Abstract: Electric vehicles can reduce the dependence on limited resources such as oil, which is conducive to the development of clean energy. An accurate battery state of health (SOH) is beneficial for the safety of electric vehicles. A multi-feature and Convolutional Neural Network–Bidirectional Long Short-Term Memory–Multi-head Attention (CNN-BiLSTM-MHA)-based lithium-ion battery SOH estimation method is proposed in this paper. First, the voltage, energy, and temperature data of the battery in the constant current charging phase are measured. Then, based on the voltage and energy data, the incremental energy analysis (IEA) is performed to calculate the incremental energy (IE) curve. The IE curve features including IE, peak value, average value, and standard deviation are extracted and combined with the thermal features of the battery to form a complete multi-feature sequence. A CNN-BiLSTM-MHA model is set up to map the features to the battery SOH. Experiments were conducted using batteries with different charging currents, and the results showed that even if the nonlinearity of battery SOH degradation is significant, this method can still achieve a fast and accurate estimation of the battery SOH. The Mean Absolute Error (MAE) is 0.1982%, 0.1873%, 0.1652%, and 0.1968%, and the Root-Mean-Square Error (RMSE) is 0.2921%, 0.2997%, 0.2130%, and 0.2625%, respectively. The average Coefficient of Determination ($R^2$) is above 96%. Compared to the BiLSTM model, the training time is reduced by an average of about 36%.

Keywords: lithium-ion battery; state of health (SOH); incremental energy analysis (IEA); Convolutional Neural Network (CNN); Multi-Head Attention (MHA); Bidirectional Long Short-Term Memory (BiLSTM)

1. Introduction

Electric vehicles have the advantages of good environmental protection, high energy efficiency, and low operating costs. They can reduce the dependence on limited resources such as oil and help reduce carbon emissions, making them widely used in personal transportation, commercial transportation, and special purpose vehicles [1–3]. As an important energy storage device for electric vehicles, lithium-ion batteries play a crucial role in the lifecycle of electric vehicles. However, with the use of batteries, their performance and state of health (SOH) will gradually deteriorate, limiting the service life of vehicles. Therefore, the SOH has become a key indicator for evaluating the degree of battery aging [4,5]. More importantly, the safety of batteries is closely related to their SOH. When batteries age or are damaged, the probability of catastrophic events such as self-ignition or explosion significantly increases [6,7]. Therefore, accurately estimating the SOH of batteries is an indispensable measure to ensure the safety of power batteries and extend their lifespan, which is conducive to improving the reliability of battery management systems [8–10].

Domestic and international scholars have proposed various methods for estimating the SOH of lithium-ion batteries, which can generally be divided into two categories: model-based methods [11–18] and data-driven methods [19–26]. Model-based methods involve establishing appropriate equivalent models of lithium-ion batteries to simulate
internal battery structures, materials, and chemical reaction processes, thereby predicting
the battery’s performance, response, and SOH. In recent years, model-based methods
have primarily included electrochemical models, equivalent circuit models, and empirical
degradation models. For instance, Li et al. [11] developed a degradation model based on
single-particle behavior, considering crack propagation caused by the formation of solid
electrolyte interface layers and volume expansion of particles in active materials. Yang
et al. [12] considered the limited operating range of batteries in practice and overpotential
cau sed by high current rates, establishing a voltage reconstruction model. Xu et al. [13]
utilized a simplified fractional impedance model and a least squares genetic algorithm to
identify model parameters based on intercell capacity utilization. These methods require
the establishment of complex mathematical or physical models, and typically, when battery
operating conditions change, the models need to be re-established, leading to significant
limitations in their applicability.

In recent years, with the rapid development of deep learning, data-driven methods
for estimating the SOH of lithium-ion batteries have become mainstream. These methods
are simple, easy to use, and widely applicable. They do not require detailed battery
equivalent models or complex parameter estimation processes, making them easier to
deploy in practical applications. Data-driven methods mainly consist of three key steps:
data collection, feature extraction, and model training. In the feature extraction step, it
is crucial to extract features strongly correlated with the battery SOH, enabling better
mapping between features and the SOH through deep learning models. The selection of
deep learning models is also crucial. A suitable deep learning model can not only train
the mapping between features and the SOH quickly and accurately but also exhibit strong
generalization to different datasets. Popular data-driven models in recent years include
Gaussian Process Regression (GPR) [19], Support Vector Machine (SVM) [20], Long Short-
Term Memory (LSTM) [21], and Gated Recurrent Unit (GRU) [22]. Building upon these
traditional data-driven models, many scholars have proposed more innovative methods
combining these models, including more innovative feature extraction and specific methods
to optimize traditional machine learning and deep learning models. For example, Zhang
et al. [23] combined the equivalent model method with the data-driven method, establishing
an improved equivalent circuit model. They used model parameters as health indicators,
utilized backpropagation neural networks to build SOH prediction models, and optimized
model parameters using an improved particle swarm optimization algorithm. Xue et al. [24]
addressed the problem of the decreased effectiveness of traditional data-driven methods
due to non-Gaussian noise by proposing a novel Robust Stacked Integrated Learning Model
(ILM). They used Mixture Correntropy Loss (MCL) for an Extreme Learning Machine (ELM)
and GRU to form new MCL-ELM and MCL-GRU models, serving as sub-models of the
proposed ILM. Xu et al. [25] established a Convolutional Neural Network–Long Short-Term
Memory (CNN-LSTM) model and introduced skip connections in the model to address the
neural network degradation caused by multiple layers of LSTM. Zhang et al. [26] proposed
an SOH estimation model based on the incremental capacity (IC) curve collection method
improved by reference voltage. They extracted important feature variables from IC curves
and combined them with LSTM to establish an SOH estimation model.

