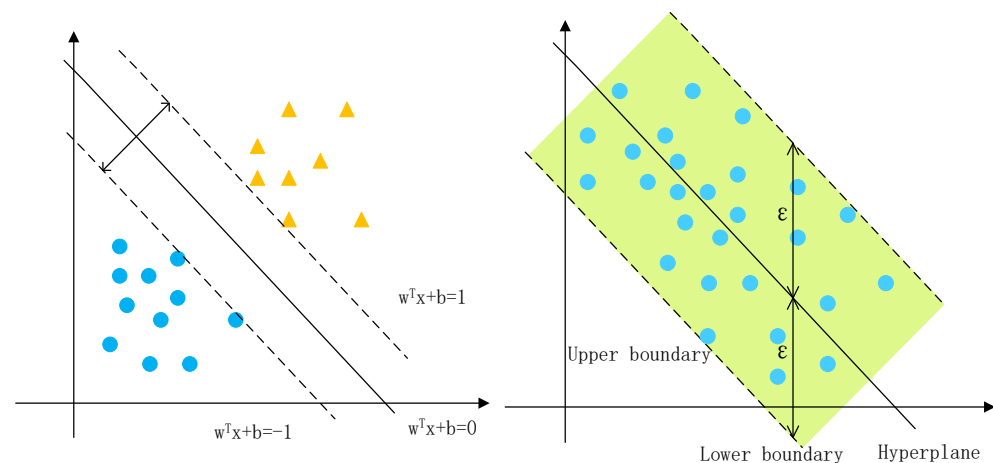


Figure 5. SVM and SVR.

To apply SVM to regression prediction, it is necessary to make changes. The same is to maximize the distance between the sample points and the hyperplane. The purpose of SVR is to maximize the distance between the farthest sample points on the hyperplane. The right side of Figure 5 is the SVR expectation. Therefore, the interval of SVR needs to be limited, that is, the deviation between the established SOH prediction model and the actual SOH value must be $\leq \epsilon$, and this deviation range is called the ϵ pipeline. The mathematical formula is:

$$\min_{w, b} \frac{1}{2} \|w\|_2^2 \quad (5)$$

$$s.t. \quad |y_i - (w^T x_i + b)| \leq \epsilon, \quad i = 1, 2, \dots, N \quad (6)$$

Among them, w represents the weight, y_i represents the virtual function, b represents the intercept term, and x_i represents the input vector. The selection of ϵ is a crucial factor in practical engineering applications. If ϵ is set too small, there is a risk that some samples will fall outside the ϵ -tube, which could result in the model being unable to capture important

information. Conversely, if ε is set too large, outliers may heavily influence the model, leading to a model with high variance and poor performance.

$$\begin{cases} y_i - (w^T x_i + b) \leq \varepsilon + \xi_i^\uparrow \\ (w^T x_i + b) - y_i \leq \varepsilon + \xi_i^\downarrow \\ \xi_i^\uparrow, \xi_i^\downarrow \geq 0 \end{cases} \quad (7)$$

Among them, ξ_i^\uparrow represents the upper bound constraint, and ξ_i^\downarrow represents the lower bound constraint.

Xiao Bin et al. combined the least squares (LS) method with SVR to developed a least squares support vector regression (LS-SVR) model with a polynomial kernel function [30]. For the LS-SVR regression model, the hyperplane can transform the problem of minimizing the loss function and solve the convex optimization problem through the Lagrangian function:

$$\begin{cases} \min R(\omega, C, \xi) = \frac{1}{2} \|w\|_2^2 + \frac{1}{2} C \sum_{i=1}^N \xi_i^2 \\ \text{s.t. } |y_i - (w^T x_i + b)| \leq \varepsilon, \quad i = 1, 2, \dots, N \\ L(\omega, b, \xi, \alpha, \beta) = R(\omega, C, \xi) - \sum_{i=1}^N \alpha_i (\omega^T \varphi(x_i) + b + \xi_i - y_i) \end{cases} \quad (8)$$

where C is the penalty factor and α and β are the Lagrange multipliers. The optimized regression model can be written as:

$$f(x) = \sum_{i=1}^N \alpha_i \varphi(x_i) \varphi(x_j) + b \quad (9)$$

where x_i and x_j are both input vectors. This method adds the K-fold verification method when optimizing the hyperparameters. The errors obtained by the LS-SVR model under the verification of multiple battery data are shown in Table 1. Although the performance of different battery groups is different, the overall performance indicators remain the same. The LS-SVR model with a polynomial kernel has proven to be effective in overcoming the nonlinear relationship between input variables and battery SOH, as demonstrated by the total error range of battery SOH estimation being between -1.83% and 2.25% .

This method adds the K-fold verification method when optimizing the hyperparameters. The errors obtained by the LS-SVR model under the verification of multiple battery data are shown in Table 1. Although the performance of different battery groups is different, the overall performance indicators remain the same. The results of the study demonstrate that the LS-SVR model with a polynomial kernel is effective in overcoming the nonlinear relationship between the input variables and battery SOH. This is evidenced by the total error range of battery SOH estimation, which falls between -1.83% and 2.25% .

Racha Khelif et al. used the SVR model of the linear kernel function [35], as shown in the following formula:

$$k(x_i, x_j) = x_i^T x_j \quad (10)$$

They also combined the time detection mechanism to input the health data of LIB and used the weighted average method to obtain the input features and make predictions. The adjustment of γ , C , ξ , combined with ten cross-validation, obtained the best MSE of 0.045. The article also stipulates a novel prediction result evaluation method, which is detailed in the Table 2.

Table 2. Results of the estimation.

Kernel	Test Case	RMSE	MAE	MAPE	R2(%)	Error (%)
Polynomial kernel	No.5	1.26	0.89	0.53	99.8	[−1.48 2.25]
	No.6	1.36	1.08	0.65	99.73	[−1.83 1.09]
	No.7	0.95	0.7	0.41	99.67	[−1.4 1.39]
	No.5.6.7	1.19	0.89	0.53	99.66	[−1.83 2.25]

In order to avoid the direct calculation of the inner product in the high-dimensional or even infinite-dimensional feature space, Zhang Yajun et al. chose the Gaussian kernel function to map the input to the high-dimensional space combined with the feature extraction method of ICA for combined modeling [36]. The formula is as follows.

$$k(x_i, x_j) = \exp\left(\frac{-||x_i - x_j||^2}{2\sigma^2}\right) \quad (11)$$

where σ represents the Gaussian parameter, which determines the height of the peak of the Gaussian distribution. The authors suggest that the SVR model for Li-ion battery SOH estimation has a low computational burden and high accuracy, as only a small number of support vectors (SV) have non-zero values and dominate the model. Furthermore, the SVR model does not require intensive mathematical operations. The authors conducted two experiments on three sets of batteries in various datasets, and the accuracy was much higher than that of the ordinary linear regression model, as shown in Table 3.

Table 3. Results of the estimation.

Battery Label	SVR		Linear Model	
	MAE (%)	RMSE (%)	MAE (%)	RMSE (%)
B5	0.3413	0.5935	3.1645	3.6222
B6	0.5979	1.0953	5.296	6.0685
CS35	0.2729	0.3663	1.4508	1.8601
CS36	0.5157	0.6701	3.4108	3.9441

Cai Lei et al. proposed a feature selection method using non-dominated sorting genetic algorithm (NSGA-II) [37], which combined the use of radial basis function (RBF), also known as Gaussian kernel function to construct SVR, and used the same kernel as Zhang Yajun et al. function.

The RBF sum function has the benefit of being able to approximate other kernel functions by parameter adjustment, making it a flexible kernel function. Figure 6 presents a novel comparison using this method to predict SOH from different battery SOC. The SOH estimation for battery 1 and battery 2 using current pulses at SOC = 20% and SOC = 50% is a preferred choice. Since the mean square error of battery 2 as the validation set is 0.0046 under all conditions, and the results of the training set battery 1, its mean square error is 0.0035, thus demonstrating the good generalization of the method.

As can be seen from the above, the kernel function is used as the core of SVR for regression prediction, and the common SVR kernel functions used for SOH prediction of LIB are shown in Table 4.

The overall advantage of the SVR algorithm lies in that the nonlinear data needs to be mapped to the high-dimensional space through the kernel function in the calculation to become linearly separable data, which cleverly avoids the complexity of the high-dimensional space. When the kernel function is known, it can be simplify the difficulty of solving high-dimensional space problems. At the same time, SVR finally solves a convex quadratic programming problem, which will obtain a global optimal solution, solve the local extreme value problem that cannot be avoided in neural networks, and have better generalization ability.

