Journal of Energy Storage 84 (2024) 110780

Contents lists available at ScienceDirect

Journal of Energy Storage

journal homepage: www.elsevier.com/locate/est

Research papers

AttMoE: Attention with Mixture of Experts for remaining useful life prediction of lithium-ion batteries

Daoquan Chen [a], Xiuze Zhou [b,*]

[a] School of Intelligent Transportation, Zhejiang Institute of Mechanical and Electrical Engineering, Hangzhou, China
[b] Shuye Tech., Hangzhou, China

ARTICLE INFO

Keywords:
Li-ion battery
Remaining useful life
Attention mechanism
Mixture of experts
Neural network

ABSTRACT

For Lithium-ion (Li-ion) batteries, problems such as material aging and capacity decay lead to battery performance degradation or even catastrophic events. Predicting Remaining Useful Life (RUL) is an effective way to indicate the health of Li-ion batteries, which helps to improve the reliability and safety of battery-powered systems. We propose a novel neural network, AttMoE, which combines an attention mechanism with Mixture of Experts (MoE), to capture the capacity fade trend for battery RUL prediction. When facing the problem that raw data collected from sensors are always full of noise, AttMoE uses a dropout mask to denoise the raw data. For RUL prediction, one key idea is that the attention mechanism captures the long-term dependencies between elements in a sequence and more attention is paid to the important features that contain more degradation information; another key idea is that MoE uses many experts to increase model capacity to achieve better representations. Finally, we conducted experiments using two public data sets to show that AttMoE is effective in RUL prediction and achieves up to 10%–20% improvement in terms of Relative Error (RE). Our projects are all open source and are available at https://github.com/XiuzeZhou/RUL.

1. Introduction

As a portable source of energy, Lithium-ion (Li-ion) batteries have been broadly used in transportation, aerospace, and defense military applications [1–3]. Usually with increasing battery usage, their capacity is reduced. Failure of Li-ion batteries can lead to performance degradation, increased maintenance costs, and even catastrophic device failure [4–7]. To fully use the advantages of Li-ion batteries and prevent them from causing catastrophic damage to human safety, it is necessary to monitor the states accurately and take maintenance measures before the failure threshold is reached [8,9].

The prediction of accurate Remaining Useful Life (RUL) effectively indicates the health of Li-ion batteries and helps provide maintenance plans to ensure the reliability and safety of the system [10–12]. Therefore, for reliable and accurate battery RUL prediction, it is important to develop methods that are divided into the following two typical categories: model-based and data-driven. Model-based methods build mathematical functions to reflect the physical and electrochemical properties of batteries [13–15]. Data-driven methods model historical data, without involving any physical properties, to describe the degradation evolution of batteries [16–18]. Because of this property, data-driven methods have been receiving more and more attention recently.

State Of Health (SOH) estimation is a critical aspect of battery performance prediction, because it indicates the overall health and degradation level of a battery. Over the past decade, machine learning based data-driven methods have achieved great success in many applications, such as Computer Vision (CV) [19–21], Natural Language Processing (NLP) [22–24], recommendation systems [25–27], and medical diagnosis [28–30]. Machine learning models analyze time series data and extract degradation patterns to estimate the remaining capacity [31,32].

Machine learning algorithms offer the ability to extract valuable insights from complex and high-dimensional time series data, thereby enabling accurate prediction, optimization, and control of battery performance. To simulate the trend in decay of a battery, many researchers have developed automated solutions with machine learning techniques. For example, to model battery degradation, by using online learning techniques, Liu et al. [33] developed the Relevance Vector Machine (RVM). To model battery degradation, Nuhic et al. [34] proposed applying the Support Vector Machine (SVM) to analyze capacity sequence. However, these traditional methods have limited ability to learn nonlinear features.

Deep learning has a powerful ability to learn nonlinear representation from raw data [35–39]. The application of deep learning for

* Corresponding author.
E-mail addresses: chendaoquan@zime.edu.cn (D. Chen), zhouxiuze@foxmail.com (X. Zhou).

https://doi.org/10.1016/j.est.2024.110780
Received 1 September 2023; Received in revised form 13 November 2023; Accepted 30 January 2024
Available online 7 February 2024

Nomenclature

$CALCE$	Center for Advanced Life Cycle Engineering
ECL	Equivalent Circle Life
EOL	End Of Life
MAE	Mean Absolute Error
MoE	Mixture of Experts
MSE	Mean Square Error
RE	Relative Error
$RMSE$	Root Mean Square Error
RUL	Remaining Useful Life
SOH	State Of Health
$\hat{y}_t$	Predicted value of model
c	Input sequence of capacity
C_k	Battery capacity in the cycle k
C_o	Initial capacity
m	Size of sliding window
n	Length of input sequence
SOH_k	SOH in the cycle k
x_c	Normalized sequence
x_t	Input of the network
y_t	Output of the network

time series analysis in battery systems has gained significant attention in recent years. To establish the connection between RUL and a charge curve, Wu et al. [1] applied a Multi-Layer Perceptron (MLP) to model both the terminal voltage curve and charge process. To assess battery reliability, Ding et al. [40] integrated Convolutional Neural Network (CNN) and wavelet packet decomposition by learning long-term dependencies of capacity.

When making predictions for RUL, Recurrent Neural Network (RNN) based networks have demonstrated their effectiveness in dealing with sequential data. For example, to learn about changes in capacities effectively, RNN is proposed to simulate the intricate nonlinear trend associated with battery degradation [41–43]. To evaluate SOH, LSTM is used to model nonlinear capacity curves [44–46]. Gate Recurrent Unit (GRU), which is developed from LSTM, is also used often in RUL prediction [47–49]. However, utilizing RNN-based networks for modeling sequences recurrently results in significant time costs during training and leads to performance degradation due to the challenges of long-term dependencies [50–52].

