



Enhancing lithium-ion battery lifespan early prediction using a multi-branch vision transformer model

Wanjie Zhao ^a, Wei Ding ^a, Shujing Zhang ^b, Zhen Zhang ^{a,*}

^a School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, 200444, China

^b State Grid Intelligence Technology Co., Ltd, Shandong, 250000, China

ARTICLE INFO

Handling editor: X Ou

ABSTRACT

Accurately predicting the lifespan of lithium-ion batteries is crucial for effective battery management systems, particularly for ensuring the safe operation and proactive maintenance of electric vehicles. However, existing methods encounter challenges due to the early weak capacity aging trend in batteries. This study proposes a novel approach using a multibranch vision transformer model to address early stage lifespan prediction issues. The proposed model leverages a multi-input data structure by considering various battery parameters during the charging and discharging phases. By employing distinct branch structures for different inputs, the model can separately extract features from individual input variables. Each vision transformer network was meticulously designed to extract high-dimensional global hidden features from the inputs and integrated to predict the battery lifespan. Comparative analyses against advanced baseline models and existing methods consistently demonstrate the superior performance and robustness of the proposed model. Compared with traditional vision transformer, the proposed model demonstrated a notable reduction in root mean square errors of 13.17 % and 6.32 % on the two public datasets, respectively, indicating the efficacy and reliability of our approach for accurately predicting the lifespan of LIBs during the early stages of capacity decline.

1. Introduction

Lithium-ion batteries stand as the primary power sources for electric vehicles, boasting high energy density, eco-friendliness, and extended longevity. Within the critical framework of a battery management system that is pivotal for electric vehicles, the lifespan of these batteries holds crucial significance. It serves as a key metric for energy management and the allocation of kinetic energy [1,2].

The lifespan of a lithium-ion battery is defined by the number of charge and discharge cycles it can endure while maintaining its initial performance under specific usage conditions [3]. Normally, when the available capacity of the battery drops to 80 % of its initial capacity or the internal resistance increases to twice its original capacity, the battery can be considered to have reached the end of its lifespan [4]. Early prediction of battery lifespan refers to predicting the lifespan of a battery in the early stage, usually within the first 100 cycles. This is beneficial for the proactive management of batteries and helps optimize their usage [5]. Although both early lifespan prediction and remaining useful life (RUL) prediction are integral components of battery management

systems (BMS), they have fundamental differences. RUL prediction involves the continuous assessment of a battery's ability to maintain an acceptable capacity level before experiencing significant degradation at any given point during its operational lifespan, based on its current health and performance. Conversely, early prediction of battery lifespan is centered around estimating the total expected life of a battery when it is new, considering a multitude of factors influencing its longevity. The precision of early battery lifespan prediction is paramount, as it can substantially contribute to enhanced resource management, cost-effectiveness, heightened safety, sustainability, and overall efficiency across diverse applications and industries reliant on battery technology.

The existing research literature delineates two broad methods for predicting battery lifespans early: model-based and data-driven methods [6]. Model-based methods rely on a comprehensive understanding of the physical and chemical properties inherent in batteries. These techniques involve creating physical models, such as electrochemical models [7] and equivalent circuit models [8]. Alternatively, they utilize estimation models based on particle filters [9] and Kalman filters [10] to forecast

* Corresponding author.

E-mail address: zhangzhen_ta@shu.edu.cn (Z. Zhang).

the lifespan of lithium-ion batteries. While model-based methods generally offer higher accuracy in predicting battery lifespans, their precision is intricately tied to the quality of the model. This dependency on model quality presents significant limitations in domain knowledge, which hampers the widespread applicability of model-based methods.

In contrast, data-driven methods circumvent the limitations imposed by prior battery knowledge and have become extensively employed in predicting battery lifespans. Chen et al. [11] introduced a hybrid prediction model that amalgamates the broad learning system with a relevance vector machine. Similarly, Fei et al. [12] presented a comprehensive framework for battery lifespan prediction utilizing machine learning, in which the support vector machine demonstrated superior performance. In a different approach, Zhang et al. [13] devised a hybrid lifespan prediction model that integrates random forest, artificial bee colony, and a general neural network. Meanwhile, Yang et al. [14] proposed a gradient boosting regression tree model to forecast battery lifespan. In another vein, Xu et al. [15] advocated for a deep-learning-based stacked denoising autoencoder approach to predict battery lifespan. However, these traditional machine learning methods typically hinge on intricate feature engineering, which significantly constrains the efficiency of model prediction.

In recent years, with advancements in computer technology, significant strides have been made in data-driven approaches for predicting battery lifespans. Deep learning, which autonomously extracts features from original data, has become increasingly prevalent. Yang et al. [16] introduced a hybrid convolutional neural network (CNN) that fused three-dimensional and two-dimensional CNNs, leveraging the terminal voltage, current, and temperature curves during battery charging to successfully predict early battery life. Tang et al. [17] proposed a series model that incorporated convolutional neural networks and long short-term memory networks (LSTM). Their model accurately predicted the battery lifespan based on differences in the capacity-voltage curves between battery cycles. Likewise, Zhang et al. [18] developed a neural network incorporating an attention mechanism and bidirectional long short-term memory (BiLSTM), focusing solely on early battery data to achieve accurate lifetime predictions. Wang et al. [19] introduced an enhanced antinoise adaptive LSTM, showcasing robust feature extraction and optimal parameter characterization to precisely predict the remaining useful life. Given the ongoing challenge of forecasting the lifespan of lithium-ion batteries amid complex operational scenarios, novel deep learning models capable of extracting thorough battery lifespan insights from intricate raw data are imperative [20–22].

In the realm of sequential data modeling, the transformer model with a self-attention mechanism stands as one of the most widely employed neural network architectures, originally introduced for machine translation tasks within natural language processing. The transformative impact of the transformer architecture has spurred the emergence of novel approaches based on transformers across various domains, including computer vision and time series analysis. Over recent years, transformer-based models have demonstrated remarkable efficacy across diverse applications, including predicting the state of health (SOH) and estimating the state of charge (SOC) of lithium-ion batteries [23–30]. Gomez et al. [31] proposed an enhanced temporal fusion transformer model that utilized Bayesian optimization, employing a tree-structured Parzen estimator for hyperparameter tuning. This approach aimed to augment the accuracy of predicting the SOH and RUL of lithium-ion batteries. Similarly, Zhang et al. [32] introduced an integrated model that combined stacked denoising autoencoders with a transformer framework to improve the precision of RUL prediction for lithium-ion batteries subjected to random charging and discharging conditions. Furthermore, Fei et al. [27] presented a deep learning-based framework featuring a bilateral branched Vision Transformer (ViT), designed to learn multi-timescale high-level latent features concurrently for online estimation of the SOH of lithium-ion batteries. These investigations employed complex deep learning architectures to extract high-dimensional features from battery data, thereby enhancing the

performance of BMS. However, despite these advancements, there remains a notable dearth of research focusing on early prediction of lithium-ion battery lifespan.

To tackle these challenges, this study introduces a multi-branch vision transformer model named MB-ViT. This model autonomously extracts feature information pertaining to battery lifespan from various battery parameters and subsequently integrates this information to directly predict the battery lifespan.

The main contributions of this study can be summarized as follows:

- (1) Introduction of MB-ViT: A novel deep learning model specifically tailored for early prediction of battery lifespan, MB-ViT is grounded in the vision transformer architecture, leveraging its robust functionality for effective feature extraction via the self-attention mechanism. The multi-branch structure of the model enables simultaneous processing of various input variables derived from voltage, current, temperature, and time parameters during the initial phases of battery charging and discharging.
- (2) Achieving relatively accurate early lifespan prediction: Using only 100 cycles of data in the early stages of battery charging and discharging, the study attained relatively accurate early lifespan prediction. The voltage, current, and temperature change curves of the battery from the first 10 to 110 cycles were respectively extracted, and an input data structure matching the model was constructed to realize early prediction of the battery lifespan.
- (3) Validation of the proposed model's superiority: Comparative analysis against baseline models and existing research methodologies using two publicly available datasets demonstrated the superior performance and predictive stability of the proposed model in battery lifespan estimation.

The subsequent sections of this paper are organized as follows: Section 2 introduces the dataset employed in this study and encompasses data analysis and the generation of input data. Section 3 delineates the specific structure of the proposed model. Section 4 details the experimental setup, including input data preprocessing and implementation specifications. Section 5 presents the experimental results, providing outcomes and an in-depth discussion. Finally, Section 6 encapsulates the summary of this work.

2. Experimental data

In this section, the description of the datasets, source of the original data, and preprocessing of the battery data are explained in detail.

2.1. Dataset

The two datasets utilized in this study were sourced from MIT [33, 34], with their fundamental parameters outlined in Table 1. All batteries underwent testing within a temperature-controlled incubator set at 30 °C. Batteries within BD-1 adhered to a two-stage charging strategy labeled as “C1 (Q1)-C2.” This entails initially charging to the state of charge Q1 at a constant current C1, followed by reaching 80 % state of charge with a constant current C2. On the other hand, batteries in BD-2 followed a charging strategy denoted as “CC1-CC2-CC3-CC4,” involving

Table 1
The experimental parameters of two datasets.