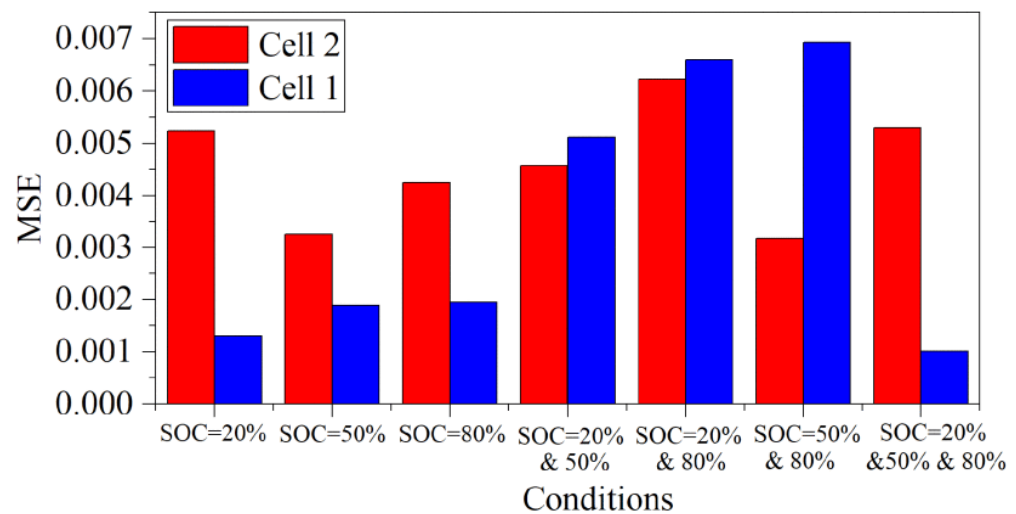


Figure 6. Results of the SOH estimation.

However, the mapping process of the kernel function brings more computation and requires a lot of storage space, which is also one of the inevitable problems of SVR. Moreover, the establishment of the SVR model and the selection of the kernel function are more dependent on the selection of LIB features, the analysis of the feature structure, and the selection of hyperparameters. Whether the kernel function and features are matched and adapted determines the accuracy and rate of SVR prediction. At present, there is no established method for selecting models, and researchers have to rely on their experience and experiments to make a choice.

Table 4. Kernel functions commonly used in regression prediction.

Function	Formula	Remark
Linear kernel	$k(x_i, x_j) = x_i^T x_j$	
Polynomial kernel	$k(x_i, x_j) = (x_i^T x_j)^2$	$d \geq 1$ is the degree of polynomial
Gaussian kernel	$k(x_i, x_j) = \exp\left(-\frac{\ x_i - x_j\ ^2}{2\tau^2}\right)$	$\tau > 0$ is the bandwidth of the Gaussian kernel
Sigmoid kernel	$k(x_i, x_j) = \tanh(\beta x_i^T x_j + \theta)$	$\beta > 0, \theta < 0$

From the results given by various researchers, SVR kernel function selection is very flexible and has a great impact on the results. It is a machine learning algorithm with great potential. Compared with LR, SVR can achieve higher prediction accuracy. With the development of mathematical theory, SVR will certainly be enhanced.

3.3. BP Neural Network

The BP neural network (BPNN) emulates biological processes to exhibit certain traits of the human brain using artificial neurons. It operates on a parallel divisional processing structure and is categorized as a machine learning algorithm. The artificial neurons within BPNN are typically referred to as “processing units” and from the perspective of the network, they are commonly known as “nodes”. As shown in the Figure 7a, the input features are weighted and summed by weights, and the whole is input to the neuron, which is mapped to another value through the activation function for output. The activation function represented by f generally uses the sigmoid or relu function.

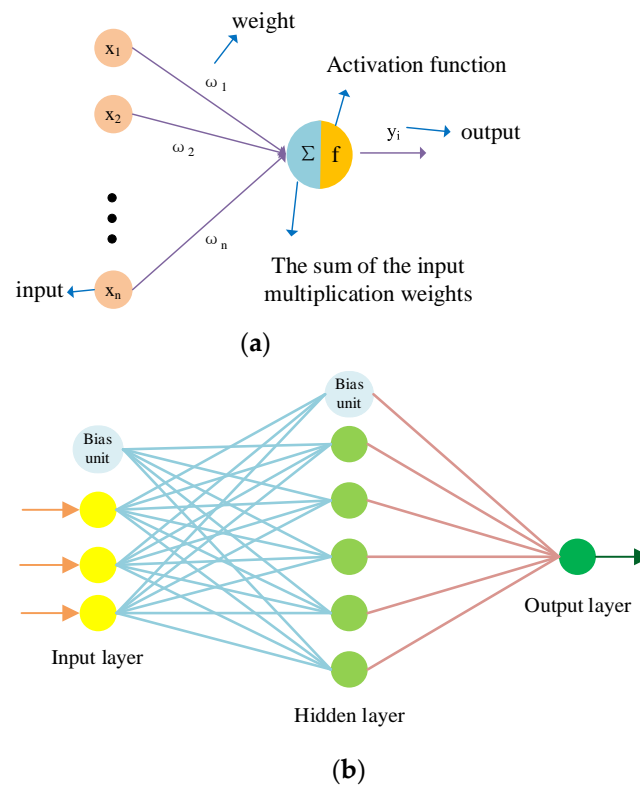


Figure 7. (a) The Specific Structure of BP Neural Network; (b) The Overall Structure of BP Neural Network.

As shown in Figure 7b, the basic BPNN architecture is formed by interconnecting processing units (artificial neurons) and linked signal channels. Each processing unit can branch into any number of identical signals, which are then connected in parallel. The architecture includes input layer, hidden layer, and output layer, of which there can be multiple hidden layers, which is also the key to self-construction and design of BPNN.

The definition of the loss function is the same as that of LR. This is the standard for evaluating the forward propagation process. The mean square error (RMSE) is often used as the loss function for regression problems. The formula is:

$$E = \sum_{i=1}^n (T_i - O_i)^2 \quad (12)$$

Among them, E represents the error, T_i represents the target value, and O_i represents the model predicted value. In the process of backpropagation, the actual error is propagated, and the error of the output layer is transmitted back to the network, and then the weight parameters are corrected. Its basic weight update formula is:

$$\delta w = \alpha \cdot E_k \cdot O_k (1 - O_k) \cdot O_j^T \quad (13)$$

Among them, α represents the learning rate, and includes all the constants in the formula, and the subscript K represents the number of layers of the neural network. Backpropagation also relies on gradient descent to find the value of w that minimizes the error. As an optimization algorithm, gradient descent is one of the most important techniques and foundations. The algorithm is constantly updated in order to minimize the loss function by finding the minimum value of the error gradient. Stochastic gradient descent updates the parameters for each training sample, with a single update executed at each step, resulting in faster speeds. However, frequent parameter updates can lead to high variance, which can be partly alleviated by mitigating local minima [38]. Mini-batch gradient descent addresses the issues associated with stochastic gradient descent by

updating parameters using sample values in batches, reducing fluctuations and ensuring stable convergence, making it a favorable optimization method. Additionally, gradient descent can leverage momentum techniques to accelerate gradient descent training by optimizing relevant directions and attenuating irrelevant directions [39]. The Adagrad algorithm optimizes by adjusting the appropriate learning rate and is particularly suitable for processing sparse data [40]. Finally, the AdaDelta algorithm builds on Adagrad to tackle the issue of learning rate decay and constrain the window of previous gradients [38].

The accuracy of the univariate model (m-AIC) based on m-AIC algorithm was further improved by Khalid, A et al. [41], firstly, by modeling and unifying with the multi-layer perceptron (MLP), and then, by modeling and unifying with the nonlinear autoregressive neural network (NARX) neural network with external input (using the previously predicted parameters). The predicted mean square error of the model is 0.1048% and 0.0175%, whereas the corresponding error of the independent model is 0.271% and 0.0236%, respectively.

M A Hannan et al. proposed a BPNN algorithm combined with backtracking search algorithm (BSA) [32]. The parameters of BPNN have an important impact on the performance of the algorithm. The learning rate parameter can be used to identify if values are stuck in local minima. In contrast, the hidden layer neurons control the time complexity of the algorithm. Overfitting may occur if the hidden neurons are higher than optimal, and underfitting and high variance may occur when the number of hidden neurons is less than optimal. These parameters are set empirically, and it is difficult to ensure an optimal solution. Therefore, the BSA algorithm is used to obtain the optimal value and learning rate of the hidden layer neurons. BSA can operate large-dimensional problems using historical populations and mapping matrices to obtain optimal solutions. Taking the data of the Dynamic Stress Test (DST) cycle as the training and testing sets, the RMSEs of the BPNN-BSA model calculated with the battery data at 0 °C, 25 °C, and 45 °C were 1.47%, 0.81%, and 0.48%, respectively.