In practical applications, raw data collected by sensors is always full of noise [41,44,53]. Many existing methods directly feed raw data into models without a denoising step, which seriously affects the performance of the model [16,54]. In our model, a dropout mask is applied to reduce noise by randomly deleting some noisy points in sequences.

In modeling RUL, existing RNN-based models have some key disadvantages: limited parallelization, difficulty in capturing long-term dependencies, and lack of attention mechanism. To solve the problem, an attention mechanism is used to model the capacity fade trend. Attention networks, with the power of parallelism and effective long-range capture of dependencies, are designed to extract degradation features of time series. The attention mechanism captures intricate long-term dependencies among elements in a sequence regardless of their distance and allocates more attention to significant features that contain crucial degradation information.

Finally, to enhance overall modeling capability and improve accuracy, Mixture of Experts (MoE) is used to better learn representations. MoE is a powerful technique that involves the integration of multiple experts to enhance the capacity and performance of a machine learning system [55,56]. By incorporating a diverse set of experts, a broader range of patterns within the data is captured, leading to improved overall performance and more comprehensive representations.

2. Proposed method

2.1. Problem setting

Li-ion batteries are widely applied in all kinds of electronic devices. Their performance directly affects reliability and safety [57–59]. However, Li-ion batteries suffer from side reactions during operation, leading to the aging of materials and capacity fading [4–7]. To ensure safety, early prediction of RUL offers crucial insights into the maintenance and replacement requirements of the batteries [45,60,61]. By accurately forecasting RUL, remaining lifespan is obtained to enable proactive maintenance and timely replacements, which not only enhances safety, but also optimizes resource allocation and minimizes potential risks associated with battery failures or malfunctions.

In RUL prediction, capacity is broadly regarded as the health indicator of a battery to quantify degradation. A battery reaches its End Of Life (EOL) threshold when its capacity is reduced to seventy or eighty percent of the initial value [44,62]. RUL is defined as the amount of time remaining before system health falls below a pre-determined failure threshold [63–65], calculated as follows:

$$N_{RUL} = N_{EOL} - N_{ECL}, \tag{1}$$

where N_{RUL} denotes the cycle number of battery RUL; N_{EOL} denotes the cycle number when the battery reaches its EOL; and N_{ECL} denotes the Equivalent Circle Life (ECL).

SOH is an important indicator to reflect the performance of batteries [44,66]. SOH in the cycle k is described as follows:

$$SOH_k = C_k / C_0 \times 100\%, \tag{2}$$

where C_o is the initial capacity, and C_k is the battery capacity in the cycle k.

2.2. Architecture

To predict RUL, we propose a novel neural network, AttMoE, which combines attention networks with MoE, to capture capacity fade trends. In AttMoE, attention is responsible for extracting features from the capacity degradation; To provide RUL prediction, MoE combines the different extracted features. AttMoE consists of four parts: inputs and dropout, attention, MoE, and outputs. The framework is shown in Fig. 1.

2.2.1. Input and dropout

First, to mitigate the impact of input data distribution changes on neural networks, it is essential to normalize the data [67,68]. Data normalization ensures that data is consistently represented across different samples and minimizes the variations caused by differences in data distribution. By normalizing input data, neural networks become more robust and less sensitive to changes in data distribution, enabling them to generalize better and make more reliable predictions. Normalization also facilitates the convergence of the training and improves the overall efficiency and effectiveness of the neural network model [69,70]. An input sequence of capacity, $c = [c_1, c_2, \dots, c_n]$, is mapped to $(0, 1]$ by $x_c = c/C_0$.

Second, raw input data often contains much noise, particularly during charge/discharge regeneration, which has a detrimental effect on the accuracy of predictions. To ensure stability and robustness, it is crucial to denoise the input data before feeding it into deep neural networks. We use a dropout mask to process the normalized data, x_c:

$$x = dropout(x_c). \tag{3}$$

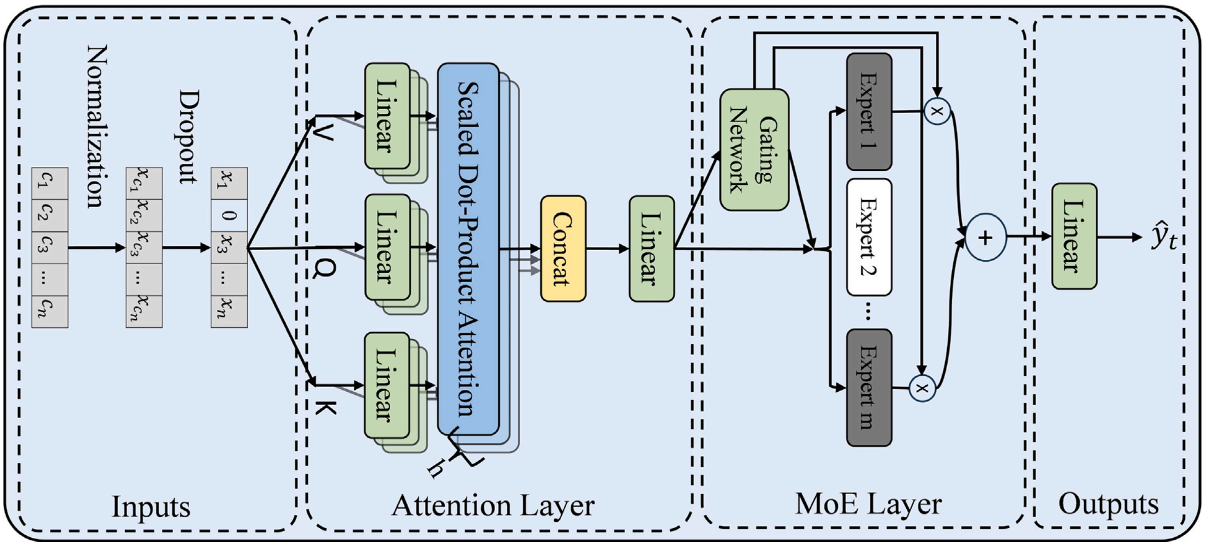


Fig. 1. AttMoE network for RUL prediction consists of four parts: inputs and dropout, attention, MoE, and outputs. Dropout is designed for denoising; Attention is designed to extract degradation features from time series; MoE is designed for improved overall performance and more comprehensive representations.