Taking into account the advantages and disadvantages of the data-driven methods
mentioned above, a lithium-ion battery SOH estimation method based on multi-feature
and Convolutional Neural Network–Bidirectional Long Short-Term Memory–Multi-head
Attention (CNN-BiLSTM-MHA) is proposed in this paper. This method has extremely high
accuracy in SOH estimation and excellent training speed. The main contributions of this
paper are summarized as follows:

(1) Based on the constant current charging phase data, comprehensive multi-features
are extracted. This paper considers that the existing methods extract relatively single
features. Firstly, the voltage and energy data of the battery during the constant current
charging phase are measured, and incremental energy analysis (IEA) is performed to
calculate the incremental energy (IE) curve. The IE curve features, including IE, peak value,
average value, and standard deviation, are extracted. It is verified through the Pearson correlation coefficient method that the IE curve features have a strong correlation with the battery SOH. Subsequently, the IE curve features are combined with the thermal features of the battery to form a complete multi-feature sequence. Compared with the relatively single features extracted by traditional feature extraction methods, the multi-features extracted in this paper contain a more comprehensive nonlinear aging trend of the battery SOH. Therefore, the deep learning model can learn richer battery SOH aging information, thereby establishing a more accurate battery SOH estimation model.

(2) A highly efficient CNN-BiLSTM-MHA deep learning model is set up. The unique capability of the Convolutional Neural Network (CNN) to extract key trends in sequences is utilized in this paper. It effectively combines the original multi-feature sequence into local abstract features, reducing the computational burden on subsequent models. Following the CNN, Bidirectional Long Short-Term Memory (BiLSTM), which has the ability to capture long-term dependencies in sequences, is used. By calculating the sequence in both the forward and backward directions, the model can keenly capture the aging trend of the battery SOH. Additionally, Multi-head Attention (MHA) is applied to the hidden states of BiLSTM, effectively alleviating the problem of limited information propagation in hidden layers caused by gradient explosion and gradient vanishing in BiLSTM, thereby enhancing the model’s generalization capability. Consequently, the CNN-BiLSTM-MHA model can achieve excellent SOH estimation performance, including accuracy, speed, and robustness.

(3) Comprehensive and rigorous SOH estimation validation and model comparison: Following the method proposed in this paper, SOH estimation is conducted using four battery aging datasets with different charging rates. The CNN-BiLSTM-MHA model is systematically compared with the CNN-BiLSTM, BiLSTM, BiGRU, and SVR models. The experimental results indicate that, even with the pronounced nonlinear degradation process of the battery SOH, the proposed method demonstrates superior SOH estimation performance. Across the four datasets, the Mean Absolute Error (MAE) remains within 0.2%, the Root-Mean-Square Error (RMSE) within 0.3%, and the average Coefficient of Determination ($R^2$) exceeds 96%. Compared to the BiLSTM model, the MAE and RMSE are reduced by approximately 36% and 35%, respectively, and the training time is reduced by an average of about 36%.

In summary, the measurement data were utilized for comprehensive feature extraction in this paper, and an SOH estimation model combining accuracy and speed was established, which was effectively demonstrated through experimental verification and comparison.

The rest of this paper is structured as follows: Section 2 provides a detailed introduction to multi-features and their extraction methods. Section 3 introduces the proposed CNN-BiLSTM-MHA model and analyzes its advantages. Section 4 introduces the acquisition of battery aging data and uses four different charging rate datasets to conduct experiments based on the method proposed in this paper. The experimental results are obtained, followed by model comparison to verify the superiority of the SOH estimation model established in this paper. Section 5 concludes this paper. Finally, Section 6 presents the limitations of this study.

2. Feature Extraction

Constant current–constant voltage (CC-CV) charging is the most widely used charging method for lithium-ion batteries. As shown in Figure 1, it depicts the curves of voltage and current over time during the CC-CV charging process of a lithium-ion battery. In this paper, data from the constant voltage charging phase are selected for a comprehensive and detailed analysis of the IE curve. Combined with the temperature during the constant voltage charging phase, comprehensive health indices are extracted, and the Pearson correlation coefficient method is utilized to analyze the correlation of features extracted from the IE curve.
2.1. Feature Extraction Based on IE Curve and Correlation Analysis

Traditional incremental capacity analysis (ICA) is commonly applied to analyze the aging mechanism of batteries during the constant current charging and discharging phases. By collecting capacity and terminal voltage data during the constant current charging and discharging phases, the original capacity–voltage (Q-V) curve is transformed into an incremental capacity (dQ/dV-V) curve through first-order differentiation, allowing for the extraction of more comprehensive features from the transformed curve. However, battery charging and discharging involve changes in both current and voltage. Battery capacity calculation is only related to current and time during charging and discharging, ignoring the voltage variable. Therefore, in this paper, the original energy–voltage (E-V) curve is transformed into an IE (dE/dV-V) curve through first-order differentiation, which is conducive to analyzing and extracting more comprehensive and effective features to reflect the battery aging process.

Taking the laboratory cycling aging dataset of batteries with a 0.5C charging rate as an example, the E-V curves and IE curves at different cycle numbers are shown in Figures 2 and 3, respectively.

Figure 1. CC-CV charging curve.

Figure 2. E-V curve under different cycles.
Figure 3. IE curve under different cycles.

2.1.1. IE during Constant Current Charging Phase

Analyzing Figure 2 reveals that the vertical span of the E-V curve, i.e., the span of IE, decreases with an increase in the cycle number. This indicates that during the constant current charging phase, the battery’s acceptable external energy decreases due to battery aging. By calculating the energy added during each cycle of constant current charging, the results shown in Figure 4 are obtained. From Figure 4, it can be observed that the overall trend of the energy added during the battery’s constant current charging phase exhibits a decreasing trend, but there are also local sharp increases and decreases. These fluctuations are somewhat associated with the phenomenon of battery capacity regeneration. Therefore, using IE as a feature can effectively establish a mapping with the battery SOH through deep learning models.

The IE is calculated as follows:

\[ \Delta E = E_{\text{end}} - E_{\text{start}} \]  

(1)

where \( \Delta E \) represents IE, and \( E_{\text{end}} \) and \( E_{\text{start}} \) respectively denote the final and initial energy during the constant current charging phase.