Ge Dongdong et al. proposed the use of BPNN to improve extreme learning machine (ELM) combined with the bat algorithm (BA) [42]. ELM is a feedforward neural network with a single hidden layer, which simplifies the original BPNN model, and has the advantages of fast learning speed and generalization. To prevent the hidden layer from having a fixed number of neurons, the connection weights and thresholds are set randomly. However, this approach may result in an imprecise estimation of the State of Health (SOH) of the battery. To overcome these shortcomings, the BA algorithm is used to optimize the connection weights and biases of the ELM model. The RMSE of the BA-ELM model is 0.5354%, and the MAE is 0.4326%, which is the smallest error among the models in the comparison test.

The structure of feedforward neural networks, including BPNN and its variants, is similar. They exhibit greater capacity for processing multi-dimensional nonlinear data and possess stronger feature mining capabilities, compared to the LR algorithm. This enhances accuracy to some extent. However, these neural networks tend to suffer from over-fitting issues and are susceptible to local minima. Furthermore, they can only mine the correlation between input and output from a one-dimensional perspective. However, with the progress of charging and discharging cycles of LIB, there are differences between the data. Even with the combination of BPNN's basic model with excellent parameter optimization or weight fitting algorithms, its accuracy cannot be improved due to the inability to mine the correlation in the time series. As a result, BPNN is typically used as the primary model, but there are limited instances of achieving outstanding results in Li-ion battery SOH prediction using this neural network.

According to the various methods proposed by researchers, BPNN is a fast fitting algorithm with general accuracy and large prediction fluctuation, which is suitable for small batch and low accuracy requirements, and is also suitable for applications such as edge computing.

3.4. Long Short-Term Memory Neural Networks

The SOH value of a LIB is a sequential data set. Therefore, it is necessary to consider not only the data from the current time step but also the data from the preceding time step during processing, as it can influence the subsequent time step. The transmission of information is one-way. The output of the network only depends on the current input and has no memory ability. The RNN can record the output of each time step, and the previous time step will affect the output of the subsequent time step, so it can effectively process data with sequence characteristics and mine time series information from it. The common loop body structure is a special neural network structure. As shown in Figure 8a, it consists of an input layer, a hidden layer, and an output layer. After the loop body is expanded, it is shown on the right.

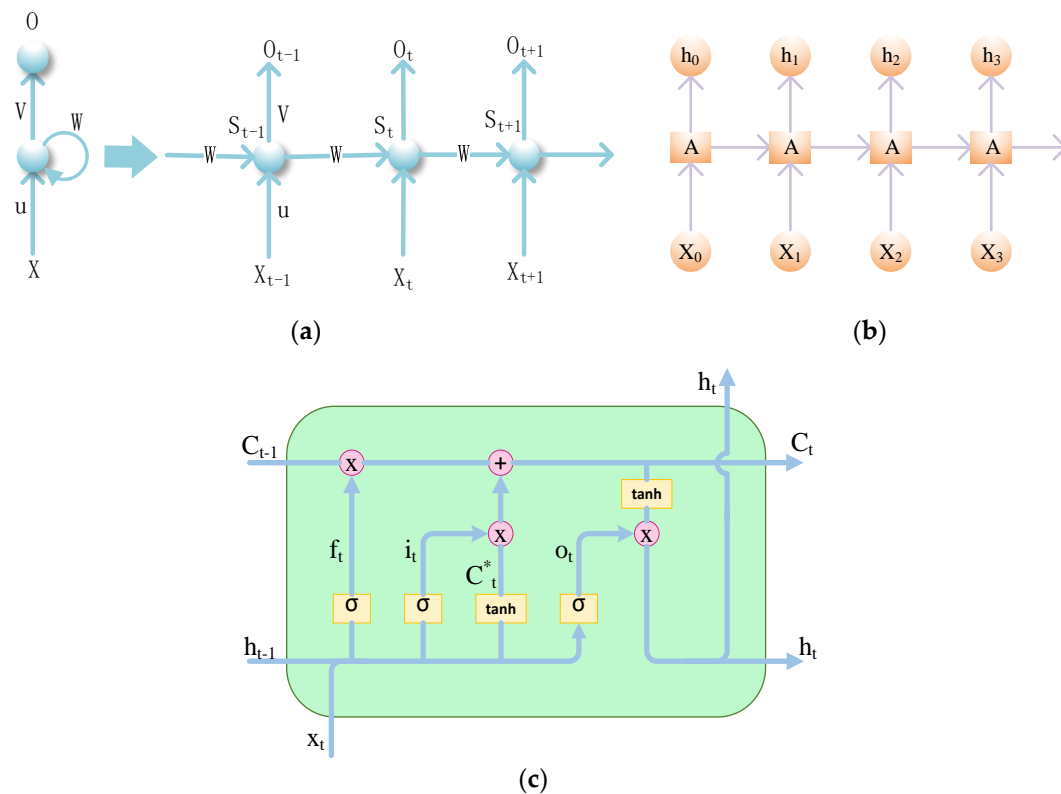


Figure 8. (a) Circulatory structure; (b) Basic structure of RNN; (c) Basic structure of LSTM.

Among them, the input of the hidden layer processing unit S_t includes X_t , S_{t-1} , X_t is the input at the current moment, and S_{t-1} is the information at the previous moment. After receiving the input, the \tanh activation function should be used to map, and O_t is the output at time t . Softmax is often used as an activation function. RNN is formed by stacking loop bodies. As shown in Figure 8b, it is a one-way RNN, which usually processes data related to the current moment and the previous moment.

Khalid, A et al. used the empirical mode decomposition (EMD)—recursive wavelet neural network (RWNN) model as a prediction method for SOH of lithium-ion batteries [43]. A charge–discharge capacity analysis method based on reference charging rate was proposed for aging degradation analysis. The prediction accuracy of neural network was improved through advanced modeling and continuous filtering of various lithium-ion battery parameters.

Long Short-Term Memory Neural Network (LSTM) is a type of RNN, which is an improvement on RNN. The parameter learning of the RNN cycle can be learned through the back-propagation algorithm over time; in other words, the error is passed forward step by step according to the reverse order of time. When the input sequence is relatively long, the gradient explosion or gradient disappearance problem will occur, which is also

called the long-term dependency problem. To address this issue, gating mechanisms are introduced to improve recurrent neural networks, namely LSTMs and Gated Recurrent Units (GRUs). Figure 8c shows the detailed internal structure of LSTM, in which LSTM has three special network structures called “gates”. The overall combined LSTM structure can more effectively determine the forgetting or retention of information, specifically:

Forgetting gate (f_t): The forgetting gate will jointly decide which part of the memory needs to be forgotten according to the current input x_t , the state C_{t-1} at the last moment, and the output h_{t-1} at the last moment. The mathematical formula is:

$$f_t = \sigma_g(W_f x_t + U h_{t-1} + b_f) \quad (14)$$

Among them, W_f represents the input matrix of the forget gate at the current moment, U represents the output matrix of the previous moment, b_f represents the bias unit, and σ_g represents the sigmoid activation function.

Input gate (i_t): After the work of the forget gate is over, some information is deleted, and the input gate determines which memories will enter the current state C_t according to x_t , C_{t-1} , h_{t-1} . The mathematical formula is:

$$i_t = \sigma_g(W_i x_t + U h_{t-1} + b_i) \quad (15)$$

Output gate (o_t): After the new state C_t is calculated, the output of the current moment is generated through the output gate according to x_t , C_{t-1} , h_{t-1} . The mathematical formula is:

$$o_t = \sigma_g(W_o x_t + U h_{t-1} + b_o) \quad (16)$$

So far, the current state C_t and output h_t can be obtained through the overall structure of LSTM. The mathematical formula is:

$$C_t = f_t * C_{t-1} + i_t * \sigma_c(W_c + U h_{t-1} + b_c) \quad (17)$$

$$h_t = o_t * \sigma_c(C_t) \quad (18)$$

Among them, σ_c represents the tanh activation function, and $*$ represents the star multiplication, which is the point-to-point multiplication between matrices.

The core idea of the GRU model is the same as that of the LSTM, with only a slight difference in the selection of the gating module. Interested readers can refer to it by themselves. It is generally believed that the LSTM and GRU only differ in the amount of calculation. The LSTM is larger and the GRU is reduced by 25%.

