This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TTE.2024.3427334, IEEE Transactions on Transportation Electrification

PFFN: A Parallel Feature Fusion Network for Remaining Useful Life Early Prediction of Lithium-ion Battery

Zhekang Dong, *Senior Member*, *IEEE*, Mengjie Yang, Junfan Wang, Hao Wang, Chun Sing Lai, *Senior Member*, *IEEE*, Xiaoyue Ji, *Member*, *IEEE*

***Abstract*—Remaining useful life (RUL) early prediction of lithium-ion battery is crucial to develop advanced battery health management and complete security assessment. However, most of existing methods still suffer from two limitations, i.e., the inadaptability to the different data distribution and the inability to capture the relationship between input series and RUL, which always make the RUL early prediction difficult and challengeable. To address these issues, this paper proposes a parallel feature fusion network (PFFN) for RUL early prediction of lithium-ion battery. Firstly, a feature selection strategy is designed to filter the optimal feature sets (containing cycle statistical features and domain knowledge-based features) that are most related to RUL of lithium-ion battery. Secondly, two specific Transformer encoders connected in parallel configuration are developed to integrate the cycle statistical features and domain knowledge-based features, respectively, achieving original RUL early prediction results. Furthermore, the Bayesian optimization is applied for global iterative optimization, aiming to enhance the prediction accuracy and generalization capability. A series of experiments are conducted with different data distributions. Experimental results demonstrate that the proposed PFFN outperforms the state-of-the-art (SOTA) methods, achieving 6.00%~27.61%, 0.58%~6.49%, and 5.95%~7.03% reduction in Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Score respectively.**

***Index Terms*— Remaining useful life (RUL), early prediction, lithium-ion battery, Bayesian optimization.**

I. Introduction

Enviromental pollution and energy crises have consistently been two significant issues faced by the international community. In recent years, many countries have vigorously developed new energy industries. Lithium-ion batteries (LIBs) are widely used in various power systems such as consumer electronics, electric vehicles, aircraft, and spacecrafts due to their high energy density, low costs, and environmental friendliness [1-4]. During the long-term repeated charge-discharge cycles, the capacity degradation of LIBs inevitably occurs. Accurately expressing the complex and dynamic degradation process of LIBs and realize the remaining useful life (RUL) early prediction can not only prevent catastrophic accidents caused by battery aging, but also avoid waste caused by premature battery replacement.

Generally, battery RUL prediction methods include physics-based approaches [8-15], data driven-based approaches [16-24], and hybrid approaches [25-29, 31, 32]. For physics-based approaches, some equivalent models of LIBs (e.g., the electrochemical models and equivalent circuit models) are required to establish by combining operational conditions and electrochemical reaction mechanisms. Based on this, physics-based approaches are usually accurate and interpretable for specific LIBs, while they still suffer from heavy computational burden and easy to fail in different LIB systems (i.e., low generalization performance). Different with physics-based approaches, data driven-based approaches mainly predict battery RUL by injecting LIB historical data to various machine/deep learning models [30]. In other words, this kind of prediction methods mainly rely on historical data to describe the battery capacity degradation process, instead of establishing equivalent models of LIBs. In recent years, hybrid approaches integrating physics-based approaches and data driven-based approaches become a search hotspot in battery RUL prediction. Although these methods have been proved effective in different LIB applications (i.e., good generalization ability), they still suffer from three limitations. Firstly, more than half of the battery life cycle data (about 60%~70%) has to be used in

This work was supported in part by the National Postdoctoral Researcher Support Program under Grant GZB20230356, Ministry of Science and Technology - Yangtze River Delta Science and Technology Innovation Program under Grant YDZX20233100004028, the National Natural Science Foundation of China under Grant 62206062, National Major Scientific Research Instrument Development Project of China under Grant 62227802. (Corresponding authors: *Xiaoyue Ji*).

Z Dong, M Yang, and J. Wang are with the School of Electronics and Information, Hangzhou Dianzi University, Hangzhou, China, 310018. (e-mail: englishp@hdu.edu.cn; 221040056@hdu.edu.cn; wangjunfan@hdu.edu.cn)

H. Wang is with the McMaster Automotive Resource Center, McMaster University, Hamilton, ON, L8P0A6 Canada. (e-mail: wangh466@mcmaster.ca)

C. S. Lai is with Department of Electronic and Computer Engineering, Brunel University London, London, UB8 3PH, UK, (e-mail: chunsing.lai@brunel.ac.uk)

X. Ji is with the Center for Brain-Inspired Computing Research (CBICR), Beijing Innovation Center for Future Chip, Optical Memory National Engineering Research Center, Department of Precision Instrument, Tsinghua University, Beijing 100084, China. (e-mail: jixiaoyue@mail.tsinghua.edu.cn).

existing RUL prediction models. Most existing research fails to capture the correlation between early-stage (or limited data) and RUL, which makes the high-precision RUL early prediction hard to achieve. Secondly, the feature selection mechanism is not comprehensively considered in RUL prediction models, which may limit the full use of multi-dimensional input information and make the accurate modeling of the LIB degradation process challenging and difficult. Finally, the difference in LIB cycle data distribution is not considered in almost all the RUL prediction models, which may lead to the inferior prediction performance on datasets with different battery cycle life distributions.

To address these aforementioned issues, this paper proposes a novel architecture for RUL early prediction of LIB. The main contributions of this paper are summarized as follows:

1) An optimal feature selection strategy is developed, the feature sets (containing both cycle statistical features and domain knowledge-based features) that are most related to RUL of LIBs can be selected, ensuring a more accurate representation of the non-linear degradation process of LIB life.

2) A parallel feature fusion network (PFFN) with two specific Transformer encoders connected in parallel configuration is developed, which can sufficiently learn the relationship between RUL and limited input data, achieving accurate RUL early prediction with a minimum of only 5% of cycle data.

3) A Bayesian-based model optimization algorithm is designed, aiming to enhance the prediction accuracy and generalization capability caused by significant differences in data distribution.

The rest of the paper is organized as follows. Section II reviews related work on RUL prediction of LIB. Section III describes the proposed prediction model in detail. Section IV begins with an introduction to the datasets used in the experiments, followed by the feature selection process and experimental analysis and results. Finally, Section V presents the conclusion and future direction of the study.

II. Related Work

The research methods for LIB RUL prediction initially relied predominantly on physics-based approaches. Physics-based approaches always combine non-linear filtering algorithms to perform state estimation and adjust model parameters to predict battery RUL [8-15]. The identification and updating of model parameters are achieved through the utilization of state data and statistical methods. Considering the limitations of low generalization ability and accuracy in physics-based methods, the alternative data driven-based methods have been widely used in different LIB applications. Data driven-based approaches can be further divided into two types: machine learning methods [16-19] and deep learning methods [20-24]. The traditional machine learning methods excessively relies on health features (mainly cycle statistical features) extracted from the working principles and characteristics. Meanwhile, the traditional machine learning models are often insufficient to capture the interrelationships within the data, making it difficult to achieve high-precision RUL prediction.

With the development of deep learning, several neural network models have been widely applied in RUL prediction of LIB. Zhang et al. [20] used long short-term memory (LSTM) recurrent neural network (RNN) to learn the long-term dependence in the degradation of LIB capacity, and the elastic mean square backpropagation method was used for adaptive optimization to construct a clear capacity-oriented RUL predictor. Ding et al. [21] proposed a convolutional neural network (CNN) to perform RUL prediction of LIBs when enough online data are available. Chen et al. [22] applied the denoising auto-encoder (DAE) and Transformer hybrid network to perform battery RUL prediction, where an objective function is designed to bridge the denoising subtask and prediction subtask. Ardeshiri et al. [23] proposed a gated recurrent unit (GRU) neural network to extract sufficient statistical features from voltage, current and temperature data. An adaptive gradient descent algorithm is used to learn parameters automatically and built capacity-oriented RUL predictor. Li et al. [24] designed an improved LSTM neural network to perform the RUL and state of health prediction respectively, which couples the input gate and forget gate through a fixed connection, and adds a peephole to the output gate to shield the unwanted error signals.