Figure 4. The curve of IE variation with cycles.

2.1.2. The Peak and Average Values of the IE Curve

Analyzing the IE curve shown in Figure 3 reveals that with an increase in the cycle number, the peak of the IE curve decreases correspondingly. This peak reflects the peak rate of IE. By extracting the peaks of the IE curves for each cycle, the results illustrated in
Figure 5 are obtained. As seen in Figure 5, the peaks of the IE curve exhibit a stair-step decreasing trend, indicating that during the battery aging process, the peak rate of IE periodically enters a stable phase after a steep decline. This reflects the maximum intensity of internal chemical reactions within the battery.

![Figure 5. The peak value of the IE curve varies with cycles.](image)

To comprehensively analyze the intensity of internal chemical reactions within the battery through the IE curve, the average values of the IE curves for each cycle are calculated. The results are presented in Figure 6. As Figure 6 shows, the average values of the IE curve, representing the average rate of IE, exhibit a decreasing trend overall, with occasional minor sharp increases and decreases. This indicates that as the battery ages, the average intensity of internal chemical reactions gradually weakens.

The average value of the IE curve is calculated as follows:

\[
\text{Average value} = \frac{1}{b-a} \int_a^b f(x)dx
\]

where Average value represents the average value of the IE curve, \(a\) and \(b\) denote the starting and ending voltages of the IE curve, respectively, and \(f(x)\) represents the function of the IE curve.

![Figure 6. The average value of the IE curve varies with cycles.](image)

By extracting the peak value and average value of the IE curve, it is possible to effectively reflect the peak rate and average rate of IE, thus establishing a certain correlation with the internal chemical reactions within the battery. Therefore, selecting the peaks and
averages of the IE curve as features can effectively establish a mapping with the battery SOH through deep learning models.

2.1.3. The Standard Deviation of the IE Curve

The standard deviation reflects the degree to which the values in the dataset deviate from the mean value. In the context of the IE curve, it indicates the stability of the IE rate. By calculating the standard deviation of the IE curve for each cycle, the results presented in Figure 7 are obtained. Figure 7 shows that the standard deviation of the IE curve exhibits a decreasing trend, indicating that as the battery ages, the IE rates of each cycle tend towards their average rate. The stability of the IE rate gradually increases, while there are also sharp increases and decreases associated with battery capacity regeneration. Therefore, selecting this feature can effectively establish a mapping with the battery SOH through deep learning models.

The standard deviation of the IE curve is calculated as follows:

\[ \text{Std} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (f(x_i) - \overline{f})^2} \]  

where \( \text{Std} \) represents the standard deviation of the IE curve, \( f(x_i) \) represents the discrete data points of the IE curve, \( \overline{f} \) denotes the mean value of the IE curve, and \( N \) is the number of data points.

![Figure 7](image)

**Figure 7.** The standard deviation of the IE curve varies with cycles.

2.1.4. Correlation Analysis

To further validate the correlation between features extracted from the IE curve and the battery SOH, the Pearson correlation coefficient method is used for correlation analysis between IE curve features and the SOH in this paper. The Pearson correlation coefficient method is commonly used to measure the degree of the linear relationship between two variables, with values ranging from \(-1\) to \(1\). A correlation coefficient of \(1\) indicates a perfect positive correlation between two variables, \(-1\) indicates a perfect negative correlation, and \(0\) indicates no linear relationship between the variables. Therefore, if the absolute value of the Pearson correlation coefficient between the extracted features and battery SOH approaches \(1\), it indicates that the extracted features can better reflect the decay of the battery SOH in terms of a linear relationship.

The Pearson correlation coefficient is defined as follows:

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

(4)
where \( r \) represents the Pearson correlation coefficient, \( X_i \) and \( Y_i \) respectively denote the \( i \)-th values in the two variables, \( \bar{X} \) and \( \bar{Y} \) represent the mean values of the two variables, and \( n \) denotes the sample size.

By calculating the Pearson correlation coefficients between the features extracted from the IE curve and battery SOH, the results are obtained as shown in Table 1.

Table 1. The Pearson correlation coefficient between the features of each IE curve and the battery SOH.

<table>
<thead>
<tr>
<th>IE</th>
<th>Peak Value</th>
<th>Average Value</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.9982</td>
<td>0.9648</td>
<td>0.9909</td>
</tr>
</tbody>
</table>

From Table 1, it can be observed that the Pearson correlation coefficients between the features extracted from the IE curve and battery SOH are all close to 1. This indicates a very strong correlation between the features and battery SOH. Therefore, selecting these features from the IE curve allows for the aging pattern of the battery to be accurately tracked by deep learning models.

2.2. Thermal Features

The temperature and SOH of lithium-ion batteries are closely related \([27,28]\), primarily due to the physical and chemical changes that occur within the battery. At elevated temperatures, secondary reactions and degradation phenomena may occur, such as the oxidation of the cathode material and decomposition of the electrolyte. These reactions and degradation phenomena accelerate the aging process of the battery, leading to a decrease in the SOH. Conversely, at low temperatures, the flowability of the electrolyte inside the battery decreases significantly, resulting in slower ion transport within the battery. This can lead to reduced efficiency during the battery charging and discharging processes, thereby rapidly decreasing the battery SOH. However, at an appropriate temperature, the flowability of the electrolyte, chemical reaction rates, and material stability within the lithium-ion battery reach optimal conditions. As a result, the aging rate of the battery slows down, and capacity regeneration phenomena may occur.

Considering the impact of temperature on the SOH of lithium-ion batteries, the average temperature of the lithium-ion battery during the constant current charging phase is extracted as a thermal feature in this paper. This enables deep learning models to learn richer information about battery aging, thereby accurately tracking the aging patterns of the battery.

The average charging temperature is calculated as follows:

\[
T_{\text{avg}} = \frac{T_1 + T_2 + \cdots + T_n}{n}
\]

where \( T_{\text{avg}} \) represents the average charging temperature, \( T_1 \) to \( T_n \) denote the temperatures of each sampling point, and \( n \) represents the number of sampling points.