Third, a sliding window of size m, used to capture local patterns and dependencies within the time series, x, divides the time series into smaller segments:

$$x \xrightarrow{\text{sliding window}} \begin{array}{cc} x_t & y_t \\ \begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ x_2 & x_3 & \cdots & x_{m+1} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n-m} & x_{n-m+1} & \cdots & x_{n-1} \end{bmatrix} & \begin{bmatrix} x_{m+1} \\ x_{m+2} \\ \cdots \\ x_n \end{bmatrix} \end{array}, \tag{4}$$

where x_t and y_t denote the input and output of the network, respectively, and $t \in [1, 2, \dots, n-m]$.

2.2.2. Attention mechanism

In extracting degradation features from time series, attention networks capture, in parallel, intricate long-term dependencies, without being constrained by the spatial separation. A key advantage of attention networks lies in their ability to allocate more attention to the significant features that contain essential degradation information. By dynamically adjusting attention weights, the attention mechanism ensures that the most relevant and informative features receive heightened focus during the modeling process. This targeted attention allocation enables a network to effectively extract and emphasize the crucial degradation-related patterns and characteristics within the time series.

By combining the power of parallel processing, efficient long-range dependency capture, and selective attention allocation, attention networks offer a robust framework for extracting degradation features from time series. We enhance the ability to discern and prioritize the most important information, ultimately leading to improved performance and accuracy in degradation feature extraction. Let $Q = W_Q x_t$, $K = W_K x_t$, and $V = W_V x_t$ denote query, key, and value, respectively. A Multi-head attention with h heads and hidden size, d, is defined as follows:

$$MultiHead(Q, K, V) = Concat[head_1, \dots, head_h]W_O, \tag{5}$$

$$head_i = \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right), \tag{6}$$

$$\text{Attention}(q, k, v) = softmax\left(\frac{qk^T}{\sqrt{d_h}}\right)v, \tag{7}$$

where $d_h = d/h$ is used to avoid generating extremely small gradients.

2.2.3. Mixture of Experts

To enhance the capacity and performance of our model, MoE is used to integrate multiple experts. Each expert focuses on a specific subset or aspect of the data, allowing them to become proficient in capturing specific patterns or features. The MoE framework dynamically assigns weights to the experts, effectively blending their predictions to produce a more accurate and comprehensive output. By leveraging the collective knowledge and capabilities of multiple experts, MoE enables the model to handle complex tasks, capture diverse patterns, and achieve superior representations, leading to improved performance and robustness.

In an MoE model, the choice of the expert network depends on factors such as the nature of the data, the complexity of the problem, and the available resources. In our model, a fully connected layer is used as an expert, $E(x)$. A gated network, $G(x)$, is used to decide which expert to activate. The gated network consists of a full connection layer and softmax. MoE is defined as follows:

$$x_e = \sum_{i=1}^{m} G(x)_i E_i(x), \tag{8}$$

where m is the number of experts.

To ensure sparsity and balance, we follow [56] to use softmax of gated network as $G(x)$ as follows: Following the technique in [56], noisy Top-K gating is used to ensure sparsity and balance:

$$G(x) = Softmax(KeepTopK(H(x), k)), \tag{9}$$

$$H(x)_i = \left(x \cdot W_g\right)_i + StandardNormal() \cdot Softplus\left(\left(x \cdot W_{noise}\right)_i\right), \tag{10}$$

$$KeepTopK(u, k)_i = \begin{cases} u_i, & \text{if } u_i \text{ is in the top } k \text{ elements of } u. \\ -\infty, & \text{otherwise.} \end{cases} \tag{11}$$

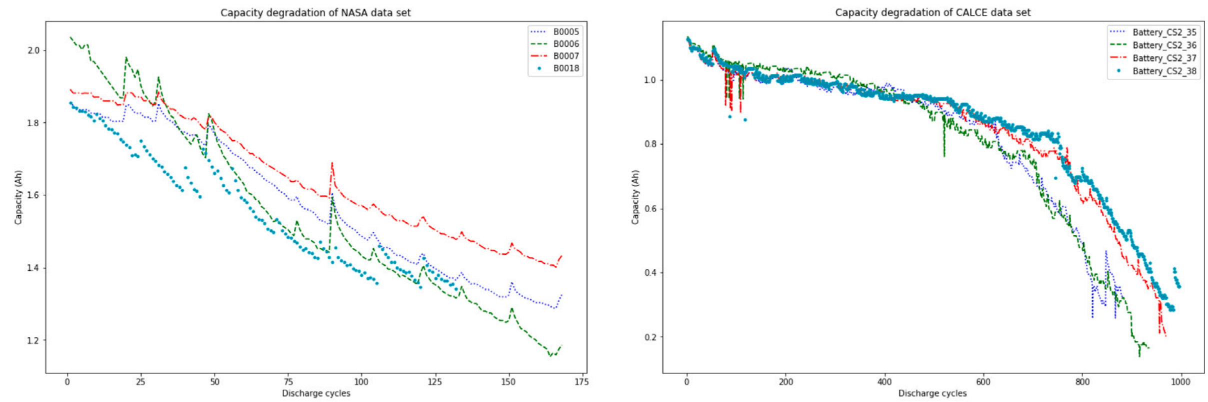
where W_g and W_{noise} are trainable weights.