Dataset	BD-1	BD-2
Number of batteries	124	45
Type and Material	LFP/graphite	LFP/graphite
Nominal capacity	1.1 Ah	1.1 Ah
Nominal voltage	3.3 V	3.3 V
Upper and lower cut-off voltage	3.6 V/2.0 V	3.6 V/2.0 V
Charge strategy	C1 (Q1)-C2	CC1-CC2-CC3-CC4
Discharge strategy	4C	4C

four distinct constant currents (CC1, CC2, CC3, and CC4) with charging current switches occurring upon every 20 % increase in battery state of charge. Subsequently, a constant current of 1C was applied to fully charge the battery. During the discharge stage, all batteries in both BD-1 and BD-2 were discharged at a constant current of 4C. The capacity degradation curves and battery lifespan distributions for the two datasets are depicted in Fig. 1.

2.2. Data analysis

From Fig. 1 (a) and (c), it is evident that, as each battery progresses through cycles in the dataset, the trend of battery capacity degradation becomes increasingly pronounced. However, within the first 110 cycles, the degree of battery capacity degradation is minimal, making it exceedingly challenging to directly predict the battery lifespan using capacity change data from the early discharge stage.

Hence, we considered all voltage, current, and temperature changes during the charging and discharging processes in the early stages of the battery, as illustrated in Fig. 2. It can be observed that, during the early stages, there is significantly more information between different cycles for a battery. For instance, regarding the voltage during the battery discharging process shown in Fig. 2 (b) and the current during the battery charging process shown in Fig. 2 (d), the internal resistance of the battery increases with the number of cycles, leading to a decline in voltage and current during charging and discharging stages. Simultaneously, the trend of the temperature data in Fig. 2 (f) aligns with the fact that the battery reaches higher temperatures more easily and quickly as the number of charge and discharge cycles increases. Therefore, we selected the cycle change curves of voltage, current, and temperature over time during battery charging and discharging as inputs for the model.

2.3. Input data

To maximize the utilization of cycle data in the early stages of battery operation, we resampled the voltage, current, and temperature cycle curves during both charging and discharging stages at uniform time intervals. Subsequently, the resampled values of all cycles in the early stage were concatenated to form the final input data. For the i -th battery, the input data structures generated by the voltage, current, and temperature cycle curves can be defined by formulas 1–3.

$$V_i = \begin{bmatrix} V_i^j(t) \\ V_i^{j+1}(t) \\ \dots \\ V_i^{j+W}(t) \end{bmatrix} = \begin{bmatrix} V_i^{j,1}, V_i^{j,2}, \dots, V_i^{j,D} \\ V_i^{j+1,1}, V_i^{j+1,2}, \dots, V_i^{j+1,D} \\ \dots \\ V_i^{j+W,1}, V_i^{j+W,2}, \dots, V_i^{j+W,D} \end{bmatrix}_{W \times D} \quad (1)$$

$$I_i = \begin{bmatrix} I_i^j(t) \\ I_i^{j+1}(t) \\ \dots \\ I_i^{j+W}(t) \end{bmatrix} = \begin{bmatrix} I_i^{j,1}, I_i^{j,2}, \dots, I_i^{j,D} \\ I_i^{j+1,1}, I_i^{j+1,2}, \dots, I_i^{j+1,D} \\ \dots \\ I_i^{j+W,1}, I_i^{j+W,2}, \dots, I_i^{j+W,D} \end{bmatrix}_{W \times D} \quad (2)$$

$$T_i = \begin{bmatrix} T_i^j(t) \\ T_i^{j+1}(t) \\ \dots \\ T_i^{j+W}(t) \end{bmatrix} = \begin{bmatrix} T_i^{j,1}, T_i^{j,2}, \dots, T_i^{j,D} \\ T_i^{j+1,1}, T_i^{j+1,2}, \dots, T_i^{j+1,D} \\ \dots \\ T_i^{j+W,1}, T_i^{j+W,2}, \dots, T_i^{j+W,D} \end{bmatrix}_{W \times D} \quad (3)$$

where W represents the window size, indicating the number of cycles in

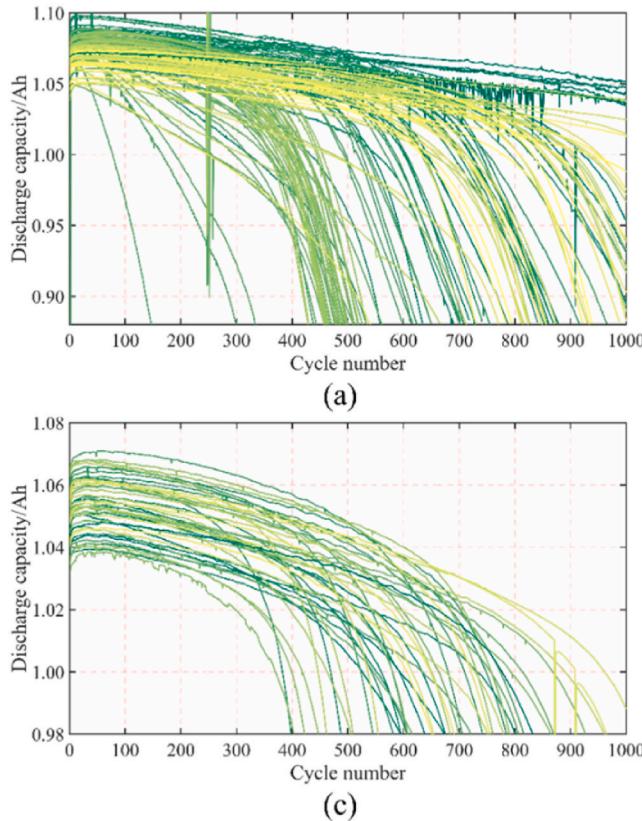


Fig. 1. The battery lifespans of two datasets. (a), (c) Discharge capacity curves of all batteries in BD-1 and BD-2. (b), (d) Histogram of battery lifespan distribution of BD-1 and BD-2.

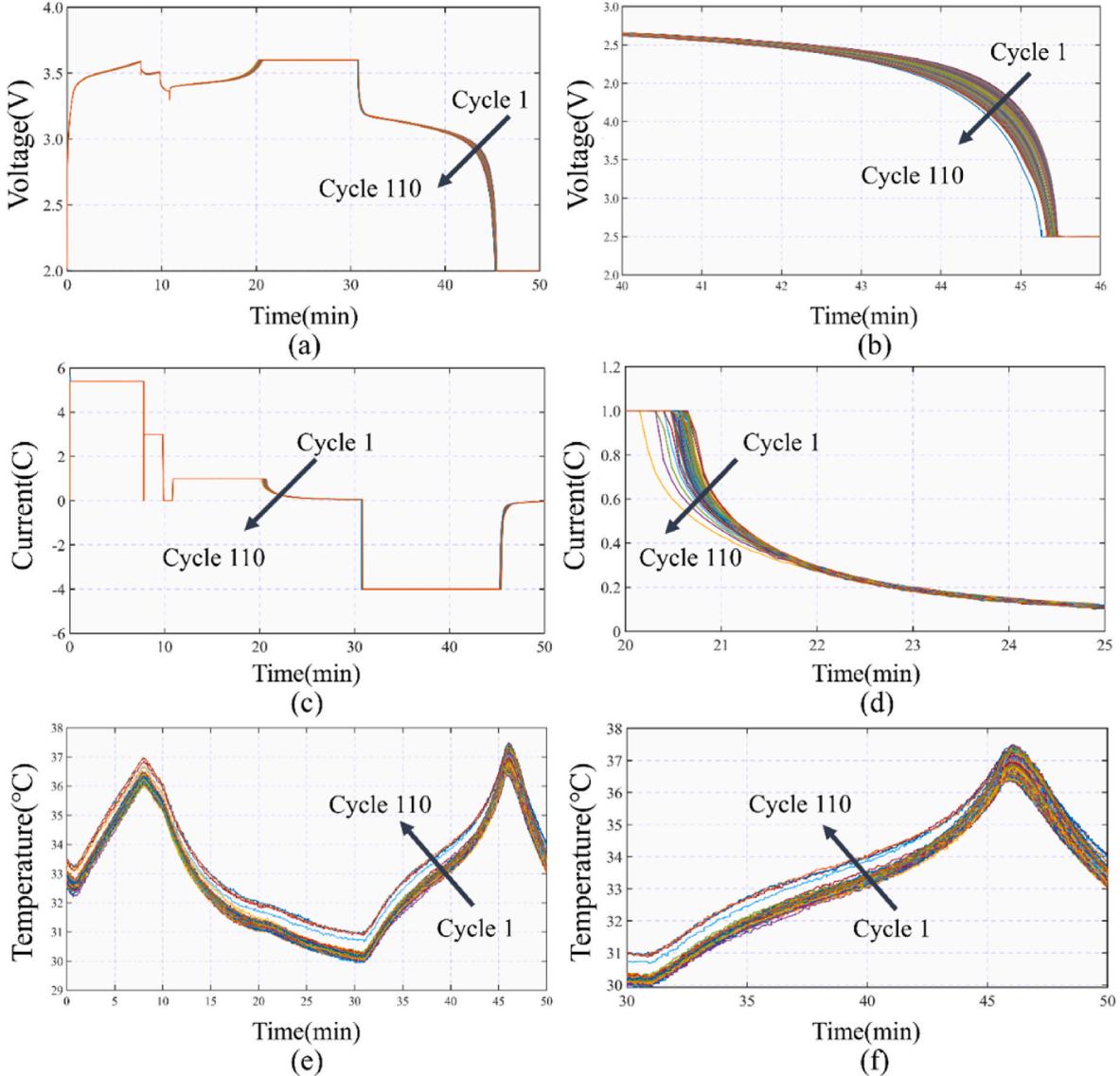


Fig. 2. Cycle-to-cycle evolution of voltage, current, and temperature in the first 110 cycles. (a), (c), and (e) are full curves, (b), (d), and (f) are partial enlarged curves corresponding to the left figure.