3. CNN-BiLSTM-MHA Model

3.1. CNN

A CNN is a deep learning algorithm inspired by the structure and principles of the visual cortex in animals. It effectively learns spatial hierarchical relationships in data by simulating the visual perception mechanism of animals. While CNNs are primarily applied to image recognition and processing tasks, their ability to capture local abstract features also makes them perform well on time series data processing tasks. Therefore, it can be effectively applied in the field of battery SOH estimation \([29]\).

Due to the excellent performance of One-Dimensional Convolutional Neural Networks (1D CNNs) in time series processing tasks \([30]\), a 1D CNN is utilized in this paper to process time series data. The 1D CNN is used to extract key trends in the IE curve features and thermal features, transforming the critical trends in the original features into abstract local
important features. This provides valuable information for the analysis and prediction of subsequent models.

The CNN in this paper includes the convolutional layer and the pooling layer. The basic mathematical formulas and descriptions of these operations are as follows:

1) Convolutional layer:

\[ y(t) = (f \times x)(t) = \sum_{s=-\infty}^{\infty} x(s) \cdot f(t-s) \tag{6} \]

where \( y \) represents the output of the convolutional layer, \( x \) represents the one-dimensional input signal, \( f \) represents the convolution kernel (filter), \( s \) represents the index within the convolution kernel, and \( t \) represents the index of time or space.

For each filter in the convolutional layer, such an operation is performed, generating the corresponding feature map.

2) Activation Function: The ReLU function is chosen as the activation function. The ReLU activation function is defined as the following:

\[ h(t) = \max(0, y(t)) \tag{7} \]

where \( h(t) \) represents the output after activation, \( \max \) denotes the maximum operation, and \( y(t) \) is the input.

This operation means that if the convolutional output \( y(t) \) is greater than 0, the output of the activation function is \( y(t) \); if it is less than or equal to 0, the output is 0. This nonlinear transformation helps the model learn complex data representations.

3) Max Pooling Layer: The max pooling operation selects the maximum value within a specified range as the output for that range. The max pooling operation is defined as the following:

\[ m(t) = \max_{s=1}^{t+p-1} h(s) \tag{8} \]

where \( m(t) \) represents the output after the pooling operation, \( h \) is the input, \( p \) represents the pooling window size, and max operation applies within the range of indices from \( t \) to \( t + p - 1 \), i.e., selecting the maximum value within each pooling window.

The structure diagram of the 1D CNN in this paper is shown in Figure 8:

![Figure 8. The 1D CNN structure.](image-url)

In the figure, the output of the convolutional layer is determined by multiple original sequence data, depending on the choice of the size of the convolutional kernel and stride. The output of the convolutional layer is activated through the ReLU activation function and finally obtained through the max pooling layer, which constitutes the output of the entire 1D CNN.

The CNN in this paper first utilizes convolutional operations to extract abstract local important features from the original feature sequences. Subsequently, pooling operations
are employed to reduce the dimensionality of the abstract features, resulting in abstract local important features. These abstract local important features can more accurately describe the periodic variations and anomalies in the sequences, thereby alleviating to some extent the overfitting risks caused by the multicollinearity of the original features. Moreover, since the abstract local important features are fewer in quantity compared to the original feature sequences, they effectively reduce the computational burden of subsequent models, thereby accelerating the overall training speed.

3.2. BiLSTM-MHA

BiLSTM is a special type of a recurrent neural network (RNN) that comprehensively captures long-term dependencies in sequence data from both the forward and backward directions. Its main applications include text generation, speech recognition, machine translation, and time series prediction, among others. In this paper, BiLSTM is utilized to process time series data, where the input sequence is the output of the aforementioned CNN.

The computation process of a standard LSTM is as follows:

(1) First, utilizing the external state from the previous time step and the current input at the current time step, the forget gate, input gate, output gate, and candidate state cell are computed. The mathematical formulations are defined as the following:

\[ f_t = \sigma \left( W_f \cdot [h_{t-1}, x_t] + b_f \right) \]  
\[ i_t = \sigma \left( W_i \cdot [h_{t-1}, x_t] + b_i \right) \]  
\[ o_t = \sigma \left( W_o \cdot [h_{t-1}, x_t] + b_o \right) \]  
\[ \tilde{C}_t = \tanh \left( W_C \cdot [h_{t-1}, x_t] + b_C \right) \]

where \( f_t, i_t, o_t, \) and \( \tilde{C}_t \) represent the forget gate, input gate, output gate, and candidate state cell, respectively. \( W_f, W_i, W_o, \) and \( W_C \) denote the weights for the forget gate, input gate, output gate, and candidate state cell, while \( b_f, b_i, b_o, \) and \( b_C \) represent the biases. \( \sigma \) denotes the Sigmoid activation function, which compresses the result between 0 and 1, and \( \tanh \) represents the hyperbolic tangent activation function, which compresses the result between \(-1\) and \(1\).

(2) Updating the memory cell is conducted by combining the forget gate and the input gate. The mathematical formulation is defined as the following:

\[ C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \]  

where \( C_t \) represents the updated memory cell, \( C_{t-1} \) denotes the memory cell to be updated, \( f_t \) and \( i_t \) respectively represent the forget gate and the input gate, and \( \tilde{C}_t \) represents the candidate cell state.

(3) Combining with the output gate, the information from the internal state is passed to the external state. The mathematical formulation is defined as the following:

\[ h_t = o_t \cdot \tanh(C_t) \]

where \( h_t \) represents the external state, \( C_t \) denotes the updated memory cell, \( o_t \) represents the output gate, and \( \tanh \) is the hyperbolic tangent activation function.

For BiLSTM, at each time step, both the forward and backward external states are computed simultaneously. Then, these two external states are combined to form the final external state.