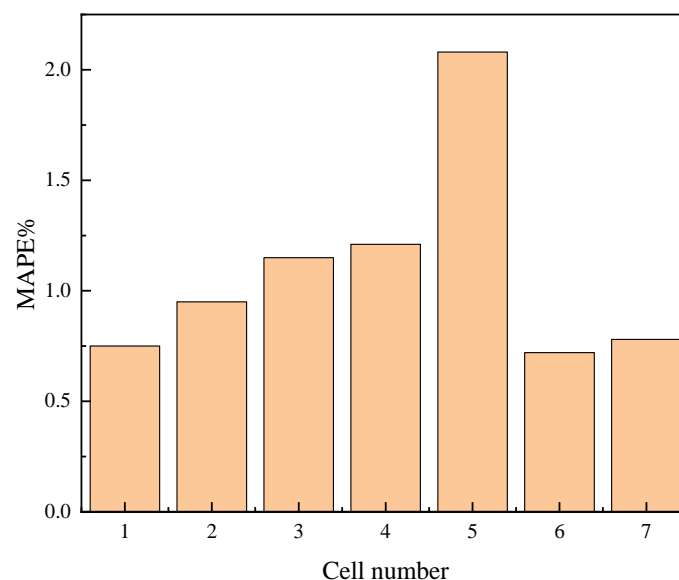
Khalid, A et al. creatively used multi-layer perceptron (MLP) combined with LSTM to estimate the available capacity of batteries, and fully considered the nonlinear and time-varying characteristics of batteries [44]. They believed that compared with the single LSTM model, the combination had higher accuracy and robustness.

Ephrem Chemali et al. proposed to use LSTM to mine deeply the characteristic data of various LIB and make accurate predictions of SOC [45]. They believe that LSTM can mine information from time series compared with BPNN and can improve the performance of the model after full fitting. Generalization, compared to RNNs, improves on their longstanding problem, where gradients explode or vanish during backpropagation, as shown in the Table 5 below. The LSTM model achieves an MAE of 0.573% at a fixed ambient temperature and a MAE of 1.606% on the dataset where the ambient temperature is increased from 10 °C to 25 °C.

Table 5. Results of the estimation.

Ambient Temperature	MAE (%)	RMSE (%)	STDDEV (%)	MAX (%)
25 °C	0.774	1.110	0.796	3.692
10 °C	0.782	0.995	0.616	4.047
0 °C	2.088	2.444	1.270	6.687
Varying Temperature	1.606	2.038	1.256	5.815

Li Weihang et al. also proposed a method to predict the SOH of LIB in multiple environments with LSTM models, and provided the trained models to embedded devices with local deep learning functions for verification [46]. Such models have been tested in future battery systems. Model validation results show that it is robust against noisy inputs and is able to generate feasible outputs even when the input set provided to it is incomplete. This results in a viable SOH estimate even when the input measurement process is interrupted or erroneous. After comparative experiments, as shown in Figure 9, the network achieves the best mean absolute percentage error of 0.76%.

**Figure 9.** Results of the estimation.

Zhang Yongzhi et al. applied the improved LSTM model to predict the RUL of LIB, combined dropout to prevent overfitting of neural networks, and used elastic mean square backpropagation for adaptive optimization [21]. Through experimental comparison, it is found that the online data required by the LSTM-RNN model to accurately predict RUL is reduced to 20–25% of the complete data, and the error is relatively small.

Yang Fangfang et al. proposed a method that uses LSTM as an algorithm and combines the unscented Kalman filter (UKF) to filter out noise [47]. Estimating battery SOC at different temperatures up to 50 °C, referring to the results in Table 6, the root mean square error is less than 1.1%, and the average error is less than 1%.

Table 6. Results of the estimation.

Methods	MAE (%)	RMSE (%)	Computation Time (s)
Proposed method	0.82	0.93	1.20
NN	5.72	8.13	0.039
SVM	5.85	8.80	0.57
GPR	4.99	7.30	2.32

Liu Kailong and colleagues decomposed the original battery capacity data using EMD. They then utilized LSTM to fit the residuals, which enabled the maintenance and updating of long-term dependence of battery capacity degradation without encountering gradient disappearance. Additionally, they employed a GPR sub-model to capture local fluctuations [26]. This approach enabled the simultaneous consideration of quantifying uncertainty caused by the phenomenon of capacity regeneration. The prediction performance of several data-driven models was investigated and compared from the perspective of kernel function and number of training inputs. The combined LSTM+GPR model outperforms other similar models. Moreover, they suggested that approach can easily be extended to other battery health diagnostics. From the fitting error of the LSTM model at different time steps, the RMSE and the maximum error are in the range of 0.0048 and 0.038, respectively, indicating that the model can achieve higher accuracy.

LSTM is a neural network that is improved and optimized on the basis of RNN. From the structure, it can be found that the network can calculate the dependencies between various observations in the time series, and its large existence in the LIB data set is closely related to the time series related data. Compared to common models such as LR, SVR, and BPNN in mining feature information, this approach has significant advantages. Additionally, it addresses the issue of the gradient disappearing during neural network training, which can cause abnormal training, and exhibits strong generalization capabilities. At the same time, adding dropout and other methods can further improve its generalization advantage.

The limitation brought by LSTM is also due to its high requirements on hardware resources, which require a large amount of data as support. At the same time, the gradient problem of RNN has been solved to a certain extent in LSTM and its variants, but the sequence order is still limited, and is not completely resolved. Moreover, the four linear layers in the LSTM structure consume significant resources for calculation during operation, and the computational efficiency will be significantly lower than that of simple neural networks or data-driven algorithms.

As one of the most established algorithms currently used, LSTM plays an important role in the research field. Many researchers use LSTM as the core algorithm for in-depth learning, and its prediction accuracy and robustness are very good. It can perform high-precision fitting for time series data, and is a well-established and in-depth learning algorithm, which can be widely used.

3.5. Convolutional Neural Network

CNN is often used in machine vision because it has great advantages in processing pixels in images. Furthermore, methods were later developed to deal with sequence prediction. As a kind of neural network, as shown in the Figure 10, it can be generally divided into input layer, convolution layer, pooling layer, fully connected layer, and output layer.

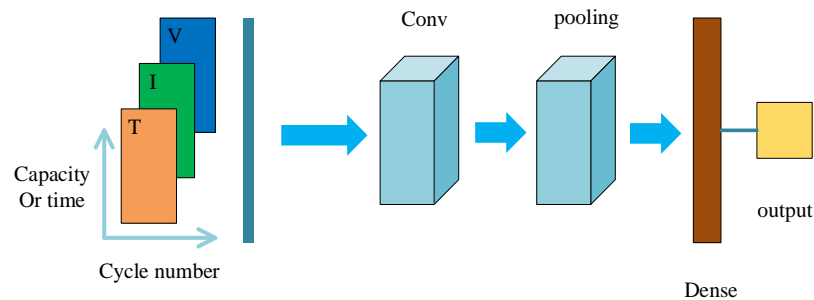


Figure 10. CNN basic structure.

The convolutional layer slides a window similar to 3*3 on the constructed input feature map, stops at each possible position and extracts the surrounding feature component weight matrix, also known as the convolution kernel. The pooling layer performs different pooling

operations. Generally, the maximum pooling operation is used to sample the feature map. Usually, a 2*2 window and a stride of 2 are used, so that the features extracted by the convolution layer can be refined. Generally speaking, two to three groups of convolution-pooling combination algorithms are used to complete the feature extraction step.

As discussed in the previous chapter, there is no clear quantitative formula for describing the relationship between LIB life and the voltage, current, and temperature curves during charging, so we need to empirically find features that can accurately describe the life of LIB; however, the current relationship between the V/I/T curve of a battery and the battery cycle life (or RUL) is difficult to fully understand due to its complex electrochemical reaction mechanism.

Therefore, CNN is used to comprehensively and automatically capture and model the features hidden in the curve. The feature map is to arrange all the features into a 3D model according to a certain rule [48]. In the case of Lithium-ion batteries (LIBs), a one-dimensional model can be developed using time or battery capacity. Alternatively, a 3D model can be created by using the number of cycles as one dimension and different features as another dimension, as reported in [49]. It can be seen from the above analysis that CNN can use convolution and pooling operations to continuously identify prominent features from the constructed feature map and perform integration optimization [50]. Compared with the features selected by experience, it is not necessarily more accurate, but it will be comprehensive. The captured features are input to the fully connected layer for neural network operations, and finally a complete LIB SOH prediction model is obtained.

Qian Cheng and colleagues developed a 1D CNN model, which includes 1D convolution, pooling, and fully connected layers [51]. The proposed model consists of three convolutional layers, one global pooling layer, and two fully connected layers. The layer composition is shown in Figure 11a,b. After each convolutional layer, a batch normalization technique is applied to improve the performance and stability of the 1D CNN model. Furthermore, as shown in Table 1, the stride in the convolutional and pooling layers is artificially set to 1, whereas the other hyperparameters should be optimized by an optimization algorithm. Before training, the 1D CNN parameters are initialized to a uniform distribution using the Xavier initialization method. Then, during training, these parameters are updated based on a gradient descent algorithm and a mean squared error loss function. When conducting the comparative experiments, the data are segmented from three different locations as input, and given the prediction results of the two batteries in Figure 11c,d, the CNN model still provides the best capacity estimates for all scenarios.