Hybrid approaches can be used to achieve different purposes at different stages of RUL prediction. Ma et al. [25] proposed a fusion prediction model using convolutional neural network (CNN) and LSTM, which can utilize the automatic feature extraction capability of CNN and combine the advantages of LSTM to capture temporal dependence. Xu et al. [26] designed a hybrid deep learning model that effectively combines domain knowledge-based features with handcrafted latent features learned by deep networks to improve the performance of RUL prediction. Zhang et al. [27] established two artificial neural network (ANN) models aiming to estimate the SOH and RUL simultaneously. Wang et al. [28] established a new RUL prediction method, consisting of optimal decision of alignment ratio and integrated prediction using Bayesian model averaging (BMA). Zhu et al. [29] proposed a hybrid approach using Grey–Markov chain model and improved Gaussian process for battery capacity and RUL prediction, aiming to deal with the problem of capacity regeneration. Karmawijaya et al. [31] proposed an effective hybrid method generating genetic algorithm (GA) and deep learning neural network (DLNN) for battery RUL prediction. Qu et al. [32] proposed a neural network-based method combining LSTM with particle swarm optimization (PSO) and an attention mechanism for RUL prediction. The Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) was utilized improve the accuracy of prediction. However, all hybrid approaches share the demerits of high computational complexity and uncertainty.

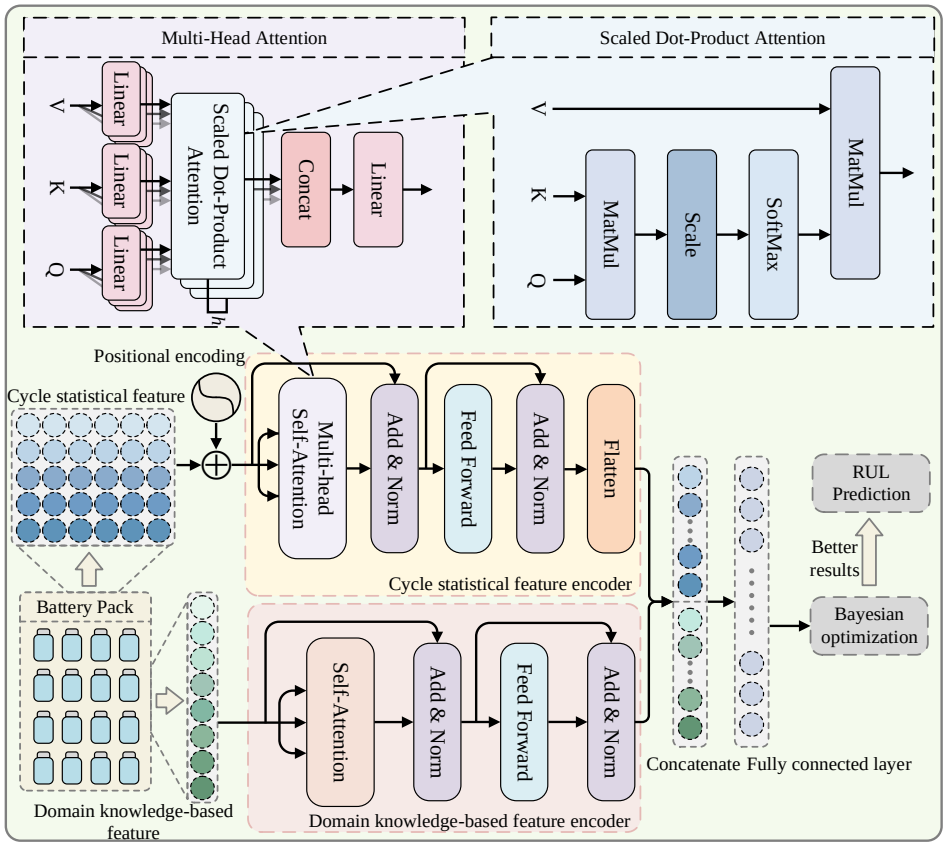


Fig. 1. Framework of PFFN

Although these above-mentioned methods [11-29, 31, 32] have achieved great success in studying battery life degradation trends, they still suffer from some drawbacks, e.g., highly relying on enough condition monitoring data. In other words, it is often costly and time-consuming to obtain sufficient monitoring data with run-to-failure labels for model development. Our research focuses on the RUL early prediction of LIB, namely predicting RUL with limited data (i.e., before the battery shows an obvious degradation trend).

III. Mathodology

In this article, the proposed PFFN takes advantage of the Transformer for powerful time-series data modeling and the multi-head attention mechanism to capture the interaction between different features. Cycle statistical features and domain knowledge-based features are processed respectively to achieve RUL early prediction of LIBs. Fig. 1 shows the framework of the proposed PFFN. The framework includes cycle statistical feature encoder, domain knowledge-based feature encoder, feature fusion and prediction output, Bayesian optimization module.

A. Feature encoder

As a time-series forecasting task, the order of input data for battery RUL prediction is particularly important. Unlike CNN and RNN, which inherently capture the position information of sequences, it is firstly required to encode the position of sequence data to inject position information. Here, we use sine and cosine functions of different frequency [33]:

$$PE(t,2k)=sin\left(t/10000^{2k/d_{model}}\right) \tag{1}$$

$$PE(t,2k+1)=cos\left(t/10000^{2k/d_{model}}\right) \tag{2}$$

where t is the number of cycle, k is the feature dimension, and d_{model} is the input dimension.

The cycle statistical feature encoder consists of three main sub-layers: a multi-head attention layer, a feedforward layer, and a flatten layer. A residual connection and layer normalization (Add & Norm) is inserted in the middle, respectively. Residual connection aims to solve the problem of gradient vanishing and weight matrix degradation. Layer normalization speeds up the training process by normalizing the layer activation values to make the model converge faster. The flatten layer transitions the high-dimensional features to a fully connected (FC) layer. Multi-head attention is a refinement of attention mechanism, designed to capture the interdependence between different features (as shown in Fig. 1). Q_c, K_c, V_c denote query, key, and value, respectively. The Scaled Dot-Product Attention is defined by:

$$Attention_c\left(Q_c,K_c,V_c\right)=softmax\left(\frac{Q_cK_c^T}{\sqrt{d_{model}}}\right)V_c \tag{3}$$

Multi-head attention enables the model to capture information from different representation spaces at different positions simultaneously. It can be defined by:

$$MultiHead\left(Q_c,K_c,V_c\right)=Concat\left(\left\{head_j\right\}_{j=1}^{h}\right)W^O \tag{4}$$

$$head_j=Attention_c\left(Q_cW_j^Q,K_cW_j^K,V_cW_j^V\right) \tag{5}$$

where h is the number of attention heads. T is the transposition operation. j denotes the j_{th} attention head. W_j^Q, W_j^K, W_j^V and W_j^O denote trainable parameter matrices.

The domain knowledge feature encoder is similar in the structure of cycle statistical feature encoder, which serves to process the extracted domain knowledge-based features. The only difference is that a self-attention mechanism is used in the domain knowledge-based feature encoder. The encoder first obtains the query, key, and value denoted as Q_d, K_d, and V_d, and then the self-attention weights for domain knowledge-based features can be calculated by:

$$Attention_d\left(Q_d, K_d, V_d\right) = softmax\left(\frac{Q_d K_d^T}{\sqrt{d_{model}}}\right)V_d \tag{6}$$

B. Feature fusion and prediction output

After the features processing is completed by the cycle statistical feature encoder and the domain knowledge-based feature encoder, the information from these two parts is fused to generate a new feature mapping:

$$F_r = Concat\left(F_c, F_d\right)W^f \tag{7}$$

where F_r is the input of the prediction layer. F_c and F_d are the feature information extracted from the two feature encoders. W^f is a trainable matrix that enables the model to capture information from both F_c and F_d.