Although BiLSTM performs well in handling time series tasks, it suffers from the problem of limited information propagation. This is because at each time step, information can only be propagated through hidden states. For sequences with long-term dependencies, the hidden states may be affected by issues like vanishing or exploding gradients, leading
to inefficient information propagation. To address the limitation of information propagation in BiLSTM, MHA is applied to the hidden state sequences of BiLSTM in this paper.

MHA is a mechanism in deep learning used to handle sequential data. It captures the relationships between different positions in the input sequence and improves the representation of the sequence by adding appropriate weights to it through the parallel computation and concatenation of multiple attention heads. This enhances the model’s generalization ability.

Assuming the input sequence is \( X = \{x_1, x_2, \ldots, x_n\} \), where \( x_i \) is the \( i \)-th element in the sequence and has a dimension of \( d \), the computation process of MHA is as follows:

(1) Linear transformations are applied to the input sequence to obtain queries (\( Q \)), keys (\( K \)), and values (\( V \)). The mathematical formulations are defined as the following:

\[
Q = XW_Q \\
K = XW_K \\
V = XW_V
\]

where \( W_Q \), \( W_K \), and \( W_V \) represent the corresponding weight matrices.

(2) For each pair of query \( q_i \) and key \( k_j \), the attention score is calculated. The mathematical formulation is defined as the following:

\[
\text{Attention}(q_i, k_j) = \frac{q_i k_j^T}{\sqrt{d_k}}
\]

where \( \text{Attention}(q_i, k_j) \) represents the attention score, \( \sqrt{d_k} \) is the scaling factor, \( d_k \) is typically equal to \( d/h \), and \( h \) is the number of attention heads.

(3) The attention scores are normalized. The mathematical formulation is defined as the following:

\[
\alpha_{ij} = \text{softmax}(\text{Attention}(q_i, k_j))
\]

where \( \alpha_{ij} \) denotes the attention weight, and the Softmax function is applied to normalize \( \text{Attention}(q_i, k_j) \).

(4) The attention weights are used to weight and sum the values \( V \), obtaining the attention representation for each query \( q_i \). The mathematical formulation is defined as the following:

\[
\text{Attention}(q_i, V) = \sum_{j=1}^{n} \alpha_{ij} v_j
\]

where \( \text{Attention}(q_i, V) \) represents the attention representation for each query \( q_i \), \( \alpha_{ij} \) denotes the attention weight, and \( v_j \) represents the value vector corresponding to the \( j \)-th element.

(5) All attention head outputs are concatenated, and another linear transformation is performed. The mathematical formulation is defined as the following:

\[
\text{MHA}(X) = \text{Concat} \left( \text{Attention}^1, \ldots, \text{Attention}^h \right) W_O
\]

where \( \text{MHA}(X) \) represents the output of the MHA, Concat denotes the concatenation operation, \( \text{Attention}^1, \ldots, \text{Attention}^h \) represent the outputs of each attention head, and \( W_O \) is the weight matrix.

The structure of BiLSTM-MHA in this paper is shown in Figure 9.

In this paper, the input of BiLSTM-MHA is the output of the CNN. The CNN effectively extracts the key variation trends of the original feature sequence as abstract local important features and reduces the amount of sequence data, thereby alleviating the computational burden of BiLSTM-MHA. MHA is applied to the hidden states of BiLSTM, allowing the model to consider the importance of input sequences at different positions at each time step. This helps alleviate the vanishing or exploding gradient problems in BiLSTM, enabling the
hidden states of BiLSTM to be updated correctly, thus improving model training stability and prediction accuracy.

Figure 9. BiLSTM-MHA structure.

4. The Experimental Data, Results, Comparisons, and Analysis

4.1. Experimental Data

This paper conducts battery charge–discharge cycle aging experiments on cylindrical lithium-ion batteries with the same specifications in indoor ambient temperature conditions. The rated capacity of the batteries is 2.5 Ah. Firstly, constant current–constant voltage charging is used at rates of 0.1C, 0.2C, 0.3C, and 0.5C until the battery terminal voltage reaches the upper cutoff voltage. Then, the voltage is kept constant for constant voltage charging until the current decreases to the cutoff current. After a 5-min rest period to stabilize the internal chemical state of the battery, a constant current discharge at 0.5C is performed until the battery terminal voltage decreases to the lower cutoff voltage. There are various definitions of the battery SOH, and the capacity ratio is selected in this paper to define the SOH of lithium-ion batteries. This definition combines reliability and intuitiveness and is defined as the following [31]:

\[
SOH = \frac{Q_{\text{max capacity}}}{Q_{\text{rated capacity}}} \times 100\% \tag{22}
\]

where \(Q_{\text{max capacity}}\) represents the current maximum discharge capacity of the lithium battery, and \(Q_{\text{rated capacity}}\) represents the rated capacity of the lithium battery when it left the factory.

Figure 10a–d respectively represent the SOH of four lithium-ion batteries charged at 0.1C, 0.2C, 0.3C, and 0.5C as the number of charge–discharge cycles increases. From the figure, it can be observed that the SOH of the batteries generally decreases with the increase in charge–discharge cycles. However, due to various comprehensive factors affecting the batteries during the charge–discharge cycle process, this process exhibits significant nonlinearity. Additionally, noticeable capacity regeneration phenomena occur at multiple different numbers of charge–discharge cycles. These phenomena pose certain difficulties for estimating the SOH of lithium-ion batteries.

The cycling aging parameters for batteries at four different charge rates are shown in Table 2.
4. The Experimental Data, Results, Comparisons, and Analysis

4.1. Experimental Data

According to the second part of this paper, the IE curve features and thermal features of lithium-ion batteries at different charge rates are extracted, combined with their corresponding SOH values, to form a complete dataset. In order to maintain the overall distribution balance of the dataset and ensure that the model has good generalization ability, reducing the risk of underfitting or overfitting and obtaining reliable evaluation results, the training set and test set are split in a 1:1 ratio. That is, the first 50% of the cycles are used as the training set, and the last 50% of the cycles are used as the test set.

The cycling aging parameters for batteries at four different charge rates are shown in Table 2. The experimental flowchart is shown in Figure 11.