the initial battery stage. D denotes the number of samples, with j representing the selected initial cycle, as the aging trend of the first 10 cycles of the battery can be largely neglected, and sampling commences from the 11th cycle. $V_i^{j+W,D}$ is defined as the D voltage sampling value on the $j + W$ cycle of the battery. Furthermore, future experiments should take into account the discontinuous nature of battery cycle data collection, a common occurrence during actual use.

3. Method

This section presents the detailed structure of the proposed MB-ViT model. The early lifespan prediction method for lithium-ion batteries is illustrated in Fig. 3.

3.1. The overall architecture of the MB-ViT model

Due to the unique data embedding method and powerful global modeling ability of the ViT model [30,35], we devised an MB-ViT model, illustrated in Fig. 4, tailored to the data characteristics of lithium-ion batteries for accurate lifespan prediction. The MB-ViT model

employs three network branches to simultaneously learn potential high-dimensional feature information related to battery lifespan from three distinct battery input data streams. These learned features are then aggregated to form the final feature vector, which is subsequently forwarded to the lifespan predictor to derive the predicted battery lifespan.

The primary structures of the three branches in the MB-ViT model were identical. The battery input data V_i , I_i , and T_i , are initially partitioned into m patches of uniform size, as depicted in formulas 4–6. Following a series of experimental considerations, the size of each patch was determined to be 20×10 .

$$(V_1, V_2, V_3, \dots, V_m) \in V_i, \quad V_m \in R^{20 \times \frac{D}{10}} \quad (4)$$

$$(I_1, I_2, I_3, \dots, I_m) \in I_i, \quad I_m \in R^{20 \times \frac{D}{10}} \quad (5)$$

$$(T_1, T_2, T_3, \dots, T_m) \in T_i, \quad T_m \in R^{20 \times \frac{D}{10}} \quad (6)$$

Next, branch 1 is considered as an example to illustrate the embedding process of the input data of the MB-ViT model. For patch $V_m \in R^{20 \times \frac{D}{10}}$ in the input sequence, it will be flattened to vector $V'_m \in R^h$, $h = \frac{W}{20} \times \frac{D}{10}$

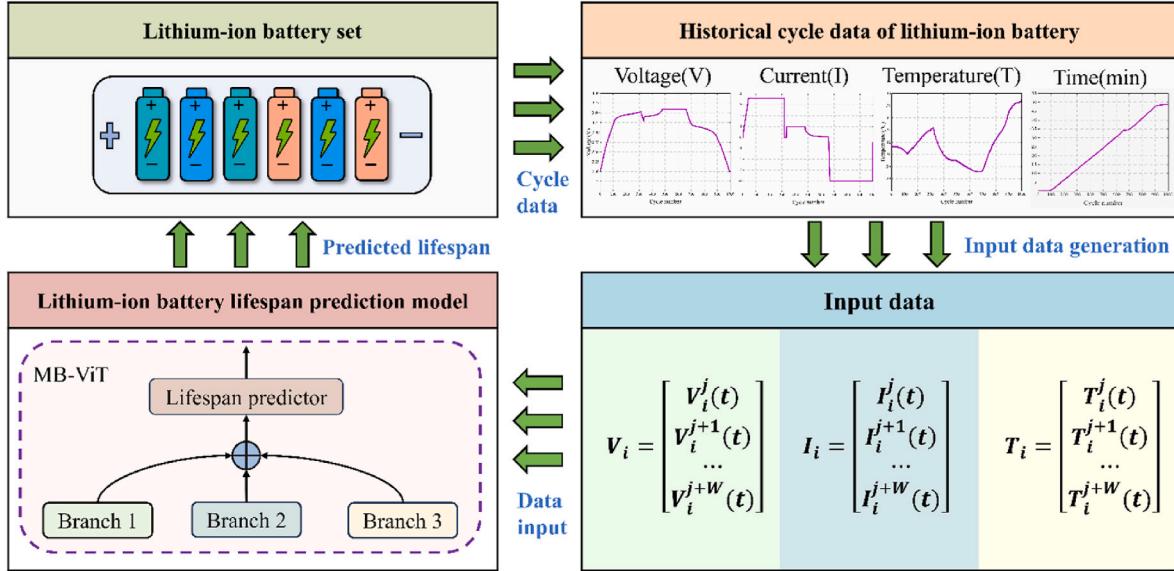


Fig. 3. The framework of lifespan early prediction method.

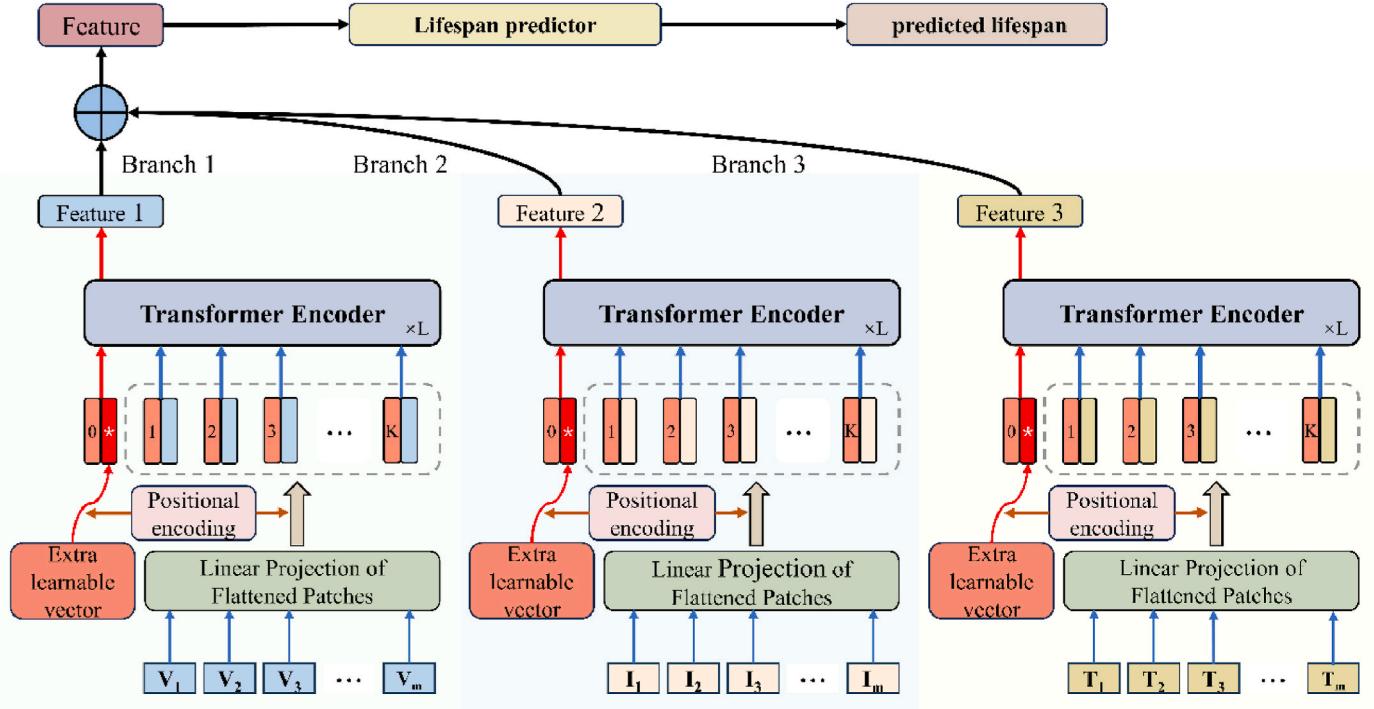


Fig. 4. Detailed structure of MB-ViT model.

first. It was then projected onto the embedded dimensions of the Transformer Encoder (TE) through a linear projection operation. The obtained vector V'_m is given by Equation (7):

$$V'_m = \text{Linear}(\text{flat}(V_m)), \quad V'_m \in R^\omega \quad (7)$$

where *flat* represents flattening operation, *Linear* represents linear projection operation, ω is defined as the embedded dimension of TE ($\omega = 128$).