The RUL prediction results are obtained through a FC layer:

$$RUL = f\left(W_p F_r + b_p\right) \tag{8}$$

where W_p, b_p, and f (.) denote the weight, bias, and mapping function of the prediction layer, respectively.

To illustrate the RUL early prediction process more intuitively and effectively, the specific pseudocode is provided in Table I.

TABLE I
OUTLINE OF PFFN FOR RUL EARLY PREDICTION OF LIBS

Input: Cycle statistical feature *train_x_cycle*, Domain knowledge-based feature *train_x_domain*
Output: RUL early prediction result
Begin
Part1: Cycle statistical feature encoder
Step1: Conduct positional encoding operation with *train_x_cycle* in Equation(1-2).
Step2: Calculat multi-head attention weights *MultiHead*(Q_c, K_c, V_c) after position encoding in Eq. (3-5).
Step3: X_c= Q_c, K_c, V_c, conduct *Add&Norm* operation with X_c=LayerNorm(X+*MultiHead*(Q_c,K_c,V_c)).
Step4: Conduct *feed forward* and *Add&Norm* operation with X_c=LayerNorm(X+*Feed_Forward*(X_c)).
Step5: Conduct *Flatten* operation to make the data one-dimensional transition to fully connected layer with F_c=Flatten(X).
Part2: Domain knowledge-based feature encoder
Step1: Calculat attention weights *Attention*$_d$(Q_d, K_d, V_d) with *train_x_domain* in Eq. (6).
Step2: Conduct *Add&Norm* operation with X_d=LayerNorm(X+*Attention*$_d$(Q_d, K_d, V_d)).
Step3: Conduct *feed forward* and *Add&Norm* operation with F_d=LayerNorm(X+*Feed_Forward*(X_d)).
Part3: Feature Fusion and Prediction
Step1: Integrate the two parts of feature information to form a new feature map in Eq. (7).
Step2: Obtain the prediction result of RUL through a fully connected layer in Eq. (8).
End

C. Bayesian optimization module

The Bayesian optimization module is employed to improve the RUL prediction performance by optimizing the hyperparameters in PFFN. Specifically, an initial set of candidate solutions is first generated, i.e., search space of hyperparameters set: $S=\{s_1, \ldots, s_i, \ldots, s_n\}$, where s_i denotes the combination of different hyperparameters (including learning rate, number of heads, number of hidden layers, and etc.). Here, the probability model for Bayesian optimization is a Gaussian process (GP). Assuming that the black-box objective function $f(x)$ is randomly drawn from the GP. The multivariate GP model can be expressed by [32]:

$$P\left(s\right) = \frac{1}{\left(2\pi\right)^{\frac{n}{2}}\left|cov\right|^{\frac{1}{2}}} exp\left(-\frac{1}{2}\left(s-\mu\right)^T cov\left(s-\mu\right)^{-1}\right) \tag{9}$$

where μ and cov denote the mean and covariance values, respectively.

A set of solutions is randomly determined to calculate the acquisition function $\alpha(s)$, which can be expressed by:

$$\alpha(s) = \int_{-\infty}^{f'}\left(f' - f\left(s\right)\right)N\left(f; \mu\left(s\right), cov\left(s,s\right)\right)df \tag{10}$$

where f' represents the current minimum value of $f(s)$.

The maximum value of the acquisition function is used to guide the updating of hyperparameter combination in PFFN, which can be written by:

$$s_{n+1} = arg\ max_s\ \alpha\left(s_n; D_n\right) \tag{11}$$

where $argmax_s(\cdot)$ is the maximum a posteriori estimation operation, D_n represents the observed hyperparameter combination set.

Then the parameters are updated until the optimal process is totally completed. Updating hyperparameters in PFFN is able to obtain the optimal RUL prediction model. The specific pseudocode is provided in Table II.

TABLE II
BAYESIAN OPTIMIZATION PROCESS

Input: Search space of hyperparameters sets: $S=\{ s_1, \ldots, s_i, \ldots, s_n\}$,
Objective function: $f(s)$~$GP(s)$.
Output: Best hyperparameter sets *s_opt*.
Begin
randomly select a set of solutions: s_i.
for n = 1, 2, …, **do**
select new s_{n+1} by optimizing acquisition function α.
$s_{n+1} = \underset{s}{argmax}\,\alpha\left(s_n; D_n\right)$
query objective function to obtain y_{n+1}.
augment data $D_{n+1}=\{D_n,(s_{n+1}, y_{n+1})\}$.
update statistical model.
end for
determine the optimal hyperparameter combination: *s_opt*.
End

IV. EXPERIMENTS

In this section, the dataset utilized for experiments and the feature selection method are described respectively. Meanwhile, a series of comparison experiments are conducted to verify the validity and effectiveness of the entire scheme. The specific experiment environment is provided below: CPU/Intel(R) Core (TM) i7-10700, memory/24G, graphics card/NVIDIA GeForce GTX 3080, programming language/Python 3.10, deep learning

framework/Pytorch. The corresponding source code is publicly available at http://github.com/yyangmj/PFFN.

The hyperparameters and structure parameters of the model are provided in Table III and Table IV, respectively.

TABLE III
HYPERPARAMETERS OF THE MODEL

Hyperparameters	Value	Hyperparameters	Value
Number of attention head$_c$	2	Learning rate	1e-6
Hidden dimension	12	Epochs	4000
Dimension of attention$_c$	32	Dimension of attention$_d$	8
Dropout	0.1	/	/

TABLE IV
STRUCTURE PARAMETERS OF THE MODEL

Model structure	Value	Model structure	Value
ID$_c$	[BS, 32, 100]	ID$_d$	[BS, 8]
Attention$_c$	[BS, 32, 2]	Attention$_d$	[BS, 8, 1]
LN$_c$	32	LN$_d$	8
FCL$_c$	[32, 12]	FCL$_d$	[8, 8]
Flatten	[BS, 3200]	FCLF	[24, 1]

Note: ID→ Input dimension; LN→ Layer normalization; FCL→ Fully connected layer; FCLF→ Fully connected layer fusion; BS→ Batch size.

A. Dataset

The A123 dataset generated by Severson et al [18] is employed for training and testing in this work. The dataset comprises 124 commercial LPF/graphite batteries (A123 system, model APR18650M1A), each with a nominal capacity of 1.10Ah and a nominal voltage of 3.3V. These batteries were subjected to continuous charging-discharging cycles under various fast charging conditions until they reach the point of failure. Cycle tests were conducted at a constant temperature of 30°C with different fast charging conditions and the same discharging conditions (4C to 2.0V, where 1C is 1.1A) until reaching end-of-life (EOL=80% BOL, begin-of-life). The dataset is divided into three sub-datasets: the first sub-dataset serves as the training set, the second sub-dataset is the primary test set, and the remaining sub-dataset is the secondary test set. The training and testing phases involve LIB data from entire battery packs rather than single cell. Due to calendar aging, the batteries in secondary test set exhibit significant deviations compared to the other two battery sub-datasets.

Fig. 2 compares the cycle life distribution and capacity distribution of individual batteries in these three sub-datasets. It can be observed that the data distribution in the training set and the primary test set is similar, while the secondary test set exhibits significant differences in data distribution compared to the other two sub-datasets. Table V summarizes the number of batteries in each cycle and initial capacity range, and Table VI lists the parameters and experimental conditions of LIBs in the dataset. Notably, the RUL prediction range is from 75% to 96%.