2.2.4. Output

Finally, to predict unknown capacity, a fully connected layer is used to map the feature of MoE to the value of capacity. To predict the output, $\hat{y}$, prediction is defined as follows:

$$\hat{y} = f\left(W_p x_e + b_p\right), \tag{12}$$

where W_p, b_p, and $f(\cdot)$ denote weight, bias, and activation function of the output layer, respectively.



(a) Capacity degradation of NASA (b) Capacity degradation of CALCE

Fig. 2. Capacity degradation of both data sets: CALCE contains more noise than NASA.

2.3. Learning

To evaluate the loss, Mean Square Error (MSE), a widely adopted metric for regression that calculates the average of the squared differences between predicted and actual values, is selected as the loss function. MSE is defined as follows:

$$\mathcal{L} = \frac{1}{s}\sum_{t=1}^{s}\left(y_t - \hat{y}_t\right)^2 + \lambda\Omega(\Theta), \tag{13}$$

where s denotes the number of predicted points; λ denotes a regularization parameter; $\Omega(\cdot)$ denotes the regularization; Θ denotes the learning parameters of the model.

3. Experimental settings

3.1. Data sets

In our study, we performed experiments on two publicly available data sets: CALCE and NASA. Both data sets provide valuable insights into the behavior and performance of Li-ion batteries. The CALCE data set was obtained from the Center for Advanced Life Cycle Engineering (CALCE) at the University of Maryland[1] [71–73]. The NASA data set, accessible from the NASA Ames Research Center website,[2] consists of records from four distinct Li-ion batteries [74,75]. Each battery undergoes a sequence of three operations: charging, discharging, and impedance measurements. These operations are repeated multiple times, generating a comprehensive data set for in-depth analysis of battery behavior and performance. Capacity degradation of both data sets are shown in Fig. 2.

3.2. Baseline approaches

To verify the effectiveness of our proposed model, we compared our model with the following baseline approaches:

- **MLP** [1]. MLP is the most used in all kinds of tasks. With multiple fully connected layers, MLP learns the temporal trend of a battery. MLP has two key hyperparameters: learning rate and the number of hidden layer. Learning rate is set at 0.01, and the number of hidden layer is set at 2 and 4 for NASA and CALCE, respectively.
- **LSTM** [45]. LSTM, which incorporates memory cells to retain information from previous time steps in a recurrent manner, learns contexts of the data and long-term dependencies of time series. LSTM has two key hyperparameters: learning rate and the number of hidden layer. Learning rate is set at 0.001, and the number of hidden layer is set at 2 for both data sets.
- **GRU** [48]. GRU, an extension of LSTM, effectively and efficiently captures and retains relevant information in the sequential data. GRU has two key hyperparameters: learning rate and the number of hidden layer. Learning rate is set at 0.001, and the number of hidden layer is set at 2 for both data sets.
- **Dual-LSTM** [76]. Dual-LSTM uses two different LSTM cells to learn both short and long-term dependencies of input signals to predict RUL. Dual-LSTM has two key hyperparameters: learning rate and the number of hidden layer. Learning rate is set at 0.001, and the number of hidden layer is set at 2 for both data sets.

3.3. Evaluation metrics

To evaluate the performance of our model, we used three metrics: Relative Error (RE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE), defined as follows:

$$RE = \frac{\left|RUL^{pred} - RUL^{true}\right|}{RUL^{true}}$$

$$MAE = \frac{1}{s}\sum_{t=1}^{n}\left\|y_t - \hat{y}_t\right\|$$

$$RMSE = \sqrt{\frac{1}{s}\sum_{t=1}^{s}\left(y_t - \hat{y}_t\right)^2}$$

where RUL_{pred} and RUL_{true} denote the predicted and true RUL, respectively. The smaller the values of RE, MAE, and RMSE, the better the performance.

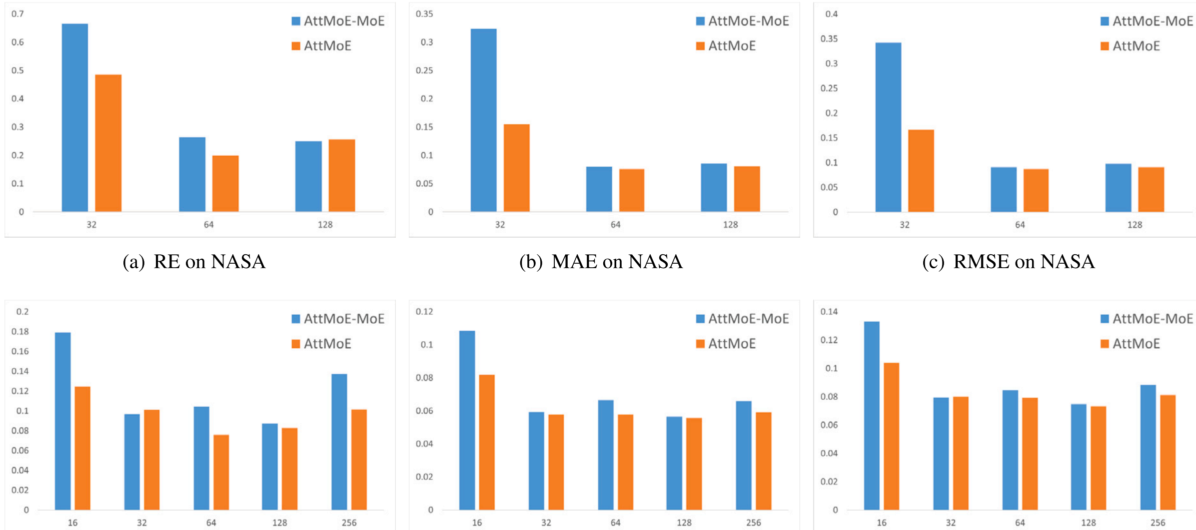
To assess the performance of all models, we used a leave-one-out evaluation method over all data as follows: each iteration, select one battery as a test sample and use the remaining batteries for training. After five iterations of this procedure, we determined the average score across all batteries.