Simultaneously, an additional learnable vector $\alpha \in R^\omega$ is incorporated into the embedding vector, and its output state at the TE layer is considered the final output feature vector. Subsequently, to preserve position information, one-dimensional position coding is added to the

entire embedding vector. The resulting final output sequence, x_v is given as

$$x_v = [\alpha, V'_1, V'_2, V'_3, \dots, V'_m] + \varphi_{pos}, \quad \varphi_{pos} \in R^{(m+1) \times \omega} \quad (8)$$

where φ_{pos} is a learnable parameter matrix.

Equivalent output sequences x_i and x_t . Finally, sequences x_v , x_i , and x_t are sent to the TE. After the TE layer extracts the feature information from the input sequence, the three feature vectors $F_v \in R^\omega$, $F_i \in R^\omega$, and $F_t \in R^\omega$ are finally obtained. The three feature vectors are then added and sent to the lifespan predictor to output the final predicted lifespan y_i of the battery. The entire process can be expressed using Equations (9) and (10):

$$F_v, F_i, F_t = \text{Transf}(x_v, x_i, x_t) \quad (9)$$

$$y_i = LP(F_v \oplus F_i \oplus F_t) \quad (10)$$

where *Transf* represents TE operation, *LP* represents lifespan predictor operation.

3.2. Transformer encoder

Fig. 5 illustrates the detailed structure of the TE, primarily comprising multihead self-attention, feed-forward blocks, layer normalization applied before each block, and a residual connection applied after each block, as depicted in **Fig. 5c**. At the core of TE lies multi-head self-attention, which employs multiple independent self-attention heads to capture global dependencies between elements in the input sequence and assigns different weights to different elements. Using the input sequence x_v as an example, the entire process can be expressed by formulas 11–13.

$$x_v^{\tau'} = MSA(LN(x_v^{\tau-1})) + x_v^{\tau-1}, \tau \in [1, L] \quad (11)$$

$$x_v^{\tau} = FF(LN(x_v^{\tau'})) + x_v^{\tau'}, \tau \in [1, L] \quad (12)$$

$$F_v = LN(x_v^{r,0}) \quad (13)$$

where *LN* denotes the layer normalization operation described in Equation (14), while *FF* represents a feed-forward operation comprising two linear layers with 256 hidden neurons, as specified in Equation (15). L denotes the number of TE layers ($L = 4$). *MSA* signifies a multihead self-attention operation, as detailed in Equation (16), performing h self-attention operations in parallel and ultimately concatenating the output of each head. The final attention weight b_1 is obtained through linear transformation, as depicted in **Fig. 5b**. In the self-attention layer, as illustrated in **Fig. 5a**, the input sequence is initially transformed into three vectors: query, key, and value. Subsequently, the dot product $\rho_{r,r}$ of the query and key vector is computed, followed by the application of the *softmax* function to derive attention weights $Att_{r,r}$. Finally, the value vectors are weighted by the attention weights to yield the final output, as described in formula 17.

$$LN(x) = \frac{\gamma(x - \mu)}{\sqrt{\sigma^2 + \epsilon}} + \beta \quad (14)$$

$$FF(x) = \delta(\omega_{ff}x + b) \quad (15)$$

$$MSA(\varepsilon) = [SA(\varepsilon_1); SA(\varepsilon_2); \dots; SA(\varepsilon_h)]W_{msa} \quad (16)$$

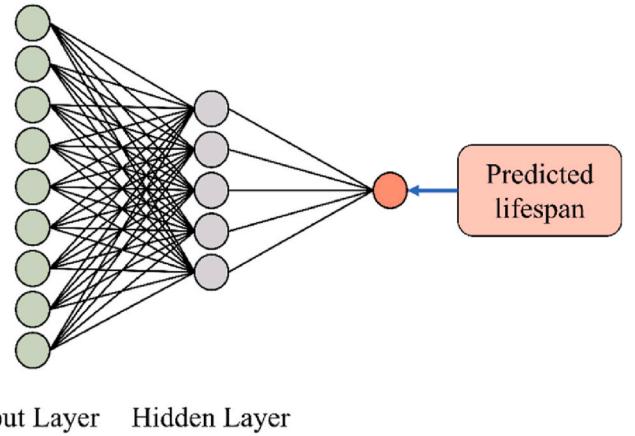
$$SA(\varepsilon) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, [Q, K, V] = \varepsilon W_{sa} \quad (17)$$

where h is the number of heads of multi-head self-attention ($h = 4$), W_{msa} and W_{sa} are trainable transformation matrix, Q , K , and V are the query, key, and value represent of the input sequence $\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_h]$, $\varepsilon_1 = [\alpha_1, \alpha_2, \dots, \alpha_r], r \in [1, m+1]$, d_k is the dimension of Q and K . Where γ and β are learnable parameters, δ represents the ReLU activation function, ω_{ff} and b are trainable parameter.

3.3. Lifespan predictor

By leveraging the potent nonlinear modeling capacity of the fully connected layer, which adeptly extracts lifespan information from the TE, a mapping relationship between feature information and battery lifespan is established. This process involves the implementation of two fully connected layers, as illustrated in **Fig. 6**, with a specific hidden layer comprising 64 neurons. Dropout technology is integrated to prevent overfitting and bolster the model's training efficiency.

Given the sensitivity of the mean square error (MSE) to prediction discrepancies, it serves as an optimal metric for gauging the disparity between the predicted and actual battery lifespans. Therefore, MSE is chosen as the loss function during the model training phase.



Input Layer Hidden Layer

Fig. 6. The structure of lifespan predictor.

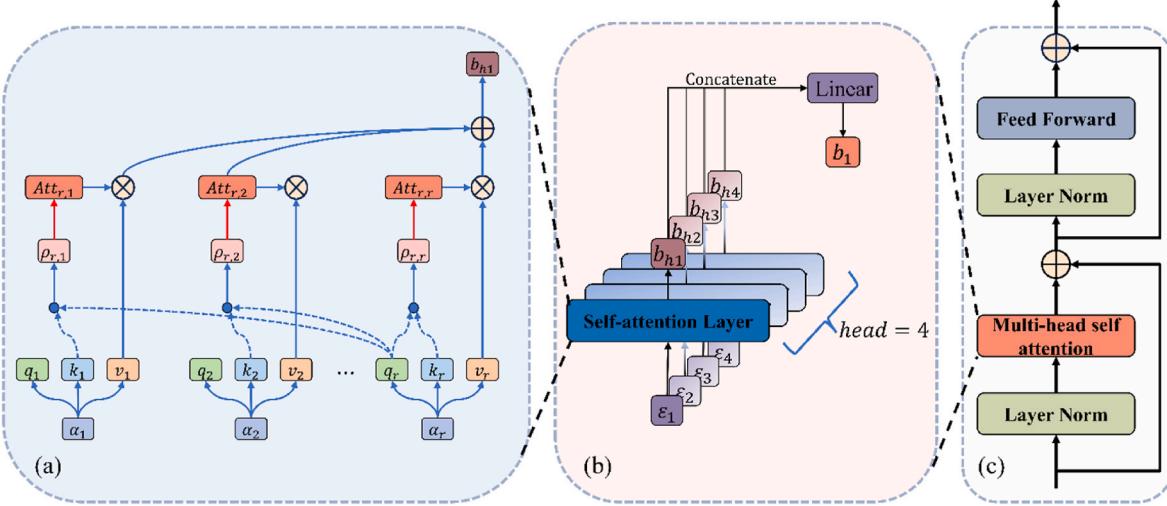


Fig. 5. The detailed structure of transformer encoder.

4. Experiments

The evaluation metrics, data preprocessing, and implementation details are described in this section.

4.1. Evaluation metrics

To more intuitively demonstrate the superiority of the model proposed in this study for early prediction of battery lifespan, we employed root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) to measure the prediction results of the model. Their mathematical formulas are as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y(i) - \hat{y}(i))^2} \quad (18)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y(i) - \hat{y}(i)| \quad (19)$$

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y(i) - \hat{y}(i)}{y(i)} \right| \quad (20)$$

where $y(i)$ and $\hat{y}(i)$ are the measured and predicted lifespans of the i -th sample, respectively, and N is the number of samples.

4.2. Data preprocessing

In our experiment, we partitioned the dataset into three distinct subsets: a training set, a validation set, and a test set. Based on the input data type, obvious abnormal samples in the dataset were excluded. To ensure the model's performance, we adopted a 3:1:1 ratio for randomly selecting batteries from the dataset for each training process. This careful selection method acknowledges the significant impact that the distribution of the training and test sets can have on the model's performance. Such an approach not only guarantees the stability and reliability of the model's predictive performance but also mitigates any unexpected anomalies resulting from specific distributions in the test sets.