**Table 2. Battery aging parameters.**

<table>
<thead>
<tr>
<th>Charging Rates</th>
<th>Cycles</th>
<th>Initial SOH (%)</th>
<th>Final SOH (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1C (250 mA)</td>
<td>309</td>
<td>100.2880</td>
<td>90.0640</td>
</tr>
<tr>
<td>0.2C (500 mA)</td>
<td>470</td>
<td>99.7120</td>
<td>88.0200</td>
</tr>
<tr>
<td>0.3C (750 mA)</td>
<td>472</td>
<td>100.5240</td>
<td>89.0920</td>
</tr>
<tr>
<td>0.5C (1250 mA)</td>
<td>507</td>
<td>101.3000</td>
<td>82.1360</td>
</tr>
</tbody>
</table>

4.2. Experimental Results

According to the second part of this paper, the IE curve features and thermal features of lithium-ion batteries at different charge rates are extracted, combined with their corresponding SOH values, to form a complete dataset. In order to maintain the overall distribution balance of the dataset and ensure that the model has good generalization ability, reducing the risk of underfitting or overfitting and obtaining reliable evaluation results, the training set and test set are split in a 1:1 ratio. That is, the first 50% of the cycles are used as the training set, and the last 50% of the cycles are used as the test set.

The training set data are input into the CNN-BiLSTM-MHA model constructed in the third part of this paper. In this model, the default number of filters for the CNN and the size of the convolutional kernel are 64 and 3, respectively. The default number of
neurons for BiLSTM is 128, which can be adjusted appropriately based on the complexity of the task. The number of attention heads and the dimension of keys are both set to 2. The learning rate is adaptively updated by the Adam optimizer, which helps stabilize the internal parameters of the model and converge faster to the global optimal solution. The training is conducted for 100 epochs.

The experimental flowchart is shown in Figure 11.

The SOH estimation results are shown in Figure 12a–d. To rigorously evaluate the accuracy and speed of SOH estimation, this experiment employs common metrics for assessing the performance of regression models, namely the MAE, RMSE, $R^2$, and training time. The MAE, RMSE, and $R^2$ are defined as the following:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$  \hspace{2cm} (23)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$ \hspace{2cm} (24)

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$ \hspace{1cm} (25)

In this experiment, $n$ represents the number of charge–discharge cycles, $y_i$ denotes the true SOH value, $\bar{y}$ represents the average true SOH value, and $\hat{y}_i$ stands for the estimated SOH value. For the MAE and RMSE, smaller values indicate a higher estimation accuracy of the model. For $R^2$, a value closer to 1 indicates a better fitting performance of the model. The SOH estimation performance of the proposed method is shown in Table 3.
Battery testing

Step1: Data measurement

Raw data

Temperature

Step2: Feature extraction

Peak value

IE curve features

\[ x_i = \frac{\sum_{n=1}^{N} x_{ni}}{N} \]

\[ y_i = \frac{\sum_{n=1}^{N} y_{ni}}{N} \]

Correlation analysis

Step3: Dataset division, Model training

0.1C, 0.2C, 0.3C, 0.5C charging batteries dataset

Multi-feature, SOH

First 50% cycles as training dataset

Input (features)

CNN-BiLSTM-MHA

Training CNN-BiLSTM-MHA model

Testing dataset

Multi-feature

True SOH

Trained CNN-BiLSTM-MHA model

Estimated SOH

Estimation error

Step4: SOH estimation, Performance analysis

1MAE

\[ \text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \]

RMSE

\[ \text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \]

\[ R^2 \]

\[ R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \]

Training Time (s)

Based on Figure 13 and Table 3, it can be observed that due to conducting comprehensive IEA, combining the IE curve features and thermal features enables the deep learning model to establish a mapping relationship between features and the battery SOH effectively. The CNN-BiLSTM-MHA model constructed in this paper leverages the excellent ability of CNNs to extract various key trends in the sequence. The abstract features extracted by the CNN make the combination of IE curve features and thermal features more reasonable and alleviate the computational burden of subsequent BiLSTM. Meanwhile, BiLSTM performs well in learning the correlation between sequences, and the utilization of MHA to weight the hidden states of BiLSTM further enhances its stability. The experimental results demonstrate that the proposed method achieves accurate and fast SOH estimation results across four different charge rates in the cycle aging experiment data, particularly for the true values of the SOH at charging rates of 0.1C and 0.2C, as these two charging rates result in longer charging times for each cycle of the battery, and the battery charge–discharge

![Figure 12](image-url)

Figure 12. (a) The SOH estimation results for charging at 0.1C; (b) the SOH estimation results for charging at 0.2C; (c) the SOH estimation results for charging at 0.3C; (d) the SOH estimation results for charging at 0.5C.

Table 3. SOH estimation performance.

<table>
<thead>
<tr>
<th>Charging Rate</th>
<th>MAE (%)</th>
<th>RMSE (%)</th>
<th>R² (%)</th>
<th>Training Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1C</td>
<td>0.1982</td>
<td>0.2921</td>
<td>92.62</td>
<td>10.47</td>
</tr>
<tr>
<td>0.2C</td>
<td>0.1873</td>
<td>0.2997</td>
<td>95.27</td>
<td>14.18</td>
</tr>
<tr>
<td>0.3C</td>
<td>0.1652</td>
<td>0.2130</td>
<td>97.69</td>
<td>13.81</td>
</tr>
<tr>
<td>0.5C</td>
<td>0.1968</td>
<td>0.2625</td>
<td>99.32</td>
<td>15.36</td>
</tr>
</tbody>
</table>
cycle aging experiment was conducted at room temperature. Considering the temperature difference between day and night, the average temperature difference between each cycle at these two charging rates is significant. Therefore, the capacity regeneration and degradation phenomena are more obvious at these two charging rates. Due to the consideration of thermal features, the proposed method can accurately estimate the SOH.