3.4. Parameter settings

Beside learning rate and the number of hidden layer, all models have a key hyperparameter, sampler size of input series. Sampler size of an input series is set at approximately ten percent of the length of the input sequence, i.e., 16 and 64 for NASA and CALCE, respectively. Our model has four key hyperparameters: learning rate, the number of heads and hidden size of attention model, and the number of

[1] https://calce.umd.edu/data#CS2.

[2] https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository.



(a) RE on NASA (b) MAE on NASA (c) RMSE on NASA

(d) RE on CALCE (e) MAE on CALCE (f) RMSE on CALCE

Fig. 3. Effect of Mixture of Experts: compared AttMoE with its simplified version, AttMoE-MoE by varying the hidden size of the attention network.

Table 1
Overall performance of all methods in terms of RE, MAE, and RMSE. Best results are shown in bold.

Data sets	Metrics	MLP	LSTM	GRU	Dual-LSTM	AttMoE
NASA	RE	0.3851	0.2648	0.3044	0.2557	**0.2000**
	MAE	0.1379	0.0829	0.0806	0.0815	**0.0760**
	RMSE	0.1541	0.0905	0.0921	0.0879	**0.0872**
CALCE	RE	0.4018	0.0902	0.1319	0.0885	**0.0761**
	MAE	0.1557	0.0582	0.0671	0.0636	**0.0577**
	RMSE	0.2038	**0.0736**	0.0946	0.0874	0.0794

experts of MoE. In our experiments, the learning rate is chosen from $\{10^{-4}, 10^{-3}, 10^{-2}\}$; The number of heads is chosen from $\{2, 4, 8\}$; Hidden size is chosen from $\{32, 64, 128, 256\}$; The number of experts is chosen from $\{4, 8, 16\}$. Adam optimizer is used to optimize all models. All models are trained on our machine equipped with 128 GB RAM, and one GeForce RTX 3090 GPU (24G).

4. Results and discussion

4.1. Overall performance

First, we conducted experiments to validate performance. The scores for all methods are shown in Table 1. Best results are highlighted in bold.

From the results presented in Table 1, the following can be concluded: (1) AttMoE achieves the best results among all the methods, indicating that it effectively extracts valuable temporal patterns from capacity sequences in modeling RUL. (2) AttMoE consistently produces accurate predictions, which is an especially significant improvement on the NASA data set. This suggests that shorter sequences with limited information are difficult for models to obtain temporal patterns. (3) AttMoE and all RNN-based models exhibit better trend prediction than MLP, indicating the importance of incorporating sequential information for accurate RUL estimation. In AttMoE, the attention network simulates the overall trend by considering the influence of past capacities in the sequence. Therefore, AttMoE demonstrates the effectiveness of extracting meaningful temporal features for accurately predicting the RUL of a battery.

4.2. Effect of Mixture of Experts

Then, we studied the performance enhancement achieved through the use of MoE. To evaluate this improvement, we compared AttMoE with its simplified version, AttMoE-MoE, which does not incorporate a MoE layer. We conducted the comparison by varying the hidden size of the attention network and measuring the average scores. The results are illustrated in Fig. 3.

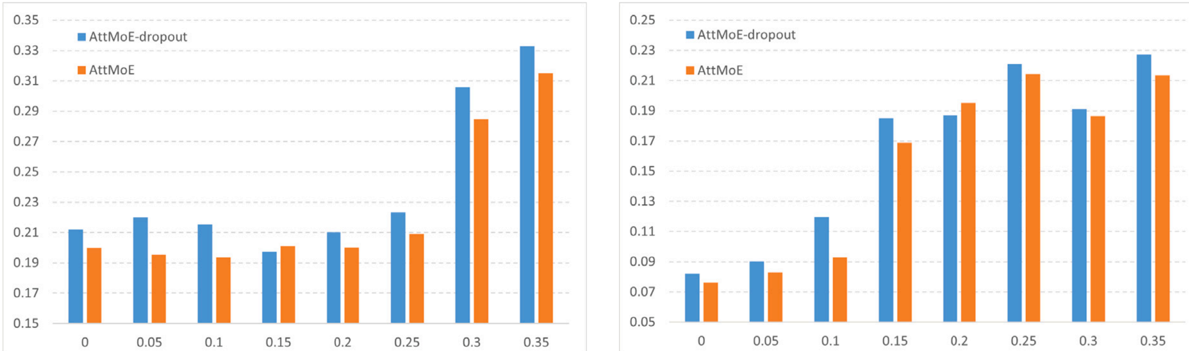
From Fig. 3, it is seen that, for all evaluation metrics, in most cases, AttMoE consistently outperformed its simplified version as the hidden size of attention increased, indicating that MoE contributes to improved performance in RUL prediction. Also, scores initially decrease and then become stable as the hidden size varies. This pattern suggests that AttMoE has a limited capacity to capture sufficient temporal information when the hidden size is too small. When the value is large, AttMoE becomes stable to learn enough temporal information by attention networks. As a result, in our model, MoE exhibits improvement over our method, leading to enhanced predictions.

4.3. Effect of dropout

Next, to assess the impact of the dropout mask, we conducted experiments by comparing AttMoE with the simplified version without dropout, AttMoE-dropout. In this experiment, we fixed the value of dropout at 0.1. Table 2 shows the average scores and the improvement of RE, MAE, and RMSE on the two data sets: relatively smooth NASA and relatively noisy CALCE.