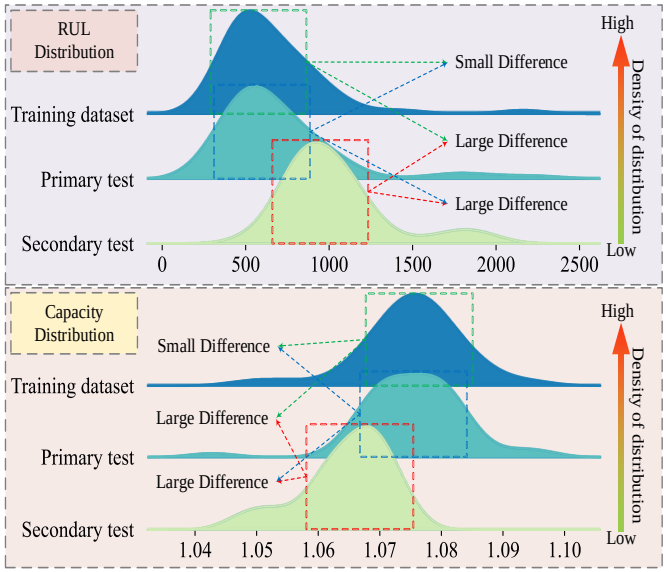


Fig. 2. Cycle life and initial capacity distribution of three sub-datasets.

TABLE VI
PARAMETERS AND EXPERIMENTAL CONDITIONS OF LIBS

Parameters	Value	Parameters	Value
Number of batteries	123	Charging conditions	72
Battery materials	LFP/graphite	Core cell weight	39g
Battery type	APR18650M1A	Nominal voltage	3.3V
Length	64.95±0.20mm	Nominal capacity	1.1Ah
Diameter	18.20±0.20mm	Test temperature	30°C

B. Feature Selection

According to [18], Severson et al. explored the relationship between domain knowledge-based features and RUL, obtaining 21 features based on domain knowledge. To reduce the computational burden, the K-means clustering is applied to group the 21 features. The basic principle is to iteratively search for K clusters that minimizes the loss function corresponding to the clustering results. The loss function is defined as the sum of the squared errors of the distance between each feature and the centroid of its corresponding cluster:

$$J(c,\mu)=\sum_{i=1}^{M}\left\|x_i-\mu_{c_i}\right\|^2 \tag{12}$$

where x_i is the ith feature, c_i is the cluster to which x_i belongs, μ_{ci} denotes the centroid of the corresponding cluster, and M is the total number of features.

The aforementioned features are categorized into four clusters. Considering the non-linear degradation of battery capacity in the early-stage, the Spearman correlation coefficient is used to assess the relationship between features and battery RUL. It can be mathematically expressed by:

TABLE V
DISTRIBUTION OF BATTERY CYCLE LIFE AND INITIAL CAPACITY IN THE DATASET

Cycle life	Training Dataset	Primary test	Secondary test	Initial capacity (Ah)	Training Dataset	Primary test	Secondary test
0-500	18	14	0	1.04-1.05	1	1	2
500-1000	19	22	21	1.05-1.06	1	0	6
1000-1500	3	3	15	1.06-1.07	7	10	26
1500-2000	0	2	4	1.07-1.08	25	24	6
2000-2500	1	1	0	1.08-1.10	7	7	0

$$\rho(X,Y)=1-\frac{6\sum_{i=1}^{N}(R(X_i)-R(Y_i))}{N(N^2-1)} \tag{13}$$

where $R(X_i)$ and $R(Y_i)$ denote the rank of each data point in vectors X and Y, and N is the total number of samples.

Here, the correlation analysis and clustering results are shown in Fig. 3.

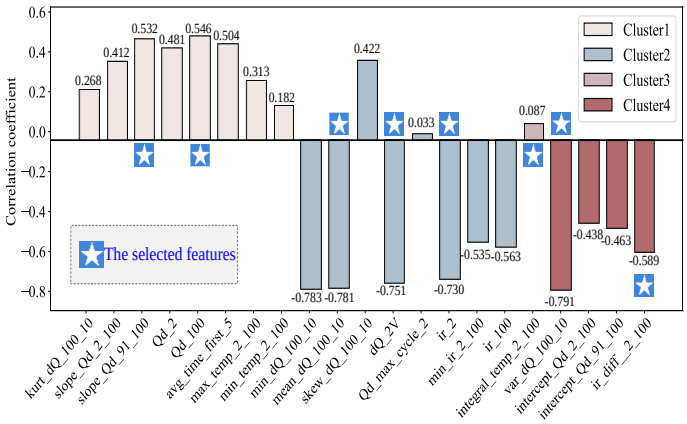


Fig. 3. Feature clustering and correlation analysis.

TABLE VII
FEATURE SELECTION RESULTS

Selected Features	Physical meaning
*mean_dQ*_100_10	Average difference in discharge capacity between the 100th and 10th cycle.
*var_dQ*_100_10	Variance of the difference in discharge Capacity between the 100th and 10th cycle.
*slope_Qd*_91_100	Slope of capacity degradation curve from the 91st to the 100th cycle.
*Qd*_100	Discharge capacity of the 100th cycle.
dQ_2V	Discharge capacity at 2V.
*integral_temp*_2_100	Integration of temperature over time from the 2nd cycle to the 100th cycle.
*ir*_2	Internal resistance of the 2nd cycle
*ir_diff*_2_100	Difference of internal resistance between 100th cycle and 2nd cycle

Then, 8 features with high correlation are selected and exhibited in Table VII. Combining the results of the correlation analysis, it can be observed that the correlation of the feature *integral_t* with RUL is extremely weak. However, when we attempt to replace this feature with others from different clusters that exhibit higher correlation, the prediction performance consistently shows different degrees of decline. The main reason may that the correlation coefficient is not the sole metric to evaluate the mutual relationship between features and battery RUL.

Furthermore, considering it is impractical to feed the cycle battery measurement data into the neural network directly, the volume of measurement data per cycle is substantial, which may lead to a significant increase in computational complexity. Hence, we process the raw measurement data at the cycle level and summarize the minimum, maximum, mean, variance, skewness, and kurtosis for each cycle. These cycle statistical features offer different perspectives on the trend of battery capacity degradation.

C. Evaluation metrics

To evaluate the performance of the proposed PFFN, three evaluation metrics are used, including root mean square error (RMSE), mean absolute percentage error (MAPE), and Score function [34, 35]. The former two metrics are defined by:

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i-y_i)^2} \tag{14}$$

$$MAPE=\frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i-y_i|}{y_i}\times 100\% \tag{15}$$

where $\hat{y}_i$ and y_i denote the predicted results and the ground truth of RUL, respectively, and n denotes the number of battery samples.

Notably, the remaining metric (i.e., Score function) is utilized in battery RUL prediction for the first time, which provides an additional perspective to assess the safety and effectiveness of the model. Specifically, it can be mathematically expressed by:

$$Score=\begin{cases}\sum_{i=1}^{n}e^{-\frac{\hat{y}_i-y_i}{100a_1}}-1, & \hat{y}_i-y_i<0\\ \sum_{i=1}^{n}e^{\frac{\hat{y}_i-y_i}{100a_2}}-1, & \hat{y}_i-y_i\geq 0\end{cases} \tag{16}$$

where a_1=13, a_2=10. For safety-critical devices like batteries, predicting a larger RUL often poses a higher risk of severe consequence compared to predicting smaller RUL. The score increases exponentially as the error between the predicted results and ground truth increases, and the score function is penalized more severely when the predicted RUL exceeds the corresponding ground truth. This asymmetrical score function reflects the preference for predicted results to be smaller than the ground truth. Notably, the parameters a_1 and a_2 can be adjusted appropriately to quantify the degree of this preference.

D. Comparison experiments

The loss during the training process is shown in Fig. 4. It can be observed that there exist some spike pulses with large loss values. This fluctuation phenomenon can be deemed as the superposition of the loss, which is caused by the dynamic adjustment of the learning rate and the Bayesian optimization. Notably, this spike pulse phenomenon has been proved having little effect on the accuracy of the RUL prediction.