Ren Lei et al. proposed an Auto-CNN-LSTM model, which uses the aforementioned AE to upgrade the feature map to improve the mineable line of LIB data, and then uses the CNN module to extract the feature map output from the autoencoder [33]. Regarding the extract depth information, they chose the most suitable CNN structure as shown in the figure below. Fully connected layers are often used to map the features learned by the convolutional and pooling layers into the sampling space, and they choose to drop them to reduce computation. The network structure is shown in Figure 11e. The CNN exploits the features in each charge and discharge cycle to find the temporal features between adjacent charge and discharge cycles that can be extracted by the autoencoder. During the forward propagation of deep CNN, the feature maps are gradually blurred, and the global information of each feature map will gradually become prominent. Finally, the information extracted by this part will be transformed into a feature vector with 480 elements through a flattening layer. This feature vector will be input into the DNN as part of the combination with the LSTM, and the LSTM model will be used to predict the mining ability of the feature time series. The predicted results are shown in the Table 7 below. The root mean square error of the prediction results obtained by using Auto-CNN-LSTM and filtering with linear filters is 4.8%, which is markedly superior to the error rates of the other two data-driven models based on the auto-encoder and deep neural network (ADNN) and Support Vector Machine (SVM), which produced error rates of 11.8% and 18.2% respectively.

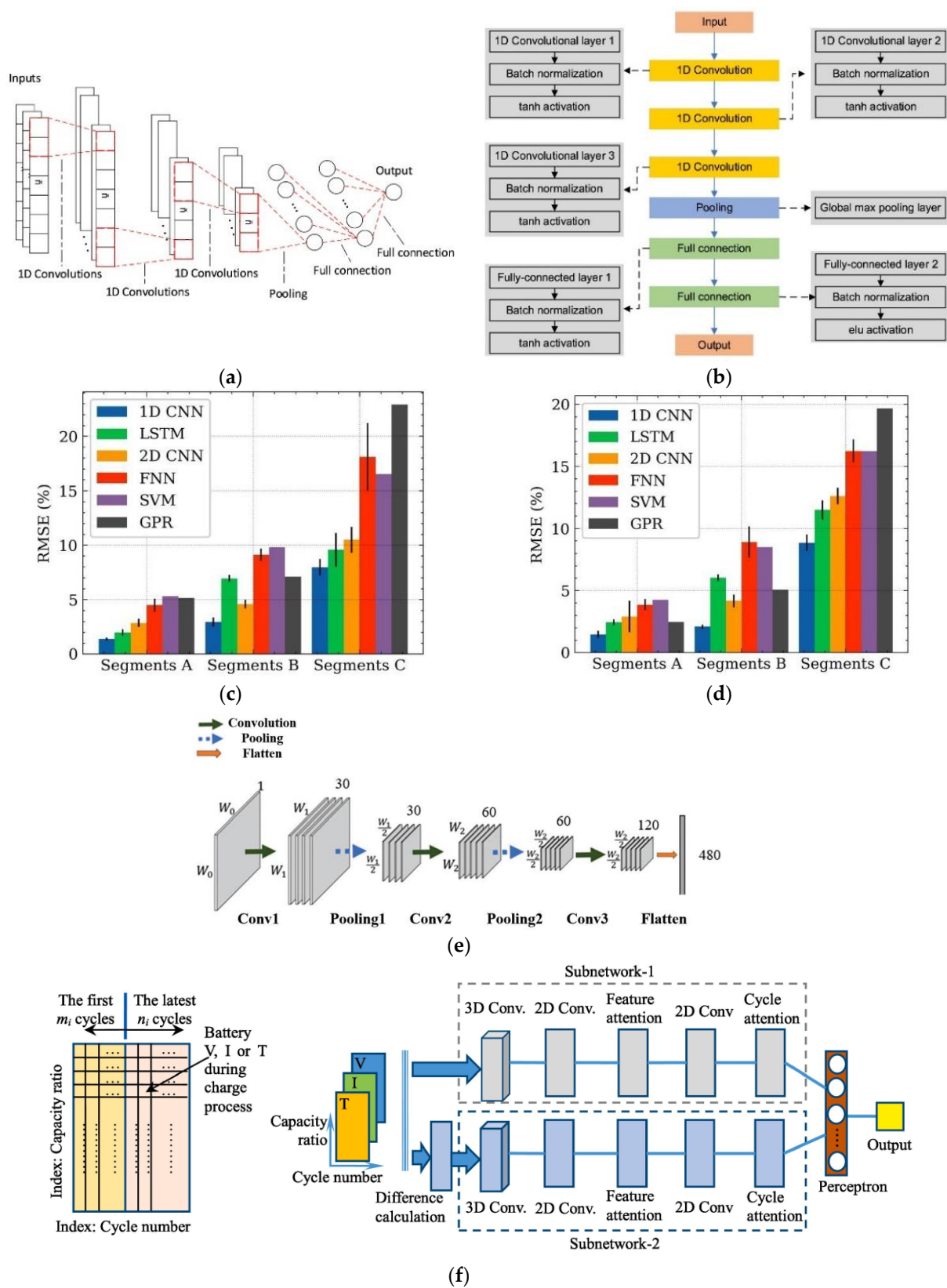


Figure 11. (a) The basic process of CNN; (b) CNN architecture based on 1D convolution; (c) Comparison of RMSE between DCNN and control experiments 1; (d) Comparison of RMSE between DCNN and control experiment 2; (e) Ren Lei's CNN Algorithm Structure; (f) Yang Yixin's CNN Algorithm Structure.

Table 7. Results of the estimation.

Method	RMSE (%)	Accuracy (%)
Auto-CNN-LSTM	5.03	94.97
Linear Filter	4.84	95.16
Second-Order Filter	4.98	95.02
Third-Order Filter	4.98	95.02
ADNN	11.80	88.20
SVM	18.23	81.77

Yang Yixin et al. proposed an HCNN model to simulate the relationship between battery charging V/I/T curves, battery cycle life, and RUL under different charging strategies [52–54]. Unlike traditional CNNs, the proposed HCNN is the first model to integrate both three-dimensional CNNs (3DCNNs) and two-dimensional CNNs (2DCNNs) for the purpose of predicting battery life. The introduced 3DCNN layer can fuse the differences between these V/I/T curves and their periods to account for the strong relationship between them. Moreover, the 2DCNN layer can fully and automatically extract and model the features hidden in these curves. The proposed method involves analyzing two distinct sets of cycles for the battery: the first m_i cycles represent the fresh state, while the subsequent n_i cycles reflect the cycled state. By considering the battery's voltage (V), current (I), and temperature (T) as functions of capacity ratio, the model predicts battery life by inputting charging data for a total of $m_i + n_i$ cycles.

In the HCNN model shown in the Figure 11f, the information in the V/I/T three-dimensional input sequence is captured by adding a 3DCNN layer. Because the simple 3DCNN layer has too many neural network parameters, the model is difficult to train. Therefore, a hybrid CNN (HCNN) was proposed which combines 3DCNN and 2DCNN, and was connected by two subnetworks in parallel. Subnetwork-1 utilized cyclic charging as input for the three matrices generated by V/I/T during the process of $m_i + n_i$. On the other hand, subnetwork-2 employed the input of each curve and the difference between the first cycle curves of battery V/I/T. Considering the strong relationship between cells V, I, and T, only one 3DCNN layer is used as the input layer to fuse the curves. It is combined with two 2D CNN layers to automatically extract and model features hidden in the input curves. The results of the battery life prediction are finally output from the outputs of the two sub-networks through the perceptron. After conducting sufficient comparison experiments, under different charging strategies, the test error of early prediction of battery cycle life of different batteries is 1.1%, and the test error of RUL prediction is 3.6%.

After the above review of scholars, it can be found that when using CNN to predict the capacity problem of LIB, the manual selection process of features can be omitted, and the correlation between the data of each charging curve and the structure of CNN can be analyzed mainly. To match, many improved CNN structures have been developed, which are more suitable for various data of LIB and reduce the subjectivity in the prediction process. In addition, there is no need to worry about the huge load of high-dimensional data on the neural network algorithm. The shared convolution kernel has no pressure on processing high-dimensional data, and can comprehensively correlate the multi-dimensional data, and the extracted features are more accurate.

The limitations of CNN are also brought about by its prominent characteristics. The data integration of the pooling process also means that some features may be lost. The variability of its structure means that it needs to be confirmed through multiple parameter adjustments, and a large number of samples are required for supply. For its analysis and extraction, training will consume more resources and take a long time. However, the training process of CNN is similar to a black box, and its physical meaning cannot be clearly defined.