From the results shown in Table 2, it is seen that in all cases, AttMoE performs better than AttMoE-dropout, indicating that the dropout mask is effective in improving a performance of the model. Also, compared with smooth NASA, noisy CALCE shows greater improvement, indicating that dropout has a larger impact on a time series with more noise and variation and more easily affects performance. These results provide valuable insight for optimizing the use of dropout and ensuring the robustness and effectiveness in noise reduction of data collected by sensors.

To examine further the denoising function of dropout, we added Gaussian noise to the raw data for sensitivity analysis. In this experiment, we selected a percentage from 5% to 35% in intervals of 5% and introduced noise to the selected percentage across all raw data. The scores obtained for RMSE are shown in Fig. 4.



(a) RE scores on NASA (b) RE scores on CALCE

Fig. 4. Effect of dropout: sensitivity analysis by introducing noise to the selected percentage across all raw data.

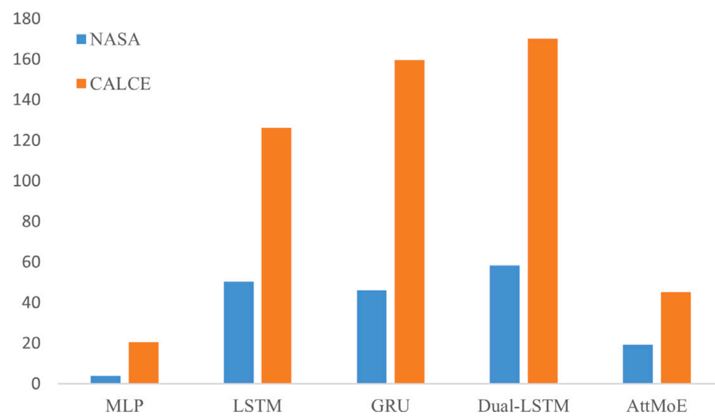


Fig. 5. Time cost (seconds) on two data sets.

Table 2
Effect of dropout: fixed the value of dropout at 0.1 to compare AttMoE with the simplified version without dropout, AttMoE-dropout.

Data sets	Metrics	AttMoE-dropout	AttMoE	Improvement
NASA	RE	0.2120	0.2000	6.00%
	MAE	0.0775	0.0760	1.94%
	RMSE	0.0885	0.0872	1.47%
CALCE	RE	0.0821	0.0761	7.31%
	MAE	0.0612	0.0577	5.72%
	RMSE	0.0836	0.0794	5.02%

From Fig. 4, it is seen that, with an increase of the proportion of noisy data, our model performs worse, indicating noisy data negatively impacts model performance. Consequently, noisy data limits or even damages the ability of a model to learn. On both data sets, in most cases, AttMoE generally outperforms AttMoE-dropout, which indicates the effectiveness of the dropout mask to enhance the performance of our model. Dropout mask is a reasonable solution to reduce noise by promoting model robustness and generalization.

4.4. Time cost

Finally, we conducted a study to analyze the time required to train different models using two datasets (See Fig. 5). From the results shown in Fig. 5, it is evident that, compared with our model, RNN-based models (LSTM, GRU, and Dual-LSTM) require significantly more time for training. This observation can be attributed to the fact that RNN-based networks, which handle sequences in a recurrent manner, result in higher time costs during training and inference. Using an attention network, AttMoE enhances the training efficiency of neural networks, thereby enabling it our model to capture sequential patterns effectively. In summary, our study concludes that with the application of an attention network, our model learning features in parallel are more suitable for predicting the RUL of a battery.

5. Conclusion

We proposed a novel network framework, AttMoE, for RUL prediction. In AttMoE, a dropout mask was applied to clear raw data by randomly removing capacities. To model the trend of capacity, an attention mechanism was used to extract features from the capacity degradation of batteries. Then, a MoE was applied to combine the different extracted features for better results. The experimental results demonstrate the improvement of up to 10%–20% achieved by our proposed model in terms of RE.

Although the proposed method has shown promising results, there are still many aspects that can be studied further. First, our model trained mainly on two data sets, leads to a limited ability for a wider application. By considering wider range of data sets, we aim to improve the accuracy and robustness of the RUL estimation for batteries. Also, the behavior of batteries is highly influenced by various factors, including temperature and current. By investigating the RUL estimation under different operating conditions, we aim to gain a comprehensive understanding of how these factors affect the remaining service life of batteries, which will enable us to develop more reliable and adaptable models for predicting RUL.

CRediT authorship contribution statement

Daoquan Chen: Methodology, Investigation, Data curation, Validation, Writing – original draft. **Xiuze Zhou:** Conceptualization, Methodology, Software, Validation, Writing – original draft, Review.

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

No data was used for the research described in the article.

Acknowledgments

The authors would like to thank Michael McAllister for proofreading this manuscript.

References

[1] Yi Wu, Wei Li, Youren Wang, Kai Zhang, Remaining useful life prediction of lithium-ion batteries using neural network and bat-based particle filter, IEEE Access 7 (2019) 54843–54854.

[2] Yuchen Song, Datong Liu, Chen Yang, Yu Peng, Data-driven hybrid remaining useful life estimation approach for spacecraft lithium-ion battery, Microelectron. Reliab. 75 (2017) 142–153.

[3] Xinyu Gu, KW See, Penghua Li, Kangheng Shan, Yunpeng Wang, Liang Zhao, Kai Chin Lim, Neng Zhang, A novel state-of-health estimation for the lithium-ion battery using a convolutional neural network and transformer model, Energy 262 (2023) 125501.

[4] Reza Rouhi Ardeshiri, Bharat Balagopal, Amro Alsabbagh, Chengbin Ma, Mo-Yuen Chow, Machine learning approaches in battery management systems: State of the art: Remaining useful life and fault detection, in: 2020 2nd IEEE International Conference on Industrial Electronics for Sustainable Energy Systems, IESES, Vol. 1, IEEE, 2020, pp. 61–66.