Firstly, we validate the effectiveness of the proposed feature selection strategy. Specifically, we arrange the selected domain knowledge-based features according to their correlation coefficients and incrementally add them as input to PFFN. After incorporating the first eight features, we further add two features with the highest correlation from the remaining thirteen features for joint validation. The experimental results are presented in Fig 5. It is evident that the performance of PFFN is optimal (i.e., the minimum RMSE, MAPE, and Score) when the input domain knowledge-based features are the eight features collected in Table VII. The prediction performance declines with different degree when the features are added or subtracted from this. Notably, the performance of the model is worst when the number of selected features is 1.

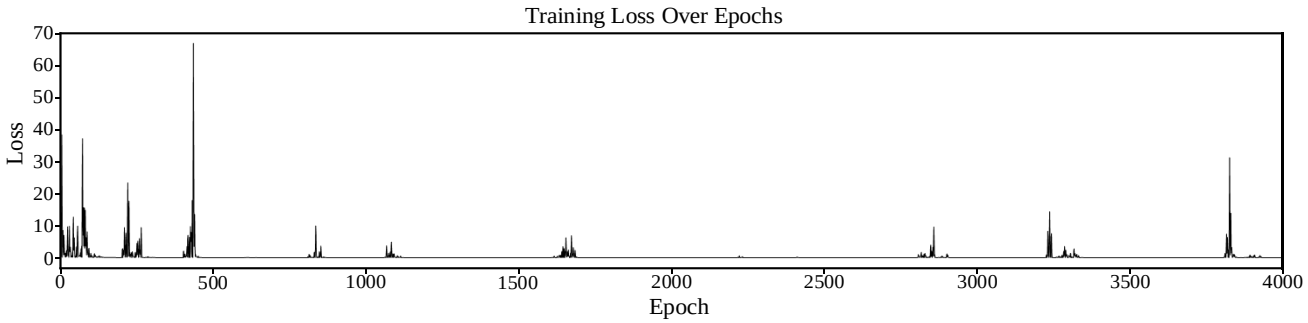


Fig. 4. Loss curve during the training process

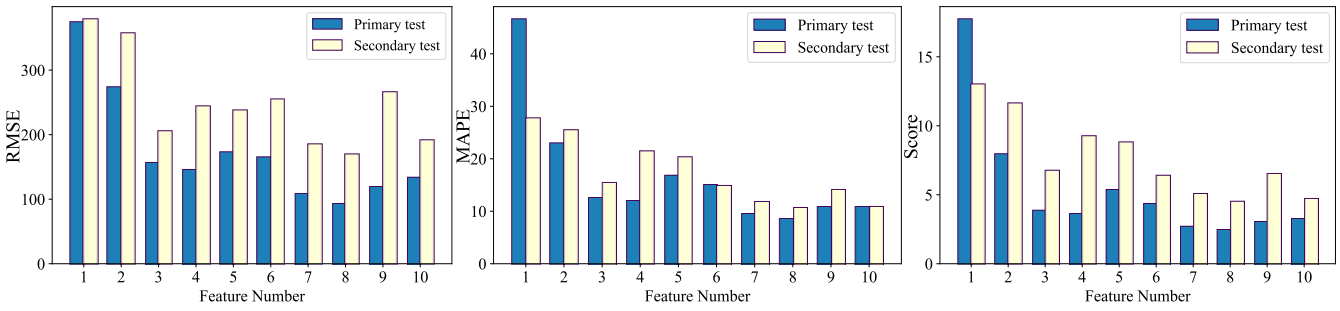


Fig. 5. Performance comparison of different input features. (a) The RMSE results. (b) The MAPE results. (c) The Score results.

Next, we compared the performance of our model with other competitors, including: Elastic [18], support vector regression (SVR) [19], LSTM [20], Transformer [22], CNN-LSTM [25], CNN-FC [26], ANNs [27], DFFNDDS [33], Multimodal-AF [36]. The specific prediction performance comparison results are presented in Table VIII.

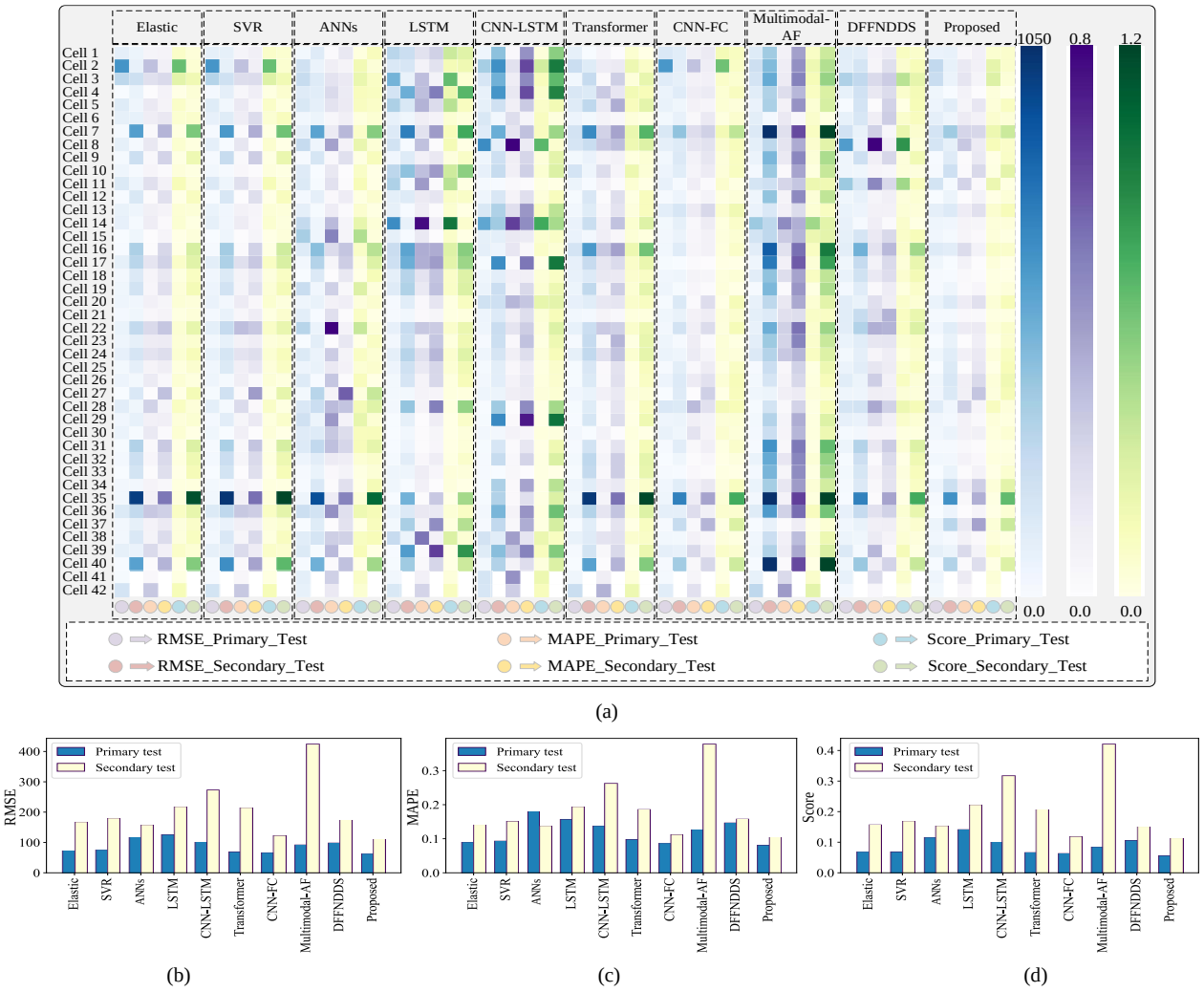
From Table VIII, when the different feature sets are input to different models, the RUL prediction results are varying. Specifically, the methods with two kinds of input features (i.e., including both the domain knowledge-based feature and cycle statistical feature) outperform the single feature-based methods (e.g., Elastic, SVR, ANNs), which means both the domain knowledge-based feature and cycle statistical feature have close connection with the RUL of LIBs. Meanwhile, compared with the CNN-FC, Multimodal-AF, and DFFNDDS, the proposed PFFN achieves best prediction performance (having the minimum RMSE, MAPE, and Score values) on primary and secondary test set. The main reason may be that the proposed PFFN has the ability to effectively express the capacity degradation of LIB in the early stage due to the parallel network configuration. Meanwhile, the proposed feature selection strategy makes the most related to RUL of LIBs are applied in the proposed PFFN, ensuring a more accurate RUL prediction.