Before commencing the training phase, we applied data normalization techniques to standardize all input data within the range of [0, 1]. This normalization fosters consistency among the input data, enhancing the efficiency and precision of model training. Additionally, to bolster the model's resilience and predictive capacity while preserving the dataset distribution, we employed a data augmentation technique. This involved introducing random values drawn from a normal distribution with mean values of zero and standard deviations of 1 % and 2 % into the original data, resulting in the generation of 20 additional sets of augmented data for each battery sample.

4.3. Implementation details

The proposed method was implemented on a computer system equipped with an Intel Core i7-8700K processor and 16 GB of RAM, utilizing PyTorch (v1.13.1). During the training process, the model's parameters were initialized via the Kaiming distribution technique. Additionally, the Adam optimizer was selected for its stability and faster convergence rate in optimizing network parameters. The initial learning rate was set at 0.001, with a reduction by a factor of 10 if the loss of the validation set did not decrease for 10 consecutive epochs during model training.

To prevent overfitting, the training epoch was capped at 300, and an early stop strategy was employed. Training halted either when the model met the early stop condition or when the number of training epochs reached 300. Each experiment was repeated ten times, and the results were averaged to enhance the reliability of the prediction

outcomes.

5. Results and discussion

In this section, a series of experimental results that verify the effectiveness of the proposed methods are presented, along with an expanded discussion.

5.1. Hyperparameters

The window size represents the early cycle range of the battery covered by the input data constructed in this study, directly impacting the computational complexity and prediction accuracy of the model. A larger window size entails more battery cycle information in the input data, while a smaller one contains less. To maximize the model's early lifespan prediction performance and achieve accurate predictions on the smallest window size possible, we conducted experiments using five groups of different window sizes.

[Fig. 7](#) illustrates the variation in prediction error of the MB-ViT model with respect to the window size. In [Fig. 7](#) (a), it is evident that when the window size is set to 100, the MB-ViT models exhibit the best prediction accuracy on the two datasets, reaching 92.14 cycles and 83.43 cycles, respectively. Similarly, from [Fig. 7](#) (b), it can be observed that when the window size is 100, the MB-ViT models displayed the best prediction stability on the two datasets, with MAPE errors concentrated within 5.90 % and 2.71 %, respectively. Therefore, a window size of 100 was selected to carry out the subsequent experiment.

A smaller window size results in insufficient battery lifespan information contained in the input data, which is the main reason for the large prediction error and increases the likelihood of encountering abnormal values, such as window sizes of 60, 80, and 140. However, while a larger window size allows the model to capture more information about the battery lifespan, it requires a stronger feature extraction ability to effectively extract information from the larger dataset. After comprehensive consideration, we set the window size to 100 in the subsequent experiments.

5.2. Influence of discretization degree of input data

The degree of input data discretization depends on the number of samples taken from the cycle curves during the battery's charging and discharging phases. In this study, the sampling quantity notably impacts the size of the model's input matrix. A greater number of samples improves the alignment between the input data and the battery cycle curve but also increases the complexity of the input matrices. To optimize computational costs while maintaining prediction accuracy, we designed five sets of varying sampling numbers for experimental purposes, aiming to utilize the fewest samples possible while ensuring predictive precision.

[Fig. 8](#) reflects the variation in prediction error of the MB-ViT model with the number of samples. As seen in [Fig. 8](#) (a), when the number of samples is small, the MB-ViT model achieves an RMSE error of approximately 92.14 cycles on BD-1. However, with an increase in the number of samples, the prediction accuracy does not further improve. Similarly, when the number of samples is 100, the MB-ViT model reaches the best prediction accuracy for BD-2, with an RMSE error of 83.43 cycles. Additionally, [Fig. 8](#) (b) illustrates that when the sampling number is 100, the prediction stability of the MB-ViT model is optimal on both datasets, with MAE errors concentrated within 14.59 cycles and 30.43 cycles, respectively. Furthermore, we analyzed the training time length of the MB-ViT model in each epoch under the same conditions when different sampling numbers are used, as shown in [Fig. 9](#). It can be concluded that in the process of increasing the number of samples from 100 to 300, the training time of the model on the two datasets increased by 101.27 % and 97.45 %, respectively. Therefore, in this study, we set the sampling number at 100.

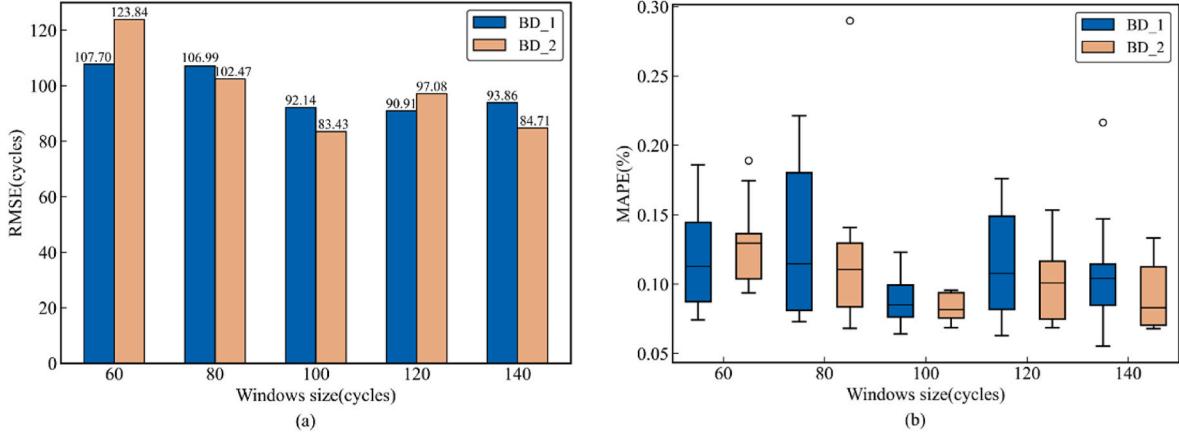


Fig. 7. Error comparison of the proposed model with different window sizes on BD-1 and BD-2. (a) is the RMSE errors, (b) is the distribution of MAPE errors.

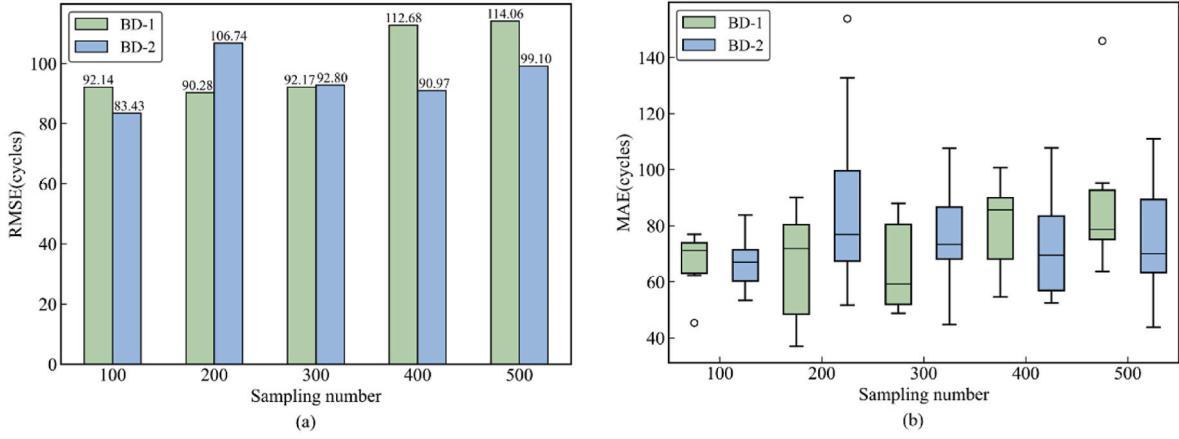


Fig. 8. Error comparison of the proposed model with different sampling number on BD-1 and BD-2. (a) is the RMSE errors, (b) is the distribution of MAE errors.

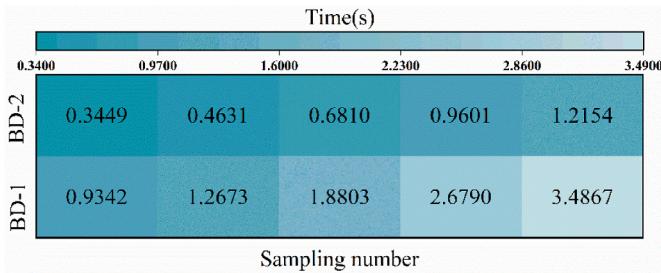


Fig. 9. The relationship between sampling number and training time of each epoch.

While a larger number of samples can enhance the ability of the input data to better depict the battery aging trend, it also heightens the risk of including abnormal values from the original battery data in the input. This scenario directly results in a phenomenon where the prediction accuracy of the model does not improve with an increase in the sampling number and is more prone to generating abnormal prediction results.