To further demonstrate the estimation performance of the proposed method, an evaluation is conducted based on the four aforementioned metrics: the MAE, RMSE, $R^2$, and training time.

From the perspective of estimation accuracy, across the four different charge rates, the MAE of the estimation results is within 0.2%, and the RMSE is within 0.3%. For the 0.2C, 0.3C, and 0.5C charge rate data, the $R^2$ are all above 95%. For the 0.1C charge rate data, due to the pronounced nonlinearity in the variation in its true SOH values, $R^2$ is slightly lower compared to the estimation results of other charge rates. However, both the MAE and RMSE reach considerable standards, and $R^2$ still reaches 92.62%, indicating that the proposed method can provide an accurate estimation of the battery SOH.
From the perspective of training time, with a significant amount of data in the battery charge–discharge cycle process, the training time varies due to the different numbers of charge–discharge cycles and the complexity of the tasks. The maximum training time is only 15.36 s.

4.3. Experimental Comparisons

In order to validate the superiority of the CNN-BiLSTM-MHA model proposed in this paper, comparative experiments are conducted with the CNN-BiLSTM, BiLSTM, BiGRU, and SVR models. Using the aforementioned laboratory data with different charge rates, the comparative experimental results are shown in Figure 13a–d.

The MAE, RMSE, and $R^2$ comparison of the estimation results for different charging rate across various models are shown in Tables 4–7.

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE (%)</th>
<th>RMSE (%)</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN-BiLSTM-MHA</td>
<td>0.1982</td>
<td>0.2921</td>
<td>92.62</td>
</tr>
<tr>
<td>CNN-BiLSTM</td>
<td>0.2301</td>
<td>0.3551</td>
<td>89.09</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>0.2827</td>
<td>0.4471</td>
<td>82.70</td>
</tr>
<tr>
<td>BiGRU</td>
<td>0.3429</td>
<td>0.4816</td>
<td>79.93</td>
</tr>
<tr>
<td>SVR</td>
<td>0.3754</td>
<td>0.5354</td>
<td>75.00</td>
</tr>
</tbody>
</table>

The comparison of the training time between the proposed model and the CNN-BiLSTM, BiLSTM, and BiGRU models at different charging rates is shown in Table 8.
Table 8. Training time comparison.

<table>
<thead>
<tr>
<th>Charging Rate</th>
<th>Model</th>
<th>Training Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1C</td>
<td>CNN-BiLSTM-MHA</td>
<td>10.47</td>
</tr>
<tr>
<td></td>
<td>CNN-BiLSTM</td>
<td>10.04</td>
</tr>
<tr>
<td></td>
<td>BiLSTM</td>
<td>15.66</td>
</tr>
<tr>
<td></td>
<td>BiGRU</td>
<td>15.35</td>
</tr>
<tr>
<td>0.2C</td>
<td>CNN-BiLSTM-MHA</td>
<td>14.18</td>
</tr>
<tr>
<td></td>
<td>CNN-BiLSTM</td>
<td>13.67</td>
</tr>
<tr>
<td></td>
<td>BiLSTM</td>
<td>22.76</td>
</tr>
<tr>
<td></td>
<td>BiGRU</td>
<td>22.42</td>
</tr>
<tr>
<td>0.3C</td>
<td>CNN-BiLSTM-MHA</td>
<td>13.81</td>
</tr>
<tr>
<td></td>
<td>CNN-BiLSTM</td>
<td>12.96</td>
</tr>
<tr>
<td></td>
<td>BiLSTM</td>
<td>22.23</td>
</tr>
<tr>
<td></td>
<td>BiGRU</td>
<td>21.65</td>
</tr>
<tr>
<td>0.5C</td>
<td>CNN-BiLSTM-MHA</td>
<td>15.36</td>
</tr>
<tr>
<td></td>
<td>CNN-BiLSTM</td>
<td>15.02</td>
</tr>
<tr>
<td></td>
<td>BiLSTM</td>
<td>24.06</td>
</tr>
<tr>
<td></td>
<td>BiGRU</td>
<td>23.47</td>
</tr>
</tbody>
</table>

From Tables 4–8, it can be observed that using IE curve features combined with thermal features, the proposed model outperforms the other four models in terms of both the MAE and RMSE (which are lower) and $R^2$ (which is higher). Additionally, the training time of the proposed CNN-BiLSTM-MHA model is significantly advantageous compared to the BiLSTM model. This indicates that the CNN-BiLSTM-MHA model has better SOH estimation accuracy and speed across different charge rates compared to other models. Furthermore, it is noted that BiGRU, having a simpler gating structure and lower computational complexity compared to BiLSTM, exhibits slightly shorter training times. However, its estimation accuracy is inferior to BiLSTM. Considering that the CNN reduces the computational burden of subsequent models, BiLSTM is selected as part of the combined model in this paper.

Comparing the estimation results of the CNN-BiLSTM model with the BiLSTM model reveals that the CNN effectively extracts key trends in the original IE curve features and thermal features, transforming them into abstract features by effectively combining the original features. This allows the subsequent BiLSTM to easily learn forward and backward dependencies between sequences. Compared to the BiLSTM model, the CNN-BiLSTM model achieves an average reduction of approximately 22% in the MAE and RMSE and an average reduction of approximately 39% in training time. This demonstrates the effectiveness of the CNN in handling original features.

In this paper, MHA is applied to the hidden states of BiLSTM, effectively alleviating the gradient vanishing and exploding problems in BiLSTM, ensuring that the hidden states are updated correctly. Comparing the estimation results of the CNN-BiLSTM-MHA model with the CNN-BiLSTM model, it is observed that while the CNN-BiLSTM-MHA model experiences an average increase of approximately 4% in training time, the MAE and RMSE are reduced by an average of approximately 18% and 17%, respectively. This indicates that combining MHA with BiLSTM can achieve significant improvements in estimation accuracy with minimal computational burden.

In conclusion, the proposed method and model can provide a rapid and accurate estimation of the SOH for data from different charging rate.