To verify the superiority of proposed method in RUL prediction task, the evaluation metrics of the prediction results on primary and secondary test sets are visualized in Fig. 6. Specifically, Fig. 6(a) shows the RMSE, MAPE, and Score of different RUL prediction methods for each cell on two test sets. In Fig. 6(a), most of baselines can obtain relatively high prediction performance on primary test set, while the generalization ability on secondary test set is worse. Compared with the baseline methods, the proposed PFFN not only achieves the best prediction performance on primary test set, but also improves the generalization ability on secondary test set. The average prediction results (i.e., RMSE, MAPE, and Score) on primary and secondary test sets are depicted in Figs.

TABLE VIII
PERFORMANCE COMPARISON OF DIFFERENT METHODS

Method	Input Feature	Primary Test			Secondary Test		
		RMSE	MAPE	Score	RMSE	MAPE	Score
Elastic [18]	Domain knowledge-based feature	129.35	9.16[3]	2.88	254.69	14.11	6.24
SVR [19]	Domain knowledge-based feature	130.74	9.46	2.96	270.44	15.11	6.71
LSTM [20]	Cycle statistical feature	191.55	15.87	5.92	291.58	19.51	8.86
Transformer [22]	Cycle statistical feature	95.81[2]	9.74	2.85[3]	293.58	18.68	8.24
CNN-LSTM [25]	Cycle statistical feature	171.22	13.93	4.27	331.73	26.23	12.67
CNN-FC [26]	Cycle statistical feature+ Domain knowledge-based feature	119.48[3]	8.61[2]	2.70[2]	179.64[2]	11.40[2]	4.87[2]
ANNs [27]	Domain knowledge-based feature	148.69	18.12	4.81	240.21	13.77[3]	6.18
DFFNDDS [33]	Cycle statistical feature+ Domain knowledge-based feature	148.23	14.66	4.48	222.53[3]	15.80	6.04[3]
Multimodal-AF [36]	Cycle statistical feature+ Domain knowledge-based feature	120.79	12.74	3.59	504.16	37.59	16.74
This work	Cycle statistical feature+ Domain knowledge-based feature	**93.03**[1]	**8.56**[1]	**2.51**[1]	**168.86**[1]	**10.66**[1]	**4.58**[1]

Note: The superscript 1, 2, and 3 denote the first, second, and third place.



(a) (b) (c) (d)

Fig. 6. Comparison of the performance of different methods for single-cell batteries. (a) The prediction performance for each cell. (b) The RMSE results. (c) The MAPE results. (d) The Score results.

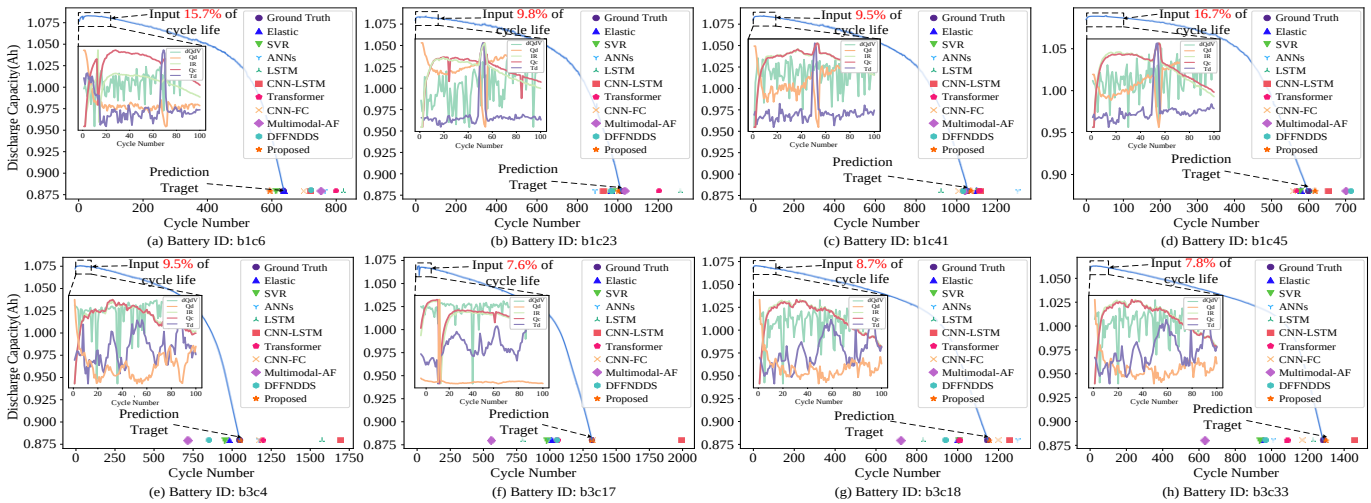


Fig. 7. The prediction results with other competitors by randomly selecting eight batteries.

6(b)~(d). The experimental results demonstrate that the proposed PFFN outperforms other competitors in terms of RMSE, MAPE, and Score.

Fig. 7 exhibits the prediction results with other competitors by randomly selecting eight batteries. In Fig. 7, the 5% to 20% battery data before the degradation is utilized. Notably, Batteries (illustrated in Figs. 7(a)~(d)) belong to primary test set, while batteries (illustrated in Figs. 7(e)~(h)) belong to secondary test set. The prediction results show that the proposed PFFN can accurately predict the RUL using limited LIB life cycles compared to other methods, enabling RUL early prediction of LIB.

E. Exploration experiments

To substantiate the efficacy of feature extraction method, as well as the contribution of cycle statistical features and domain knowledge-based features, the ablation experiment is conducted, as shown in Table IX.

TABLE IX
PERFORMANCE COMPARISON OF DIFFERENT MODULE

Feature		Primary Test			Secondary Test		
CSF*	DKBF*	RMSE	MAPE	Score	RMSE	MAPE	Score
Encoder	FC	163.47	11.06	3.50	235.99	20.03	8.44
CNN	FC	106.88	9.55	2.82	224.98	17.00	7.54
CNN	Encoder	150.77	14.13	4.19	344.41	22.12	9.72
Encoder	Encoder	**93.03**	**8.56**	**2.51**	**168.86**	**10.66**	**4.58**

Note: CSF→ Cycle statistical features; DKBF→ Domain knowledge-based features.

From Table IX, for the extraction of cycle statistical features, the performance of CNN model and Transformer encoder is compared. As for the domain knowledge-based features, we compare the performance of fully connected network and Transformer encoder. The experimental results demonstrate that the proposed method achieves the best prediction results compared with the other three combination strategies. The main reason maybe that the multi-head attention mechanism used in Transformer encoder can effectively capture interrelationship between features with long-term dependencies.