5.3. Lifespan early prediction with different deep learning models

To ascertain the superiority of the MB-ViT model in predicting the lifespan of lithium-ion batteries, we selected several typical baseline models along with existing research methods. These models comprised a pure CNN model, a series model combining CNN and LSTM, a series model combining CNN and Gated Recurrent Unit (GRU), and a pure ViT

model lacking a multi-branching structure. The structural parameters of the CNN are outlined in Table 2. Additionally, we employed two layers of LSTM and GRU to extract potential feature information, each with 32 hidden neurons. The structure of the pure ViT model mirrored the parameters of the MB-ViT model. Moreover, all models utilized the input data structure developed in this study and adhered to a consistent training strategy.

The prediction results of the MB-ViT and baseline models for the two datasets are summarized in Table 3. Notably, the MB-ViT model exhibited superior prediction performance across both datasets, as indicated by the three evaluation indexes. Specifically, concerning the RMSE error, the prediction accuracy of the MB-ViT model on BD-1 improved by 20.07 %, 14.94 %, and 16.93 % when compared with the combination of the typical CNN, LSTM, and GRU models, respectively, and by 13.17 % compared to ViT. Similarly, on BD-2, the prediction accuracy of the MB-ViT model improved by 18.06 %, 10.60 %, and 14.36 % when compared with the combination of the typical CNN, LSTM, and GRU models, respectively, and by 6.32 % compared to ViT. These findings underscore the superiority of the MB-ViT model and the

Table 2
The structure parameters of CNN.

Layer	Input size	Kernel size	Stride	Output
Conv-1	100×100	$8 \times 3 \times 3$	2×2	49×49
Conv-2	49×49	$16 \times 3 \times 3$	2×2	24×24
Conv-3	24×24	$32 \times 3 \times 3$	2×2	11×11
Conv-4	11×11	$32 \times 3 \times 3$	1×1	11×11

Table 3

The numerical comparison of the proposed MB-ViT and baseline models.

Model	BD-1				BD-2			
	RMSE (cycles)	MAPE (%)	MAE (cycles)	Test time (s)	RMSE (cycles)	MAPE (%)	MAE (cycles)	Test time (s)
CNN	115.27	12.31	81.84	0.0010	101.82	11.19	82.97	0.0010
CNN-LSTM	108.32	11.76	82.18	0.0020	93.32	9.26	73.16	0.0020
CNN-GRU	110.92	10.86	77.95	0.0050	97.42	10.52	80.17	0.0050
ViT	106.12	9.94	76.94	0.0150	89.06	8.97	66.90	0.0050
MB-ViT	92.14	8.89	67.66	0.1437	83.43	8.33	66.68	0.0484

efficacy of its multi-branch structure. Additionally, we conducted a visual analysis of the MAE errors of the MB-ViT and baseline models on the two datasets, as depicted in Fig. 10. It is evident that the prediction stability of the MB-ViT model across both datasets surpasses that of the baseline models.

Furthermore, we conducted a comparison between a typical model based on CNN and one based on ViT, revealing the superiority of the latter in terms of early lifespan prediction performance for lithium-ion batteries. While CNNs excel in handling local features due to their local receptive field, ViT models possess a unique ability to model input features globally. Given that information pertaining to battery lifespan is typically embedded across the entire charge-discharge cycle curve, the MB-ViT model demonstrates superior lifespan prediction performance. Moreover, as illustrated in Table 4, despite the proposed model featuring a larger number of trainable parameters compared to other baseline models, the disparity in time consumption during the actual model training process is negligible, being less than 1 min. This suggests that increasing model complexity to a limited extent is worthwhile in exchange for heightened accuracy and reliability in lifespan prediction.

Finally, we compared the prediction results of the MB-ViT model with the two better-performing baseline models on the test set, as depicted in Fig. 11. Across most samples, the MB-ViT model demonstrated a lower error rate compared to the other two baseline models. While there were fluctuations in the prediction errors of the three models for samples with medium lifespans (500 cycles < lifetime < 800 cycles), overall, they remained relatively stable. The MB-ViT model exhibited a slight advantage over the other two models for batteries with short lifespans (lifetime < 500 cycles), particularly evident in BD-1. However, for batteries with long lifespans (> 800 cycles), the MB-ViT model showcased better generalization ability than the other two models, as observed in battery samples #1–#4 in BD-1 and battery sample #1 in BD-2. This phenomenon can be attributed to the distribution of battery sample lifetimes in the dataset, characterized by a “few at both ends, more in the middle” pattern, as depicted in Fig. 11. Moreover, due to the fewer long- and short-lifetime battery samples in BD-2, the probability of large errors occurring is smaller, resulting in the

Table 4

Trainable parameters of the model proposed in this paper and other baseline models.

Model	CNN	CNN-LSTM	CNN-GRU	ViT	MB-ViT
Total_params	267553	159489	30273	626161	1697521

MB-ViT model's prediction performance on BD-2 being superior to that on BD-1.

5.4. Ablation study

This section examines the necessity of a multi-branch structure in the MB-ViT model and explores the impact of different branches on the early prediction performance of lithium-ion battery lifespans through an ablation study. The MB-ViT model will be compared with four versions: Branch 1 with only voltage data input, Branch 2 with only current data input, Branch 3 with only temperature data input, and a double-branch model combining Branches 1 and 2.

Table 5 summarizes the prediction results of the different branches on the two datasets. The multi-branch structure exhibited the highest prediction accuracy on both datasets. Additionally, the prediction performance of a single branch was weaker than that of the double-branch and multi-branch structures. Among the single-branch structures, Branch 3 showed the poorest prediction performance, indicating that the temperature data contained less information related to battery lifespan than the voltage and current data. However, it can be inferred that temperature data still contributes to improving the accuracy of early battery lifespan prediction by comparing the multi-branch structure with the double-branch structure combining Branches 1 and 2. Furthermore, we visually analyzed the MAE error distribution of all models on the two datasets, as depicted in Fig. 12. Clearly, the MB-ViT model demonstrated superior prediction stability due to its unique multi-branch structure. Simultaneously, Branch 3 exhibited the largest prediction fluctuation, consistent with the previous discussion. In summary, the predictive accuracy of battery lifespan is influenced by the

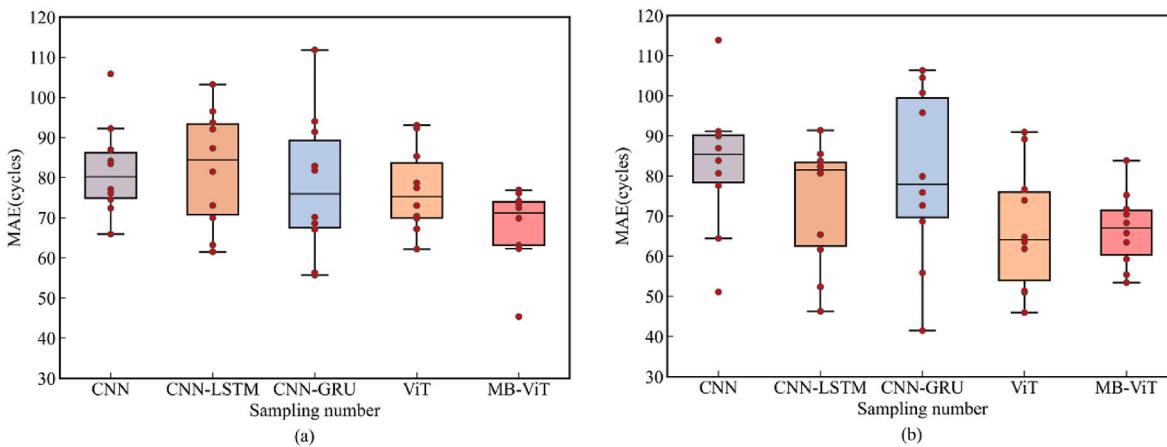


Fig. 10. The distribution of MAE errors. (a) is on BD-1, (b) is on BD-2.