[5] Bin Duan, Qi Zhang, Fei Geng, Chenghui Zhang, Remaining useful life prediction of lithium-ion battery based on extended Kalman particle filter, Int. J. Energy Res. 44 (3) (2020) 1724–1734.

[6] Guijun Ma, Yong Zhang, Cheng Cheng, Beitong Zhou, Pengchao Hu, Ye Yuan, Remaining useful life prediction of lithium-ion batteries based on false nearest neighbors and a hybrid neural network, Appl. Energy 253 (2019) 113626.

[7] Guangquan Zhao, Guohui Zhang, Yuefeng Liu, Bin Zhang, Cong Hu, Lithium-ion battery remaining useful life prediction with deep belief network and relevance vector machine, in: 2017 IEEE International Conference on Prognostics and Health Management, ICPHM, IEEE, 2017, pp. 7–13.

[8] Huzaifa Rauf, Muhammad Khalid, Naveed Arshad, Machine learning in state of health and remaining useful life estimation: Theoretical and technological development in battery degradation modelling, Renew. Sustain. Energy Rev. 156 (2022) 111903.

[9] Beitong Zhou, Cheng Cheng, Guijun Ma, Yong Zhang, Remaining useful life prediction of lithium-ion battery based on attention mechanism with positional encoding, in: IOP Conference Series: Materials Science and Engineering, Vol. 895, IOP Publishing, 2020, 012006.

[10] Yongzhi Zhang, Rui Xiong, Hongwen He, Michael Pecht, Validation and verification of a hybrid method for remaining useful life prediction of lithium-ion batteries, J. Clean. Prod. 212 (2019) 240–249.

[11] Lijun Zhang, Zhongqiang Mu, Changyan Sun, Remaining useful life prediction for lithium-ion batteries based on exponential model and particle filter, Ieee Access 6 (2018) 17729–17740.

[12] Kamran Javed, Rafael Gouriveau, Noureddine Zerhouni, State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels, Mech. Syst. Signal Process. 94 (2017) 214–236.

[13] Adam Thelen, Meng Li, Chao Hu, Elena Bekyarova, Sergey Kalinin, Mohan Sanghadasa, Augmented model-based framework for battery remaining useful life prediction, Appl. Energy 324 (2022) 119624.

[14] Linxia Liao, Felix Köttig, A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction, Appl. Soft Comput. 44 (2016) 191–199.

[15] Yaguo Lei, Naipeng Li, Szymon Gontarz, Jing Lin, Stanislaw Radkowski, Jacek Dybala, A model-based method for remaining useful life prediction of machinery, IEEE Trans. Reliab. 65 (3) (2016) 1314–1326.

[16] Daoquan Chen, Weicong Hong, Xiuze Zhou, Transformer network for remaining useful life prediction of lithium-ion batteries, IEEE Access 10 (2022) 19621–19628.

[17] Kailong Liu, Yunlong Shang, Quan Ouyang, Widanalage Dhammika Widanage, A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery, IEEE Trans. Ind. Electron. 68 (4) (2020) 3170–3180.

[18] Lei Ren, Jiabao Dong, Xiaokang Wang, Zihao Meng, Li Zhao, M Jamal Deen, A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life, IEEE Trans. Ind. Inform. 17 (5) (2020) 3478–3487.

[19] Junyi Chai, Hao Zeng, Anming Li, Eric W.T. Ngai, Deep learning in computer vision: A critical review of emerging techniques and application scenarios, Mach. Learn. Appl. 6 (2021) 100134.

[20] Abdullah Ayub Khan, Asif Ali Laghari, Shafique Ahmed Awan, Machine learning in computer vision: a review, EAI Endorsed Trans. Scalable Inf. Syst. 8 (32) (2021) e4–e4.

[21] Elizabeth A Holm, Ryan Cohn, Nan Gao, Andrew R Kitahara, Thomas P Matson, Bo Lei, Srujana Rao Yarasi, Overview: Computer vision and machine learning for microstructural characterization and analysis, Metall. Mater. Trans. A 51 (2020) 5985–5999.

[22] Daniel W. Otter, Julian R. Medina, Jugal K. Kalita, A survey of the usages of deep learning for natural language processing, IEEE Trans. Neural Netw. Learn. Syst. 32 (2) (2020) 604–624.

[23] Fredrik Olsson, A literature survey of active machine learning in the context of natural language processing, 2009.

[24] Vera Sorin, Yiftach Barash, Eli Konen, Eyal Klang, Deep learning for natural language processing in radiology—fundamentals and a systematic review, J. Am. College Radiol. 17 (5) (2020) 639–648.

[25] Xiuze Zhou, Shunxiang Wu, Rating LDA model for collaborative filtering, Knowl.-Based Syst. 110 (2016) 135–143.

[26] Xiuze Zhou, Weibo Shu, Fan Lin, Beizhan Wang, Confidence-weighted bias model for online collaborative filtering, Appl. Soft Comput. 70 (2018) 1042–1053.

[27] Fan Lin, Xiuze Zhou, Wenhua Zeng, Sparse online learning for collaborative filtering, Int. J. Comput. Commun. Control 11 (2) (2016) 248–258.

[28] Mihalj Bakator, Dragica Radosav, Deep learning and medical diagnosis: A review of literature, Multimodal Technol. Interact. 2 (3) (2018) 47.

[29] Xiujin Wu, Wenhua Zeng, Fan Lin, Xiuze Zhou, NeuRank: Learning to rank with neural networks for drug–target interaction prediction, BMC Bioinformatics 22 (1) (2021) 1–17.