The linear correlation method (i.e., Pearson) and non-linear correlation method (i.e., Spearman) are used to measure the monotonic relationship between features and RUL, respectively. Meanwhile, these two correlation methods without/with K-means clustering algorithm are compared to evaluate different feature selection strategies on RUL prediction performance, as shown in Table X. Compared to non-linear correlation method without/with K-means clustering algorithm, linear correlation method without/with K-means clustering algorithm generally achieves better prediction performance. Meanwhile, the K-means clustering algorithm can further improve the prediction performance in both scenarios.

TABLE X
PERFORMANCE OF DIFFERENT FEATURE SELECTION METHODS

Methods	Primary Test			Secondary Test		
	RMSE	MAPE	Score	RMSE	MAPE	Score
Pearson	124.43	9.42	2.81	268.98	22.14	10.06
P&K*	106.88	9.55	2.82	224.98	17.00	7.54
Spearman	128.36	11.62	3.41	313.70	18.47	8.28
S&K*	**93.03**	**8.56**	**2.51**	**168.86**	**10.66**	**4.58**

Note: P&K→ Pearson with K-means clustering algorithm; S&K→ Spearman with K-means clustering algorithm.

To demonstrate the effectiveness of core components in different modules, the ablation study is carried out, as presented in Table XI. Specifically, the proposed PFFN is performed under cycle statistics features, and a discernible decline in prediction performance can be observed. The experimental results demonstrate that the domain knowledge-based features can provide supplementary information in RUL prediction task. Then, the impact of the Bayesian optimization module is investigated by either ablating it or replacing it with other optimization methods (i.e., random search and grid search). The experimental results reveal that Bayesian optimization module significantly improves prediction performance, with varying degrees of error optimization on primary and secondary test sets.

TABLE XI
MODULAR ABLATION EXPERIMENTS

Methods		Primary Test			Secondary Test		
		RMSE	MAPE	Score	RMSE	MAPE	Score
Full Model		**93.03**	**8.56**	**2.51**	**168.86**	**10.66**	**4.57**
w/o DKBF		366.34	27.72	17.83	564.01	43.90	19.94
w/o BOA	×	115.15	11.04	3.17	217.87	16.04	6.89
	RS	106.50	9.80	2.88	186.95	11.98	5.17
	GS	101.62	9.22	2.72	275.44	11.39	4.88

Note: BOA→Bayesian optimization; ×→Without optimization; RS→Random search; GS→Grid search.

To explore the relationship between prediction range and accuracy, the minimum RMSE, maximum RMSE, average RMSE, and median RMSE for each prediction range are calculated and recorded in Table XII.

TABLE XII
RELATIONSHIP BETWEEN PREDICTION RANGE AND CORRESPONDING PREDICTION ACCURACY

Prediction range	Minimum RMSE	Average RMSE	Median RMSE	Maximum RMSE
75%~80%	2.14	31.54	36.39	93.78
80%~90%	1.61	78.89	58.36	222.89
90%~93.3%	1.77	63.76	54.85	219.6
93.3%~95%	71.14	300.39	286.37	657.38
95%~96%	278.28	278.28	278.28	278.28

From Table XII, the proposed model has higher prediction accuracy when more LIB data is used (i.e., the lower prediction range). Otherwise, when less LIB data is used (i.e., the higher prediction range), the prediction accuracy is lower. The main reason may be that the less LIB data of long-life batteries contains relatively little useful information for RUL prediction, which makes it challenging to learn the capacity degradation process.

Furthermore, we explore the impact of different input cycles, and the corresponding results are illustrated in Fig. 8.

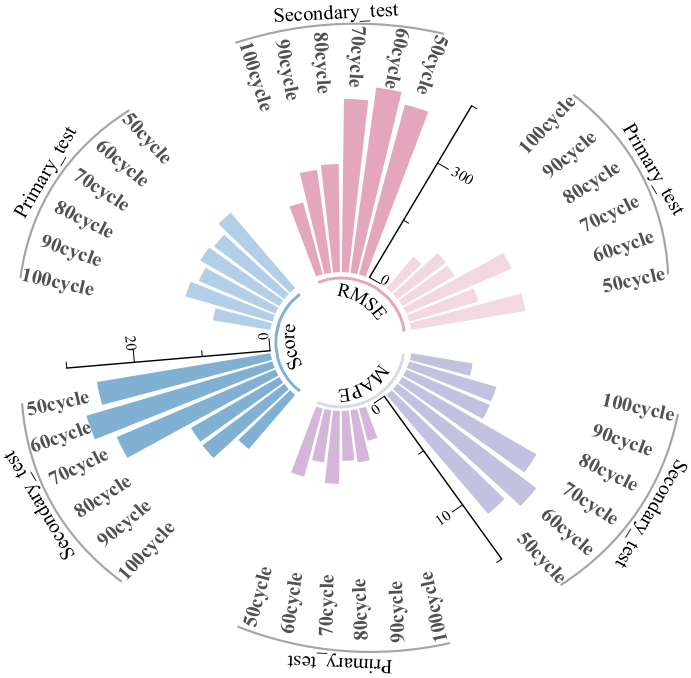


Fig. 8. Performance comparison of different input cycles.

From Fig. 8, as the number of input cycles decreases (100 cycles to 50 cycles), the information received and processed by the model also reduce, making it challenging to learn the complex trend of capacity degradation. At the same time, the model fails to capture the data distribution adequately, resulting in poorer prediction performance for battery packs with significant differences in the distribution of cycle life compared to training data.

V. Conclusion

In this paper, we proposed a parallel feature fusion network (PFFN) that can effectively integrate the cycle statistical feature and domain knowledge-based feature for battery RUL prediction. Specifically, a feature selection strategy is proposed to optimize feature sets that are most related to RUL of LIB. After feature extraction and feature fusion based on two Transformer encoders with parallel connection, an original RUL prediction result can be achieved. Then, the Bayesian optimization module is developed to boost the prediction performance and generalization ability of RUL prediction task. To demonstrate the superiority of proposed PFFN in the RUL prediction task, a series of experiments are conducted on A123 dataset with different data distributions. The experimental results demonstrate that the proposed PFFN not only achieves best prediction performance in terms of three evaluation metrics, but also exhibits a good generalization ability compared with other competitors, achieving high-precision prediction with 6.00%~27.61%, 0.58%~6.49%, and 5.95%~7.03% reduction in RMSE, MAPE, and Score respectively. The future direction of this research includes the feature exploration related to battery RUL prediction and the balance between input cycles and RUL prediction performance.

References

[1] Y. Wang, J. Tian, Z. Sun, L. Wang, R. Xu, M. Li, and Z. Chen, "A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems," *Renew. Sustain. Energy Rev.*, vol. 131, Oct. 2020.

[2] Y. Zhang, R. Xiong, H. He and W. Shen, "Lithium-ion battery pack state of charge and state of energy estimation algorithms using a hardware-in-the-loop validation," *IEEE Trans. Power Electron.*, vol. 32, no. 6, pp. 4421-4431, Jun. 2017.

[3] X. Hu, L. Xu, X. Lin, and M. Pecht, "Battery lifetime prognostics," *Joule*, vol. 4, pp. 310-346, Feb. 2020.

[4] P. M. Attia et al., "Review—'Knees' in lithium-ion battery aging trajectories," *J. Electrochem. Soc.*, vol. 169, no. 6, Jan. 2022.

[5] L. Liao and F. Köttig, "Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction," *IEEE Trans. Reliab.*, vol. 63, no. 1, pp. 191-207, Mar. 2014.

[6] M.S. H. Lipu, M.A. Hannan, A. Hussain, M.M. Hoque, P. J. Ker, M.H.M. Saad, and A. Ayob, "A review of state of health and remaining useful life estimation methods for lithium-ion battery in electric vehicles: Challenges and recommendations," *J. Clean. Prod.*, vol. 205, pp. 115-133, Dec. 2018.

[7] P. Sharma, B. J. Bora, "A review of modern machine learning techniques in the prediction of remaining useful life of lithium-ion batteries," *Batteries*, vol. 9, no. 13, Dec. 2023.