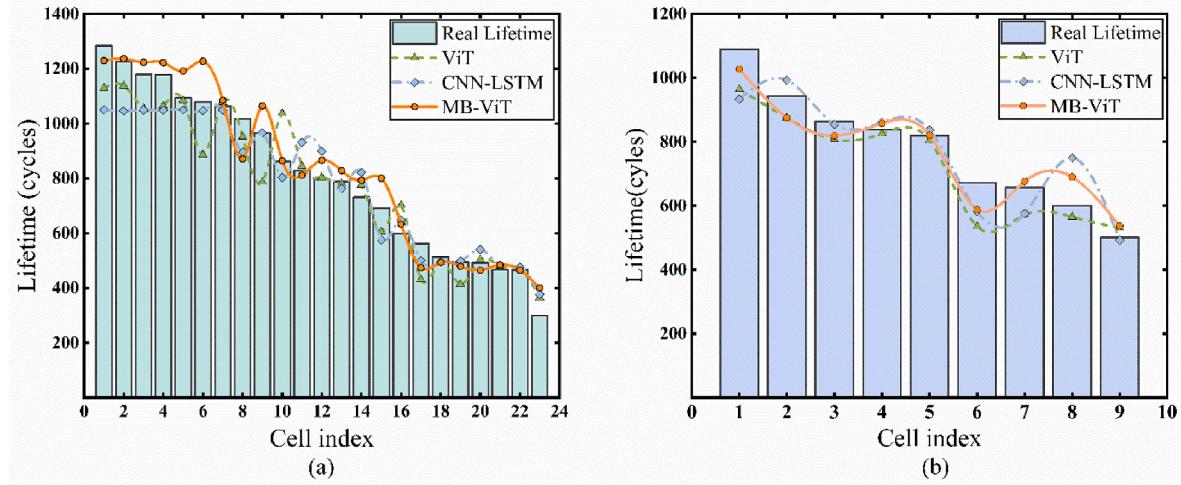


Fig. 11. Prediction results on the proposed MB-ViT model and two deep learning-based models. (a) is on BD-1, (b) is on BD-2.

Table 5

The numerical comparison under different branches.

	BD-1				BD-2			
	RMSE (cycles)	MAPE (%)	MAE (cycles)	Test time (s)	RMSE (cycles)	MAPE (%)	MAE (cycles)	Test time (s)
Branch 1	99.26	11.18	73.14	0.0091	107.54	11.45	83.38	0.0060
Branch 2	106.88	10.73	77.95	0.0080	91.10	9.55	74.86	0.0062
Branch 3	139.51	15.49	104.35	0.0090	113.83	12.37	90.51	0.0070
Branch 1 & Branch 2	95.75	10.69	69.31	0.0124	85.14	8.53	67.58	0.0103
This work	92.14	8.89	67.66	0.1437	83.43	8.33	66.68	0.0484

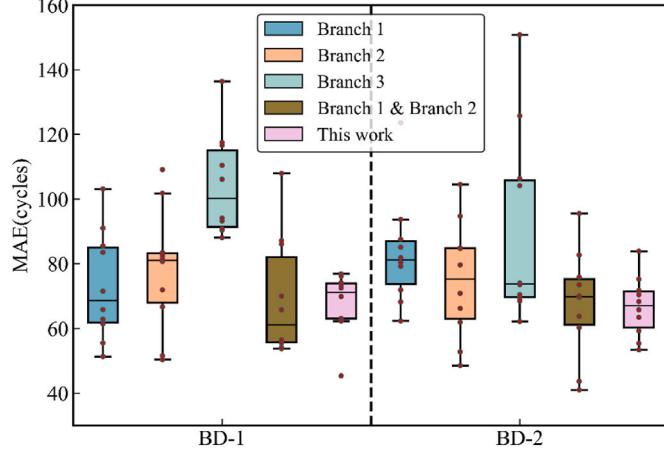


Fig. 12. The distribution of MAE errors under different branches on two datasets.

distinct contributions of various battery parameters, warranting further investigation into optimal integration methods for feature information across different analytical branches.

5.5. Comparison with other studies

Table 6 compares the performance difference in early prediction of battery lifespan between the model proposed in this study and existing advanced research methods. Similarly, the battery datasets utilized in these research methods were obtained from MIT. A comprehensive analysis of the three evaluation indicators demonstrates the superiority of the model proposed in this study over existing lithium-ion battery lifespan prediction models. Simultaneously, this study exclusively

Table 6
The comparison between this work with other advanced studies.

Reference	Model	Input data type	RMSE (cycles)	MAPE (%)	MAE (cycles)
[36]	CNN	$Q_{1-100}(V)$	–	17	106
[17]	PCLN	$\Delta Q_{10-100}(V)$	145.7	13.40	–
[37]	CNN-GPR	$[V(t), I(t), T(t)]_{1-100}$	112	8.2	–
[15]	SDAE	$\Delta Q_{10-300}(V)$	93.31	–	76.94
This work	MB-ViT	$V_{10-110}(t), I_{10-110}(t), T_{10-110}(t)$	92.14	8.89	67.66

selected historical cycle data of 100 cycles in the early stage of the battery as the model input and fully utilized it. A unique multi-branch input mode was adopted to accurately predict battery lifespan in the early stages.

5.6. Integration of deep learning model and BMS

The lithium-ion battery lifespan prediction module is a crucial component of the battery management system, providing essential information on battery lifespan, aiding in battery health assessment, and estimating remaining battery service time. Illustrated in Fig. 13, the integration of the deep learning model into a battery management system primarily involves data sampling, memory, and health management units. Once the data sampling unit completes the collection of cycle data such as voltage, current, and temperature in the early battery stage and stores them in the data memory unit, it triggers a lifespan prediction model in the health management unit to predict the battery lifespan. This process determines the battery health status and remaining cycle times. Subsequently, based on the prediction results, battery maintenance and scheduling can be executed to enhance battery utilization efficiency and service time.

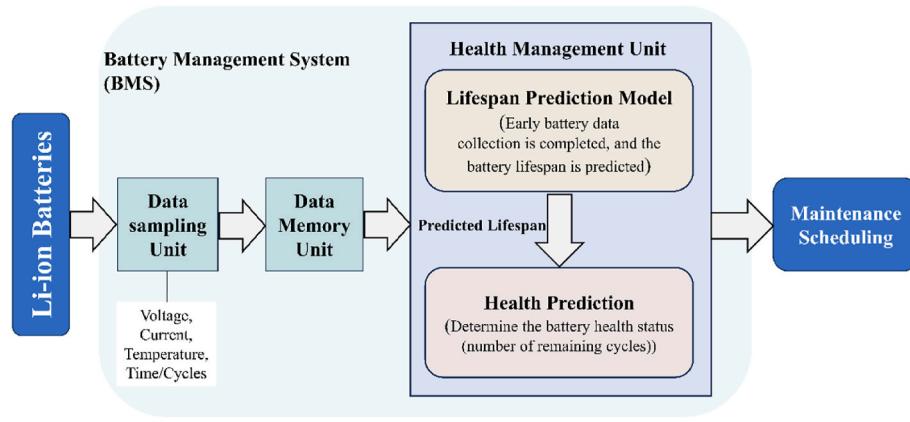


Fig. 13. Integration of battery lifespan prediction model and BMS.

6. Conclusion

To enhance the accuracy of lithium-ion battery lifespan prediction and provide theoretical support for the safe operation and preventive maintenance of electric vehicles, a data structure for multi-branch input was constructed based on various battery parameters in the early stages of battery charging and discharging. Consequently, we propose a novel deep learning-based multi-branch vision transformer model with a multi-head self-attention mechanism, named MB-ViT. This model aims to extract potentially high-dimensional hidden feature information from different types of input data and generate the final predicted lifespan using fused features.

The robustness and predictive performance of the model were verified based on two public datasets. Achieving RMSE prediction errors of 92.14 cycles and 83.43 cycles on the two datasets, respectively, using only historical data from 100 cycles in the early stage demonstrates the effectiveness of MB-ViT. Additionally, the ablation experiment shows that the multi-branch input structure combined with different battery parameters yields the best prediction effect. Furthermore, compared with a set of baseline models and existing research methods, our proposed method exhibits promising early lifespan prediction performance.

In future research, we aim to collect battery datasets with varying material properties and limited charge-discharge cycle data to further validate the model's performance. Additionally, we plan to employ advanced optimization algorithms to weigh and fuse multi-branch feature information more effectively. Ultimately, we intend to integrate the model with the BMS and extend its application to a wider range of tasks, including SOH estimation and RUL prediction for lithium-ion batteries.