According to the improvement methods of CNN provided by researchers, the core problem of the algorithm is how to form a feature map and improve the internal logic of CNN. CNN has several structures designed to address various prediction problems and is

known for its strong robustness. The accuracy of predictions heavily relies on the chosen feature map and the structure of the CNN, both of which are determined by the researchers. Similarly, LSTM also possesses high accuracy and research potential, and is a deep learning algorithm that will likely become mainstream in the future.

After the above analysis, we summarize the advantages and disadvantages of common data driven algorithms in Table 8. Please refer to it in detail.

Table 8. Comparison of commonly used data-driven algorithms.

Algorithm	Advantage	Limitation	Prediction Accuracy
Linear regression	Simple linear model, supports fast fitting of linear data, direct thinking, strong interpretability, and high computational efficiency.	Accuracy depends on the quality of linear feature selection, requires strict linear assumptions, is prone to overfitting, and is sensitive to outliers.	LR algorithm is generally not suitable for fitting large quantities of nonlinear data and needs the support of linear features. It has good fitting effect for small batches of linear datasets. Taking MAE as an example, the error is generally about 10%.
Support Vector Regression	The high-dimensional data are skillfully processed through the kernel function, avoiding the complexity of the data, avoiding the local extreme value or the global optimal solution, and has a good generalization ability.	The mapping of the kernel function requires a lot of storage space. There is no fixed scheme for the selection of the kernel function, which is not mature enough, and requires a lot of adjustment parameters for comparison experiments. The problem of model overfitting still exists, and it is prone to local minima rather than global optimal solutions, and the ability to mine the relationship between features and output is general, Unable to mine more information from the timeline.	The progress of SVR algorithm depends on the selection of kernel function. Under reasonable selection, the MAE error of SVR prediction is about 5%.
Backpropagation Neural Network	Can directly process multi-dimensional nonlinear data, strong mining ability, improved generalization, and high computational efficiency.	The high requirements for hardware resources require a large amount of data to support, and the problem of gradient disappearance still occurs in large-scale data, and the calculation efficiency is low.	As a basic type of neural network, BPNN can generally reach about 7% of MAE error in LIB life prediction problems with less characteristic dimensions and non-explosive gradients.
Long Short-Term Memory Neural Network	It can calculate the dependencies between various observations in the time series and has great advantages in mining feature information. It solves the problem of gradient disappearance that has occurred when fitting small-scale data volumes, and the model accuracy and generalization are further improved.	The high requirements for hardware resources require a large amount of data to support, and the problem of gradient disappearance still occurs in large-scale data, and the calculation efficiency is low.	As a relatively mature neural network type, LSTM has solved many problems left by BPNN. With the support of various improved models, the MAE error of LIB life prediction can generally reach about 1%.
Convolutional Neural Network	The process of manual feature selection is omitted, and there is a large room for structural improvement. It can process large-scale data and fully mine the correlation between multi-dimensional data. The extracted features are more objective and the calculation accuracy is high.	The fitting process operating in a “black box” mode results in a prediction process that is uncontrollable and opaque, necessitating a large volume of data and exhibiting low computational efficiency.	As a neural network with a high degree of automation, CNN can achieve the same accuracy as LSTM, that is, about 1% MAE error, under the support of reasonable input and output, by focusing on the way to optimize its input characteristics

4. Conclusions

Due to the challenging engineering requirements, high precision, and harsh environmental conditions during LIB charge and discharge testing, acquiring data during actual operation is typically difficult, and can lead to significant errors and fluctuations. As such, we suggest that readers consider using publicly available datasets for their modeling operations.

In data processing, according to different data types, corresponding methods can be selected. The three mentioned in this paper are all common data processing methods with filtering properties, and different parameters can be set to adjust the required smoothness.

In terms of feature selection, since the overall charge and discharge data of LIB is nonlinear, we do not recommend selecting linear features. The research conducted by Kristen A. Severson and others is highly promising, but it is challenging to identify comparable features. In terms of practicality, methods akin to Xiao Bin's approach feature a small workload, convenient feature extraction, and high efficiency, albeit at the cost of a certain level of accuracy. Chengqi et al. and Dai Houde et al. utilized ICA to extract features and uncover those that possessed physical significance and high correlation. Despite the increase in computational workload, the use of ICA ensured accuracy, and the IC curve provided the capability for deeper feature exploration. Both MA HBPNNan et al. and Ren Lei et al. utilized a data feature map input that aligns with CNN. This approach replaces manual empirical feature extraction with AI-based feature mining. By providing a comprehensive data feature map, researchers have the option to select from multiple improved feature mapping techniques, along with the flexibility to choose post-input algorithms. This strategy enables deep input feature mining of LIB using CNN, with the primary focus being to ensure CNN can accurately accept input data. We believe this method has significant potential.

The most critical step is selecting a suitable data-driven algorithm. If a proper linear feature can be chosen, the LR algorithm is a convenient and efficient prediction method, but it is limited by its large error and tendency to overfit, so caution must be exercised in its selection. On the other hand, SVR is known for its excellent performance, particularly in its ability to process high-dimensional data and map it from nonlinearity to linearly separable spaces, while avoiding local extremes and producing good generalization performance. Choosing a suitable sum function can result in a highly precise algorithm. BP neural network, being a relatively simple neural network, performs well when combined with a better gradient descent model, which significantly improves its generalization and accuracy. Although it has high efficiency compared to other advanced neural network algorithms, its local extreme value problem is not easily solvable, and its information mining ability is moderate, making it a suitable method for basic work. LSTM and GRU are advanced algorithms of RNN, which excel at mining time series information from features and achieving high accuracy, essentially solving the problem of gradient disappearance. Adding dropout further improves its generalization, though gradient disappearance may still occur with large amounts of data, making its computational requirements significant. After the improvement of computer hardware, it has become an excellent data-driven algorithm. The CNN algorithm, even after being adapted from machine vision to regression problems, retains its strong ability to mine data. The combination of convolution and pooling can extract and refine features, solving the subjectivity and incompleteness of manual feature selection. When input features and network structure are adjusted, it becomes a highly powerful algorithm. The fully connected layer after feature mining can be combined with neural networks such as LSTM and GRU to form a reliable prediction model. The innovation focus of research is also mostly focused on optimization. In terms of the input method of features and changes in network structure, the disadvantage is that it is difficult to understand the operation process, in addition to the huge amount of calculation involved. At present, there are relatively novel CNN structures such as Resnet and Inception, which can be tried to be applied to the SOH prediction of LIB [52]. After referring to many experimental verification methods, we believe that the development trend of AI is complete objectification and automation. When combined with the experimental process, it can be concluded that utilizing CNN and its enhanced structures such as TCN

and Resnet represents a data mining algorithm that possesses both automation and accuracy. For example, Sun et al. used the improved DAG structure based on CNN to complete the high-precision prediction of lithium-ion battery SOH. The accurate mining of time sequence information by LSTM and GRU can ensure the analysis and processing of the above extracted information and solve the problem of time dependence and gradient explosion. For example, Zhang et al. used the LSTM-RNN algorithm to realize the high-precision prediction of RUL of lithium-ion batteries. Although the combination of the two will consume a large amount of computing resources, with the rapid improvement of hardware and software performance, these problems will be solved.

The main contribution of this paper is to present a comprehensive prediction process by integrating numerous relevant literature, including the comparison and demonstration of various methods. This approach helps readers stay updated on the latest research trends and future advancements in the field of SOH prediction for LIB Direction. In particular, the paper provides insights into the principles and strategies for using powerful data-driven algorithms. Therefore, this paper has significant practical implications and can guide future research in this field.

Author Contributions: Conceptualization, M.Z.; Formal analysis, D.Y.; Data curation, J.D.; Writing—original draft, H.S.; Methodology, L.L.; Investigation, W; Writing—review and editing, L.W.; Project administration, K.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Youth Fund of Shandong Province Natural Science Foundation (No. ZR2020QE212), Key Projects of Shandong Province Natural Science Foundation (No. ZR2020KF020), the Guangdong Provincial Key Lab of Green Chemical Product Technology (GC202111), Zhejiang Province Natural Science Foundation (No. LY22E070007) and National Natural Science Foundation of China (No. 52007170).

Data Availability Statement: There is no data that needs to be disclosed in this article.

Acknowledgments: The authors would like to thank Boyang Yu (Qingdao No. 58 high school) and Wenbo Wan (Qingdao No. 58 high school) for the literature search and analysis of data-driven algorithms in this paper.