[8] Q. Miao, L. Xie, H. Cui, W. Liang, and M. Pecht, "Remaining useful life prediction of lithium-ion battery with unscented particle filter technique," *Microelectron. Reliab.*, vol. 53, pp. 805-810, Jun. 2013.

[9] X. Zheng, and H. Fang, "An integrated unscented kalman filter and relevance vector regression approach for lithium-ion battery remaining useful life and short-term capacity prediction," *Reliab. Eng. Syst. Saf.*, vol. 144, pp. 74-82, Dec. 2015.

[10] D. Guo, G. Yang, G. Zhao, M. Yi, X. Feng, X. Han, L. Lu, and M. Ouyang. "Determination of the differential capacity of lithium-ion batteries by the deconvolution of electrochemical impedance spectra," *Energies*, vol.13, no. 4, Jan. 2020.

[11] C. Lyu, Q. Lai, T. Ge, H. Yu, L. Wang, and N. Ma, "A lead-acid battery's remaining useful life prediction by using electrochemical model in the particle filtering framework," *Energy*, vol. 120, pp. 975-984, Feb. 2017.

[12] H. Zhang, Q. Miao, X. Zhang, and Z. Liu, "An improved unscented particle filter approach for lithium-ion battery remaining useful life prediction," *Microelectron. Reliab.*, vol. 81, pp. 288-298, Feb. 2018.

[13] B. Mo, J. Yu, D. Tang, H. Liu, and J. Yu, "A remaining useful life prediction approach for lithium-ion batteries using kalman filter and an improved particle filter," in *Proc. 2016 IEEE Int. Conf. Progn. Health Manag. (ICPHM)*, Ottawa, ON, Canada, pp. 1-5, Jun. 2016.

[14] Y. Xing, E. W.M. Ma, K. Tsui, and M. Pecht, "An ensemble model for predicting the remaining useful performance of lithium-ion batteries," *Microelectron. Reliab.*, vol. 53, pp. 811-820, Jun. 2013.

[15] J. Zhang, Y. Jiang, X. Li, M. Huo, H. Luo, and S. Yin, "An adaptive remaining useful life prediction approach for single battery with unlabeled small sample data and parameter uncertainty," *Reliab. Eng. Syst. Saf.*, vol. 222, Jun. 2022.

[16] M. A. Patil, P. Tagade, K. S. Hariharan, S. M. Kolake, T. Song, T. Yeo, and S. Doo, "A novel multistage support vector machine based approach for li ion battery remaining useful life estimation," *Appl. Energy*, vol. 159, pp. 285-297, Dec. 2015.

[17] L. Chen, Y. Zhang, Y. Zheng, X. Li, and X. Zheng, "Remaining useful life prediction of lithium-ion battery with optimal input sequence selection and error compensation," *Neurocomputing*, vol. 414, pp. 245-254, Nov. 2020.

[18] Severson, K.A., Attia, P.M., Jin, N. et al. "Data-driven prediction of battery cycle life before capacity degradation," *Nat. Energy*, vol. 4, pp. 383–391, Mar. 2019.

[19] C. Weng, Y. Cui, J. Sun, and H. Peng, "On-board state of health monitoring of lithium-ion batteries using incremental capacity analysis with support vector regression," *J. Power Sources*, vol. 235, pp. 36-44, Aug. 2013.

[20] Y. Zhang, R. Xiong, H. He and M. G. Pecht, "Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 5695-5705, Jul. 2018.

[21] P. Ding, X. Liu, H. Li, et al. "Useful life prediction based on wavelet packet decomposition and two-dimensional convolutional neural network for lithium-ion batteries," *Renew. Sust. Energy Rev.*, vol. 148, pp. 111287, Sept., 2021.

[22] D. Chen, W. Hong, and X. Zhou, "Transformer network for remaining useful life prediction of lithium-ion batteries," *IEEE Access*, vol. 10, pp. 19621-19628, Feb. 2022.

[23] R. R. Ardeshiri, and C. Ma, "Multivariate gated recurrent unit for battery remaining useful life prediction: A deep learning approach," *Int. J. Energy Res.*, vol. 45, no. 11, pp. 16633–16648, Sept. 2021.

[24] P. Li, Z. Zhang, Q. Xiong, B. Ding, J. Hou, D. Luo, Y. Rong, and S. Li, "State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network," *J. Power Sources*, vol. 459, May 2020.

[25] G. Ma, Y. Zhang, C. Cheng, B. Zhou, and Y. Yuan, "Remaining useful life prediction of lithium-ion batteries based on false nearest neighbors and a hybrid neural network," *Appl. Energy*, vol. 253, Nov. 2019.

[26] Q. Xu, M. Wu, E. Khoo, Z. Chen and X. Li, "A hybrid ensemble deep learning approach for early prediction of battery remaining useful life," *IEEE/CAA J. Autom. Sin.*, vol. 10, no. 1, pp. 177-187, Jan. 2023.

[27] S. Zhang, B. Zhai, X. Guo, K. Wang, N. Peng, and X. Zhang, "Synchronous estimation of state of health and remaining useful lifetime for lithium-ion battery using the incremental capacity and artificial neural networks," *J. Energy Storage*, vol. 26, Dec. 2019.

[28] F. Wang, Z. Zhao, J. Ren, Z. Zhai, S. Wang, and X. Chen, "A transferable lithium-ion battery remaining useful life prediction method from cycle-consistency of degradation trend," *J. Power Sources*, vol. 521, Feb. 2022.

[29] M. Zhu, Q. Ouyang, Y. Wan and Z. Wang, "Remaining useful life prediction of lithium-ion batteries: A hybrid approach of grey–markov chain model and improved gaussian process," *IEEE J. Emerg. Sel. Top. Power Electron.*, vol. 11, no. 1, pp. 143-153, Feb. 2023.

[30] J. Zha, Y. Zhu, B. Zhang, M. Liu, J. Wang, C. Liu, X. Hao, "Review of State Estimation and Remaining Useful Life Prediction Methods for Lithium–Ion Batteries," *Sustainability*, vol. 15, no. 6, Mar. 2023.
[31] M. I. Karmawijaya, I. N. Haq, E. Leksono and A. Widyotriatmo, "Development of Remaining Useful Life (RUL) Prediction of Lithium-ion Battery Using Genetic Algorithm-Deep Learning Neural Network (GADNN) Hybrid Model," *2022 7th International Conference on Electric Vehicular Technology (ICEVT)*, Bali, Indonesia, pp. 13-19, Sept. 2022.
[32] J. Qu, F. Liu, Y. Ma and J. Fan, "A Neural-Network-Based Method for RUL Prediction and SOH Monitoring of Lithium-Ion Battery," *IEEE Access*, vol. 7, pp. 87178-87191, Jun. 2019.
[33] M. Xu, X. Zhao, J. Wang, et al. "DFFNDDS: prediction of synergistic drug combinations with dual feature fusion networks," *J. Cheminform.*, vol. 15, no. 1, pp. 33-44, Mar. 2023.
[34] F. Zhang, B. Wang, Z. Gong, Z. Qin, Y. Yin, T. Guo, F. Wang, B. Zu, and K. Jiao, "Short-term performance degradation prediction of proton exchange membrane fuel cell based on discrete wavelet transform and gaussian process regression," *Next Energy*, vol 1, Sept. 2023.
[35] J. Lee, H. Sun, Y. Liu, and X. Li, "A machine learning framework for remaining useful lifetime prediction of li-ion batteries using diverse neural networks," *Energy and AI*, vol. 15, Jan. 2024.
[36] L. Song, D. Wang, X. Liu, A. Yin, and Z. Long, "Prediction of mechanical properties of composite materials using multimodal fusion learning," *Sens. Actuators A: Phys.*, vol. 358, Aug. 2023.