5. Conclusions

A lithium-ion battery SOH estimation method based on multi-feature and CNN-BiLSTM-MHA is proposed in this paper, significantly enhancing the accuracy and speed of SOH estimation.
In this paper, the IE curves are obtained from the constant current charging phase data of the battery. A comprehensive and detailed analysis of the IE curves is conducted, and IE curve features including IE, peak value, average value, and standard deviation are extracted. These IE curve features provide comprehensive and accurate information about the battery SOH. The Pearson correlation coefficient method is also used to validate the high correlation between IE curve features and the battery SOH. In addition to the IE curve features, thermal features of the battery are also incorporated. Therefore, the deep learning model can learn the relationship between the internal chemical reaction changes during the charge–discharge cycle of the battery and SOH, leading to an accurate SOH estimation model.

Based on comprehensive IEA, the extraction of IE curve features combined with thermal features, an outstanding CNN-BiLSTM-MHA model is proposed in this paper. This model combines the advantages of CNNs in extracting important abstract features and BiLSTM in learning long-term sequential dependencies. Therefore, the model achieves both speed and accuracy in battery SOH estimation. Additionally, the use of MHA enhances the stability of BiLSTM, further improving the generalization and robustness of the model. The model undergoes rigorous experimental validation on datasets with four different charging currents. Comparative experiments with the CNN-BiLSTM, BiLSTM, BiGRU, and SVR models further confirm the excellent estimation accuracy and speed of the CNN-BiLSTM-MHA model. Even in the presence of significant nonlinearity in the battery SOH decay process, the MAE is 0.1982%, 0.1873%, 0.1652%, and 0.1968%, and the RMSE is 0.2921%, 0.2997%, 0.2130%, and 0.2625%, respectively. The average $R^2$ exceeds 96%. Compared to the BiLSTM model, the CNN-BiLSTM-MHA model achieves an average reduction of approximately 36% in the MAE and RMSE and an average reduction of approximately 36% in training time.

In addition, the results of this study provide valuable references for future research work. By revealing the relationship between voltage, energy, and temperature data and the battery SOH, it lays the foundation for developing more accurate and efficient SOH estimation models and stimulates researchers to explore methods for using measurement data for accurate and fast SOH estimation. In the future, combining more types of sensor data and adopting more advanced deep learning algorithms may further improve the accuracy and real-time performance of SOH estimation. Optimizing the data collection and processing process to achieve SOH estimation using fragmented data will be an important development direction. In summary, this paper not only provides an effective solution for the accurate and rapid estimation of battery SOH but also points out the direction for future research and application, with important academic and practical value. We will optimize data processing and strengthen the establishment of SOH estimation models in the future, exploring the use of fragmented data to achieve the accurate and fast estimation of the SOH.

6. Limitations of Study

Although this study employed a data-driven approach to estimate the battery SOH by measuring voltage, energy, and temperature data, there are still some limitations. Firstly, the source and diversity of the dataset are key factors. Although this study used four datasets with different charging currents, the data used mainly came from specific types and models of batteries, which may result in the unknown performance of the estimated model when faced with other types of batteries, affecting the universality of the results. Secondly, the uncertainty and errors in the data measurement process may also have an impact on the research results. The measurement of voltage, energy, and temperature is influenced by various factors such as sensor accuracy, environmental conditions, and operating specifications, which may lead to noise and bias in the measurement data, thereby affecting the accuracy of SOH estimation. We have taken corresponding measures to avoid the influence of environmental conditions and operating specifications as much as possible. Finally, the time range and experimental conditions of this study are relatively limited, and it has not
fully covered the long-term use and various complex working conditions of batteries in practical applications. Therefore, future research needs to further validate and enhance the robustness and accuracy of SOH estimation models by introducing more diverse datasets, improving measurement techniques, and conducting more comprehensive experiments.

**Author Contributions:** Conceptualization, C.Z. and Y.Z.; methodology, C.Z.; software, Y.Z.; validation, Y.Z.; investigation, X.Z.; data curation, Z.Z.; writing—original draft preparation, Y.Z.; writing—review and editing, C.Z. and Y.Z.; visualization, Z.Z.; project administration, C.Z.; funding acquisition, C.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Major Project of Basic Science (Natural Science) Research in Colleges and Universities of Jiangsu Province, Grant No. 23KJA480002; “Qinglan Project” for universities in Jiangsu Province; the Scientific Research Foundation for High-level Personnel in Jinling Institute of Technology under Grant No. jit-rcyj-202202.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The raw data supporting the conclusions of this article will be made available by the corresponding author upon reasonable request.

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

**Abbreviations**
The following abbreviations are used in this manuscript:

- **SOH** state of health
- **CNN** Convolutional Neural Network
- **BiLSTM** Bidirectional Long Short-Term Memory
- **MHA** Multi-head Attention
- **IEA** incremental energy analysis
- **IE** incremental energy
- **MAE** Mean Absolute Error
- **RMSE** Root-Mean-Square Error
- **$R^2$** Coefficient of Determination
- **GPR** Gaussian Process Regression
- **SVM** Support Vector Machine
- **LSTM** Long Short-Term Memory
- **GRU** Gated Recurrent Unit
- **ILM** Integrated Learning Model
- **MCL** Mixture Correntropy Loss
- **ELM** Extreme Learning Machine
- **IC** incremental capacity
- **CC-CV** constant current–constant voltage
- **ICA** incremental capacity analysis
- **Q-V** capacity–voltage
- **E-V** energy–voltage
- **1D CNN** One-Dimensional Convolutional Neural Network
- **RNN** recurrent neural network
- **BiGRU** Bidirectional Gated Recurrent Unit
- **SVR** Support Vector Regression

**References**


18. Tran, M.-K.; Mathew, M.; Janhunen, S.; Panchal, S.; Raahemifar, K.; Fraser, R.; Fowler, M. A comprehensive equivalent circuit model for lithium-ion batteries, incorporating the effects of state of health, state of charge, and temperature on model parameters. *J. Energy Storage* **2021**, *43*, 103252. [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.