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

Abbreviations

LTO	lithium titanium oxide	AE	autoencoder
LFP	LiFePO ₄ lithium iron phosphate	EMD	empirical mode decomposition
SEI	solid electrolyte interface	SVR	support vector regression
SV	support vectors	SVM	support vector machine
SOH	state of health	Simple RNN	simple recurrent neural network
RUL	remaining useful life	SOC	state of charge
ML	machine learning	LR	linear regression
DL	deep learning	BPTT	back propagation algorithm
LMO	lithium manganese oxide	LSTM RNN	long short-term recurrent neural network
IMF	intrinsic mode functions	MSE	mean square error
ANN	artificial neural networks	ICA	incremental capacity analysis
GRU	gated recurrent unit	GRA	grey relational analysis
RNN	recurrent neural network	NASA	National Aeronautics and Space Administration
CC	constant current	CV	constant voltage
GW	Gaussian window	LIB	Lithium-ion battery
CNN	convolutional neural network	BMS	battery management system
OCV	open circuit voltage	EOL	end of life
MA	moving average	LS	least squares
RBF	radial basis function	UKF	unscented Kalman filter

References

1. Zhang, M.; Liu, Y.; Li, D.; Cui, X.; Wang, L.; Li, L.; Wang, K. Electrochemical Impedance Spectroscopy: A New Chapter in the Fast and Accurate Estimation of the State of Health for Lithium-Ion Batteries. *Energies* **2023**, *16*, 1599.
2. Vetter, J.; Novák, P.; Wagner, M.R.; Veit, C.; Möller, K.C.; Besenhard, J.O.; Winter, M.; Wohlfahrt-Mehrens, M.; Vogler, C.; Hammouche, A. Ageing mechanisms in lithium-ion batteries. *J. Power Sources* **2005**, *147*, 269–281. [\[CrossRef\]](#)
3. Sarasketa-Zabala, E.; Aguesse, F.; Villarreal, I.; Rodriguez-Martinez, L.M.; Lopez, C.M.; Kubiak, P. Understanding Lithium Inventory Loss and Sudden Performance Fade in Cylindrical Cells during Cycling with Deep-Discharge Steps. *J. Phys. Chem. C* **2015**, *119*, 896–906. [\[CrossRef\]](#)
4. She, C.Q.; Wang, Z.P.; Sun, F.C.; Liu, P.; Zhang, L. Battery Aging Assessment for Real-World Electric Buses Based on Incremental Capacity Analysis and Radial Basis Function Neural Network. *IEEE Trans. Ind. Inform.* **2020**, *16*, 3345–3354. [\[CrossRef\]](#)
5. Liu, C.; Li, D.; Wang, L.; Li, L.; Wang, K. Strong robustness and high accuracy in predicting remaining useful life of supercapacitors. *APL Mater.* **2022**, *10*, 061106. [\[CrossRef\]](#)
6. Han, X.B.; Lu, L.G.; Zheng, Y.J.; Feng, X.N.; Li, Z.; Li, J.Q.; Ouyang, M.G. A review on the key issues of the lithium ion battery degradation among the whole life cycle. *Etransportation* **2019**, *1*, 100005. [\[CrossRef\]](#)
7. Wang, W.; Yang, D.; Huang, Z.; Hu, H.; Wang, L.; Wang, K. Electrodeless Nanogenerator for Dust Recover. *Energy Technol.* **2022**, *10*, 2200699. [\[CrossRef\]](#)
8. Verma, P.; Maire, P.; Novak, P. A review of the features and analyses of the solid electrolyte interphase in Li-ion batteries. *Electrochim. Acta* **2010**, *55*, 6332–6341. [\[CrossRef\]](#)
9. Ma, N.; Yang, D.; Riaz, S.; Wang, L.; Wang, K. Aging Mechanism and Models of Supercapacitors: A Review. *Technologies* **2023**, *11*, 38. [\[CrossRef\]](#)
10. Panchal, S.; Khasow, R.; Dincer, I.; Agelin-Chaab, M.; Fraser, R.; Fowler, M. Thermal design and simulation of mini-channel cold plate for water cooled large sized prismatic lithium-ion battery. *Appl. Therm. Eng.* **2017**, *122*, 80–90. [\[CrossRef\]](#)
11. Yu, X.; Ma, N.; Zheng, L.; Wang, L.; Wang, K. Developments and Applications of Artificial Intelligence in Music Education. *Technologies* **2023**, *11*, 42. [\[CrossRef\]](#)
12. Hu, X.S.; Che, Y.H.; Lin, X.K.; Onori, S. Battery Health Prediction Using Fusion-Based Feature Selection and Machine Learning. *IEEE Trans. Transp. Electrification* **2021**, *7*, 382–398. [\[CrossRef\]](#)
13. Sun, F.C.; Hu, X.S.; Zou, Y.; Li, S.G. Adaptive unscented Kalman filtering for state of charge estimation of a lithium-ion battery for electric vehicles. *Energy* **2011**, *36*, 3531–3540. [\[CrossRef\]](#)
14. Wang, D.; Yang, F.F.; Tsui, K.L.; Zhou, Q.; Bae, S.J. Remaining Useful Life Prediction of Lithium-Ion Batteries Based on Spherical Cubature Particle Filter. *IEEE Trans. Instrum. Meas.* **2016**, *65*, 1282–1291. [\[CrossRef\]](#)
15. Long, B.; Li, X.N.; Gao, X.Y.; Liu, Z. Prognostics Comparison of Lithium-Ion Battery Based on the Shallow and Deep Neural Networks Model. *Energies* **2019**, *12*, 3271. [\[CrossRef\]](#)
16. Guo, Y.; Yu, P.; Zhu, C.; Zhao, K.; Wang, L.; Wang, K. A state-of-health estimation method considering capacity recovery of lithium batteries. *Int. J. Energy Res.* **2022**, *46*, 23730–23745. [\[CrossRef\]](#)
17. Hu, X.S.; Jiang, J.C.; Cao, D.P.; Egardt, B. Battery Health Prognosis for Electric Vehicles Using Sample Entropy and Sparse Bayesian Predictive Modeling. *IEEE Trans. Ind. Electron.* **2016**, *63*, 2645–2656. [\[CrossRef\]](#)
18. Qu, J.T.; Liu, F.; Ma, Y.X.; Fan, J.M. A Neural-Network-Based Method for RUL Prediction and SOH Monitoring of Lithium-Ion Battery. *IEEE Access* **2019**, *7*, 87178–87191. [\[CrossRef\]](#)
19. Guo, Y.; Yang, D.; Zhang, Y.; Wang, L.; Wang, K. Online estimation of SOH for lithium-ion battery based on SSA-Elman neural network. *Prot. Control Mod. Power Syst.* **2022**, *7*, 40. [\[CrossRef\]](#)
20. Cui, Z.H.; Kang, L.; Li, L.W.; Wang, L.C.; Wang, K. A combined state-of-charge estimation method for lithium-ion battery using an improved BGRU network and UKF. *Energy* **2022**, *259*, 124933. [\[CrossRef\]](#)
21. Zhang, Y.Z.; Xiong, R.; He, H.W.; Pecht, M.G. Long Short-Term Memory Recurrent Neural Network for Remaining Useful Life Prediction of Lithium-Ion Batteries. *IEEE Trans. Veh. Technol.* **2018**, *67*, 5695–5705. [\[CrossRef\]](#)
22. Deng, Y.W.; Ying, H.J.; Jiaqiang, E.; Zhu, H.; Wei, K.X.; Chen, J.W.; Zhang, F.; Liao, G.L. Feature parameter extraction and intelligent estimation of the State-of-Health of lithium-ion batteries. *Energy* **2019**, *176*, 91–102. [\[CrossRef\]](#)
23. Liu, D.T.; Luo, Y.; Liu, J.; Peng, Y.; Guo, L.M.; Pecht, M. Lithium-ion battery remaining useful life estimation based on fusion nonlinear degradation AR model and RPF algorithm. *Neural Comput. Appl.* **2014**, *25*, 557–572. [\[CrossRef\]](#)
24. Zhao, G.Q.; Zhang, G.H.; Liu, Y.F.; Zhang, B.; Hu, C. Lithium-ion Battery Remaining Useful Life Prediction with Deep Belief Network and Relevance Vector Machine. In Proceedings of the IEEE International Conference on Prognostics and Health Management (ICPHM), Dallas, TX, USA, 19–21 June 2017; pp. 7–13.
25. Tian, H.X.; Qin, P.L.; Li, K.; Zhao, Z. A review of the state of health for lithium-ion batteries: Research status and suggestions. *J. Clean. Prod.* **2020**, *261*, 120813. [\[CrossRef\]](#)
26. Liu, K.L.; Shang, Y.L.; Ouyang, Q.; Widanage, W.D. A Data-Driven Approach with Uncertainty Quantification for Predicting Future Capacities and Remaining Useful Life of Lithium-ion Battery. *IEEE Trans. Ind. Electron.* **2021**, *68*, 3170–3180. [\[CrossRef\]](#)
27. Zhang, C.L.; He, Y.G.; Yuan, L.F.; Xiang, S. Capacity Prognostics of Lithium-Ion Batteries using EMD Denoising and Multiple Kernel RVM. *IEEE Access* **2017**, *5*, 12061–12070. [\[CrossRef\]](#)
28. Zhang, T.; Yu, M.; Li, B.; Liu, Z. Capacity Prediction of Lithium-Ion Batteries Based on Wavelet Noise Reduction and Support Vector Machine. *Trans. China Electrotech. Soc.* **2020**, *35*, 3126–3136.