CRediT authorship contribution statement

Wanjie Zhao: Writing – original draft, Validation, Software, Methodology. **Wei Ding:** Visualization, Validation, Investigation, Data curation. **Shujing Zhang:** Validation, Resources, Data curation. **Zhen Zhang:** Writing – review & editing, Supervision, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- [1] Wang S, Jin S, Deng D, Fernandez C. A critical review of online battery remaining useful lifetime prediction methods. *Front Mech Eng* 2021;7:719718. <https://doi.org/10.3389/fmec.2021.719718>.
- [2] Song K, Hu D, Tong Y, Yue X. Remaining life prediction of lithium-ion batteries based on health management: a review. *J Energy Storage* 2023;57:106193. <https://doi.org/10.1016/j.est.2022.106193>.
- [3] Atalay S, Sheikh M, Mariani A, Merla Y, Bower E, Widanage WD. Theory of battery ageing in a lithium-ion battery: capacity fade, nonlinear ageing and lifetime prediction. *J Power Sources* 2020;478:229026. <https://doi.org/10.1016/j.jpowsour.2020.229026>.
- [4] Tang A, Jiang Y, Nie Y, Yu Q, Shen W, Pecht MG. Health and lifespan prediction considering degradation patterns of lithium-ion batteries based on transferable attention neural network. *Energy* 2023;279:128137. <https://doi.org/10.1016/j.energy.2023.128137>.
- [5] Li Y, Liu K, Foley AM, Zülke A, Berecibar M, Nanini-Maury E, Van Mierlo J, Hosten HE. Data-driven health estimation and lifetime prediction of lithium-ion batteries: a review. *Renew Sustain Energy Rev* 2019;113:109254. <https://doi.org/10.1016/j.rser.2019.109254>.
- [6] Elmahallawy M, Elfouly T, Alouani A, Massoud AM. A comprehensive review of lithium-ion batteries modeling, and state of health and remaining useful lifetime prediction. *IEEE Access* 2022;10:119040–70. <https://doi.org/10.1109/ACCESS.2022.3221137>.
- [7] Lui YH, Li M, Downey A, Shen S, Nemani VP, Ye H, VanElzen C, Jain G, Hu S, Laflamme S, Hu C. Physics-based prognostics of implantable-grade lithium-ion battery for remaining useful life prediction. *J Power Sources* 2021;485:229327. <https://doi.org/10.1016/j.jpowsour.2020.229327>.
- [8] Lai X, Yi W, Cui Y, Qin C, Han X, Sun T, Zhou L, Zheng Y. Capacity estimation of lithium-ion cells by combining model-based and data-driven methods based on a sequential extended Kalman filter. *Energy* 2021;216:119233. <https://doi.org/10.1016/j.energy.2020.119233>.
- [9] Yang F, Song X, Dong G, Tsui K-L. A coulombic efficiency-based model for prognostics and health estimation of lithium-ion batteries. *Energy* 2019;171:1173–82. <https://doi.org/10.1016/j.energy.2019.01.083>.
- [10] Chang Y, Fang H, Zhang Y. A new hybrid method for the prediction of the remaining useful life of a lithium-ion battery. *Appl Energy* 2017;206:1564–78. <https://doi.org/10.1016/j.apenergy.2017.09.106>.
- [11] Chen Z, Shi N, Ji Y, Niu M, Wang Y. Lithium-ion batteries remaining useful life prediction based on BLS-RVM. *Energy* 2021;234:121269. <https://doi.org/10.1016/j.energy.2021.121269>.
- [12] Fei Z, Yang F, Tsui K-L, Li L, Zhang Z. Early prediction of battery lifetime via a machine learning based framework. *Energy* 2021;225:120205. <https://doi.org/10.1016/j.energy.2021.120205>.
- [13] Zhang Y, Peng Z, Guan Y, Wu L. Prognostics of battery cycle life in the early-cycle stage based on hybrid model. *Energy* 2021;221:119901. <https://doi.org/10.1016/j.energy.2021.119901>.
- [14] Yang F, Wang D, Xu F, Huang Z, Tsui K-L. Lifespan prediction of lithium-ion batteries based on various extracted features and gradient boosting regression tree model. *J Power Sources* 2020;476:228654. <https://doi.org/10.1016/j.jpowsour.2020.228654>.
- [15] Xu F, Yang F, Fei Z, Huang Z, Tsui K-L. Life prediction of lithium-ion batteries based on stacked denoising autoencoders. *Reliab Eng Syst Saf* 2021;208:107396. <https://doi.org/10.1016/j.ress.2020.107396>.
- [16] Yang Y. A machine-learning prediction method of lithium-ion battery life based on charge process for different applications. *Appl Energy* 2021;292:116897. <https://doi.org/10.1016/j.apenergy.2021.116897>.
- [17] Tang Y, Yang K, Zheng H, Zhang S, Zhang Z. Early prediction of lithium-ion battery lifetime via a hybrid deep learning model. *Measurement* 2022;199:111530. <https://doi.org/10.1016/j.measurement.2022.111530>.
- [18] Zhang Z, Zhang W, Yang K, Zhang S. Remaining useful life prediction of lithium-ion batteries based on attention mechanism and bidirectional long short-term

- memory network. Measurement 2022;204:112093. <https://doi.org/10.1016/j.measurement.2022.112093>.
- [19] Wang S, Fan Y, Jin S, Takyi-Aninakwa P, Fernandez C. Improved anti-noise adaptive long short-term memory neural network modeling for the robust remaining useful life prediction of lithium-ion batteries. Reliab Eng Syst Saf 2023; 230:108920. <https://doi.org/10.1016/j.ress.2022.108920>.
- [20] Zhang C, Wang H, Wu L. Life prediction model for lithium-ion battery considering fast-charging protocol. Energy 2023;263:126109. <https://doi.org/10.1016/j.energy.2022.126109>.
- [21] Ma G, Wang Z, Liu W, Fang J, Zhang Y, Ding H, Yuan Y. A two-stage integrated method for early prediction of remaining useful life of lithium-ion batteries. Knowl Base Syst 2023;259:110012. <https://doi.org/10.1016/j.knosys.2022.110012>.
- [22] Ma Y, Shan C, Gao J, Chen H. A novel method for state of health estimation of lithium-ion batteries based on improved LSTM and health indicators extraction. Energy 2022;251:123973. <https://doi.org/10.1016/j.energy.2022.123973>.
- [23] Tang A, Jiang Y, Yu Q, Zhang Z. A hybrid neural network model with attention mechanism for state of health estimation of lithium-ion batteries. J Energy Storage 2023;68:107734. <https://doi.org/10.1016/j.est.2023.107734>.
- [24] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. 2017. <http://arxiv.org/abs/1706.03762>. [Accessed 27 March 2023].
- [25] Fu P, Chu L, Hou Z, Hu J, Zhang Y. Transfer learning and vision transformer based state-of-health prediction of lithium-ion batteries, vol. 14; 2021.
- [26] Gu X, See KW, Li P, Shan K, Wang Y, Zhao L, Lim KC, Zhang N. A novel state-of-health estimation for the lithium-ion battery using a convolutional neural network and transformer model. Energy 2023;262:125501. <https://doi.org/10.1016/j.energy.2022.125501>.
- [27] Fei Z, Zhang Z, Tsui K-L. Deep learning powered online battery health estimation considering multi-timescale ageing dynamics and partial charging information. IEEE Transactions on Transportation Electrification 2023. <https://doi.org/10.1109/TTE.2023.3264438>. 1–1.
- [28] Yang K, Tang Y, Zhang S, Zhang Z. A deep learning approach to state of charge estimation of lithium-ion batteries based on dual-stage attention mechanism. Energy 2022;244:123233. <https://doi.org/10.1016/j.energy.2022.123233>.
- [29] Yang K, Wang Y, Tang Y, Zhang S, Zhang Z. A temporal convolution and gated recurrent unit network with attention for state of charge estimation of lithium-ion batteries. J Energy Storage 2023;72:108774. <https://doi.org/10.1016/j.est.2023.108774>.
- [30] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N. An image is worth 16x16 words: transformers for image recognition at scale. <http://arxiv.org/abs/2010.11929>. [Accessed 21 September 2023].
- [31] Gomez W, Wang F-K, Choi J-H. Li-ion battery capacity prediction using improved temporal fusion transformer model. Energy 2024;296:131114. <https://doi.org/10.1016/j.energy.2024.131114>.
- [32] Zhang W, Jia J, Pang X, Wen J, Shi Y, Zeng J. An improved transformer model for remaining useful life prediction of lithium-ion batteries under random charging and discharging. Electronics 2024;13:1423. <https://doi.org/10.3390/electronics13081423>.
- [33] Severson KA, Attia PM, Jin N, Perkins N, Jiang B, Yang Z, Chen MH, Aykol M, Herring PK, Fragedakis D, Bazant MZ, Harris SJ, Chueh WC, Braatz RD. Data-driven prediction of battery cycle life before capacity degradation. Nat Energy 2019;4:383–91. <https://doi.org/10.1038/s41560-019-0356-8>.
- [34] Attia PM, Grover A, Jin N, Severson KA, Markov TM, Liao Y-H, Chen MH, Cheong B, Perkins N, Yang Z, Herring PK, Aykol M, Harris SJ, Braatz RD, Ermon S, Chueh WC. Closed-loop optimization of fast-charging protocols for batteries with machine learning. Nature 2020;578:397–402. <https://doi.org/10.1038/s41586-020-1994-5>.
- [35] Chen C-F, Fan Q, Panda R. CrossViT: cross-attention multi-scale vision transformer for image classification. <http://arxiv.org/abs/2103.14899>. [Accessed 4 October 2023].
- [36] Saxena S, Ward L, Kubal J, Lu W, Babinec S, Paulson N. A convolutional neural network model for battery capacity fade curve prediction using early life data. J Power Sources 2022;542:231736. <https://doi.org/10.1016/j.jpowsour.2022.231736>.
- [37] Fei Z, Zhang Z, Yang F, Tsui K-L, Li L. Early-stage lifetime prediction for lithium-ion batteries: a deep learning framework jointly considering machine-learned and handcrafted data features. J Energy Storage 2022;52:104936. <https://doi.org/10.1016/j.est.2022.104936>.