29. Zhang, Z.L.; Cheng, X.; Lu, Z.Y.; Gu, D.J. SOC Estimation of Lithium-Ion Batteries with AEKF and Wavelet Transform Matrix. *IEEE Trans. Power Electron.* **2017**, *32*, 7626–7634. [\[CrossRef\]](#)
30. Wang, Z.K.; Zeng, S.K.; Guo, J.B.; Qin, T.C. State of health estimation of lithium-ion batteries based on the constant voltage charging curve. *Energy* **2019**, *167*, 661–669. [\[CrossRef\]](#)
31. Dai, H.D.; Zhao, G.C.; Lin, M.Q.; Wu, J.; Zheng, G.F. A Novel Estimation Method for the State of Health of Lithium-Ion Battery Using Prior Knowledge-Based Neural Network and Markov Chain. *IEEE Trans. Ind. Electron.* **2019**, *66*, 7706–7716. [\[CrossRef\]](#)
32. Hannan, M.A.; How, D.N.T.; Lipu, M.S.H.; Ker, P.J.; Dong, Z.Y.; Mansur, M.; Blaabjerg, F. SOC Estimation of Li-ion Batteries With Learning Rate-Optimized Deep Fully Convolutional Network. *IEEE Trans. Power Electron.* **2021**, *36*, 7349–7353. [\[CrossRef\]](#)
33. Ren, L.; Dong, J.B.; Wang, X.K.; Meng, Z.H.; Zhao, L.; Deen, M.J. A Data-Driven Auto-CNN-LSTM Prediction Model for Lithium-Ion Battery Remaining Useful Life. *IEEE Trans. Ind. Inform.* **2021**, *17*, 3478–3487. [\[CrossRef\]](#)
34. Severson, K.A.; Attia, P.M.; Jin, N.; Perkins, N.; Jiang, B.; Yang, Z.; Chen, M.H.; Aykol, M.; Herring, P.K.; Fraggadakis, D.; et al. Data-driven prediction of battery cycle life before capacity degradation. *Nat. Energy* **2019**, *4*, 383–391. [\[CrossRef\]](#)
35. Khelif, R.; Chebel-Morello, B.; Malinowski, S.; Laajili, E.; Fnaiech, F.; Zerhouni, N. Direct Remaining Useful Life Estimation Based on Support Vector Regression. *IEEE Trans. Ind. Electron.* **2017**, *64*, 2276–2285. [\[CrossRef\]](#)
36. Zhang, Y.J.; Liu, Y.J.; Wang, J.; Zhang, T. State-of-health estimation for lithium-ion batteries by combining model-based incremental capacity analysis with support vector regression. *Energy* **2022**, *239*, 121986. [\[CrossRef\]](#)
37. Cai, L.; Meng, J.H.; Stroe, D.I.; Peng, J.C.; Luo, G.Z.; Teodorescu, R. Multiobjective Optimization of Data-Driven Model for Lithium-Ion Battery SOH Estimation With Short-Term Feature. *IEEE Trans. Power Electron.* **2020**, *35*, 11855–11864. [\[CrossRef\]](#)
38. Fatima, N. Enhancing Performance of a Deep Neural Network: A Comparative Analysis of Optimization Algorithms. *Adcaij-Adv. Distrib. Comput. Artif. Intell. J.* **2020**, *9*, 79–90. [\[CrossRef\]](#)
39. Wang, L.; Xie, L.; Yang, Y.; Zhang, Y.; Wang, K.; Cheng, S. Distributed Online Voltage Control with Fast PV Power Fluctuations and Imperfect Communication. *IEEE Trans. Smart Grid* **2023**. [\[CrossRef\]](#)
40. Senior, A.; Heigold, G.; Ranzato, M.; Yang, K. An empirical study of learning rates in deep neural networks for speech recognition. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vancouver, BC, Canada, 26–31 May 2013; pp. 6724–6728.
41. Khalid, A.; Sarwat, A.I. Unified Univariate-Neural Network Models for Lithium-Ion Battery State-of-Charge Forecasting Using Minimized Akaike Information Criterion Algorithm. *IEEE Access* **2021**, *9*, 39154–39170. [\[CrossRef\]](#)
42. Ge, D.D.; Zhang, Z.D.; Kong, X.D.; Wan, Z.P. Extreme Learning Machine Using Bat Optimization Algorithm for Estimating State of Health of Lithium-Ion Batteries. *Appl. Sci.-Basel* **2022**, *12*, 1398. [\[CrossRef\]](#)
43. Khalid, A.; Sundararajan, A.; Acharya, I.; Sarwat, A.I. Prediction of Li-Ion Battery State of Charge Using Multilayer Perceptron and Long Short-Term Memory Models. In Proceedings of the IEEE Transportation Electrification Conference and Expo (ITEC), Detroit, MI, USA, 19–21 June 2019.
44. Khalid, A.; Sarwat, A.I. Fast Charging Li-Ion Battery Capacity Fade Prognostic Modeling Using Correlated Parameters' Decomposition and Recurrent Wavelet Neural Network. In Proceedings of the IEEE Transportation Electrification Conference and Expo (ITEC), Chicago, IL, USA, 21–25 June 2021; pp. 27–32.
45. Chemali, E.; Kollmeyer, P.J.; Preindl, M.; Ahmed, R.; Emadi, A. Long Short-Term Memory Networks for Accurate State-of-Charge Estimation of Li-ion Batteries. *IEEE Trans. Ind. Electron.* **2018**, *65*, 6730–6739. [\[CrossRef\]](#)
46. Li, W.H.; Sengupta, N.; Dechent, P.; Howey, D.; Annaswamy, A.; Sauer, D.U. Online capacity estimation of lithium-ion batteries with deep long short-term memory networks. *J. Power Sources* **2021**, *482*, 228863. [\[CrossRef\]](#)
47. Yang, F.F.; Zhang, S.H.; Li, W.H.; Miao, Q. State-of-charge estimation of lithium-ion batteries using LSTM and UKF. *Energy* **2020**, *201*, 117664. [\[CrossRef\]](#)
48. Cui, Z.H.; Kang, L.; Li, L.W.; Wang, L.C.; Wang, K. A hybrid neural network model with improved input for state of charge estimation of lithium-ion battery at low temperatures. *Renew. Energy* **2022**, *198*, 1328–1340. [\[CrossRef\]](#)
49. Li, D.; Yang, D.; Li, L.; Wang, L.; Wang, K. Electrochemical Impedance Spectroscopy Based on the State of Health Estimation for Lithium-Ion Batteries. *Energies* **2022**, *15*, 6665. [\[CrossRef\]](#)
50. Sun, H.L.; Yang, D.F.; Wang, L.C.; Wang, K. A method for estimating the aging state of lithium-ion batteries based on a multi-linear integrated model. *Int. J. Energy Res.* **2022**, *46*, 24091–24104. [\[CrossRef\]](#)
51. Qian, C.; Xu, B.H.; Chang, L.; Sun, B.; Feng, Q.; Yang, D.Z.; Ren, Y.; Wang, Z.L. Convolutional neural network based capacity estimation using random segments of the charging curves for lithium-ion batteries. *Energy* **2021**, *227*, 120333. [\[CrossRef\]](#)
52. Yang, Y.X. A machine-learning prediction method of lithium-ion battery life based on charge process for different applications. *Appl. Energy* **2021**, *292*, 116897. [\[CrossRef\]](#)
53. Wang, W.L.; Yang, D.F.; Yan, X.R.; Wang, L.C.; Hu, H.; Wang, K. Triboelectric nanogenerators: The beginning of blue dream. *Front. Chem. Sci. Eng.* **2023**. [\[CrossRef\]](#)
54. Zhang, M.; Wang, W.; Xia, G.; Wang, L.; Wang, K. Self-Powered Electronic Skin for Remote Human–Machine Synchronization. *ACS Appl. Electron. Mater.* **2023**, *5*, 498–508. [\[CrossRef\]](#)

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.