[30] Fei Wang, Lawrence Peter Casalino, Dhruv Khullar, Deep learning in medicine—promise, progress, and challenges, JAMA Intern. Med. 179 (3) (2019) 293–294.

[31] Mehdi Bagheri, Venera Nurmanova, Oveis Abedinia, Mohammad Salay Naderi, Noradin Ghadimi, Mehdi Salay Naderi, Renewable energy sources and battery forecasting effects in smart power system performance, Energies 12 (3) (2019) 373.

[32] Mehdi Bagheri, Venera Nurmanova, Oveis Abedinia, Mohammad Salay Naderi, Noradin Ghadimi, Mehdi Salay Naderi, Impacts of renewable energy sources by battery forecasting on smart power systems, in: 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe, EEEIC/I&CPS Europe, IEEE, 2018, pp. 1–6.

[33] Datong Liu, Jianbao Zhou, Haitao Liao, Yu Peng, Xiyuan Peng, A health indicator extraction and optimization framework for lithium-ion battery degradation modeling and prognostics, IEEE Trans. Syst. Man Cybern. Syst. 45 (6) (2015) 915–928.

[34] Adnan Nuhic, Tarik Terzimehic, Thomas Soczka-Guth, Michael Buchholz, Klaus Dietmayer, Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods, J. Power Sources 239 (2013) 680–688.

[35] Wei Zhang, Yuanguo Lin, Yong Liu, Huanyu You, Pengcheng Wu, Fan Lin, Xiuze Zhou, Self-supervised reinforcement learning with dual-reward for knowledge-aware recommendation, Appl. Soft Comput. 131 (2022) 109745.

[36] Ming Chen, Yunhao Li, Xiuze Zhou, Conet: Co-occurrence neural networks for recommendation, Future Gener. Comput. Syst. 124 (2021) 308–314.

[37] Ming Chen, Xiuze Zhou, DeepRank: Learning to rank with neural networks for recommendation, Knowl.-Based Syst. 209 (2020) 106478.

[38] Shengyu Lu, Hangping Chen, Xiuze Zhou, Beizhan Wang, Hongji Wang, Qingqi Hong, Graph-based collaborative filtering with MLP, Math. Probl. Eng. 2018 (2018).

[39] Ming Chen, Tianyi Ma, Xiuze Zhou, CoCNN: Co-occurrence CNN for recommendation, Expert Syst. Appl. (2022) 116595.

[40] Pan Ding, Xiaojuan Liu, Huiqin Li, Zequan Huang, Ke Zhang, Long Shao, Oveis Abedinia, Useful life prediction based on wavelet packet decomposition and two-dimensional convolutional neural network for lithium-ion batteries, Renew. Sustain. Energy Rev. 148 (2021) 111287.

[41] Narendhar Gugulothu, TV Vishnu, Pankaj Malhotra, Lovekesh Vig, Puneet Agarwal, Gautam Shroff, Predicting remaining useful life using time series embeddings based on recurrent neural networks, Int. J. Progn. Health Manage. 9 (1) (2018).

[42] Jie Liu, Abhinav Saxena, Kai Goebel, Bhaskar Saha, Wilson Wang, An adaptive recurrent neural network for remaining useful life prediction of lithium-ion batteries, in: Annual Conference of the Prognostics and Health Management Society, 2010, pp. 1–9.
[43] Marcantonio Catelani, Lorenzo Ciani, Romano Fantacci, Gabriele Patrizi, Benedetta Picano, Remaining useful life estimation for prognostics of lithium-ion batteries based on recurrent neural network, IEEE Trans. Instrum. Meas. 70 (2021) 1–11.
[44] Kyungnam Park, Yohwan Choi, Won Jae Choi, Hee-Yeon Ryu, Hongseok Kim, LSTM-based battery remaining useful life prediction with multi-channel charging profiles, IEEE Access 8 (2020) 20786–20798.
[45] Yongzhi Zhang, Rui Xiong, Hongwen He, Michael G. Pecht, Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries, IEEE Trans. Veh. Technol. 67 (7) (2018) 5695–5705.
[46] Asadullah Khalid, Aditya Sundararajan, Ipsita Acharya, Arif I. Sarwat, Prediction of li-ion battery state of charge using multilayer perceptron and long short-term memory models, in: 2019 IEEE Transportation Electrification Conference and Expo, ITEC, IEEE, 2019, pp. 1–6.
[47] Guorong Ding, Wenbo Wang, Ting Zhu, Remaining useful life prediction for lithium-ion batteries based on CS-VMD and GRU, IEEE Access 10 (2022) 89402–89413.
[48] Bin Xiao, Yonggui Liu, Bing Xiao, Accurate state-of-charge estimation approach for lithium-ion batteries by gated recurrent unit with ensemble optimizer, IEEE Access 7 (2019) 54192–54202.
[49] Yuchen Song, Lyu Li, Yu Peng, Datong Liu, Lithium-ion battery remaining useful life prediction based on GRU-RNN, in: 2018 12th International Conference on Reliability, Maintainability, and Safety, ICRMS, IEEE, 2018, pp. 317–322.
[50] Yu Mo, Qianhui Wu, Xiu Li, Biqing Huang, Remaining useful life estimation via transformer encoder enhanced by a gated convolutional unit, J. Intell. Manuf. (2021) 1–10.
[51] Jie Hao, Xing Wang, Baosong Yang, Longyue Wang, Jinfeng Zhang, Zhaopeng Tu, Modeling recurrence for transformer, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019, pp. 1198–1207.
[52] Elozino Egonmwan, Yllias Chali, Transformer and seq2seq model for paraphrase generation, in: Proceedings of the 3rd Workshop on Neural Generation and Translation, 2019, pp. 249–255.
[53] Taichun Qin, Shengkui Zeng, Jianbin Guo, Zakwan Skaf, A rest time-based prognostic framework for state of health estimation of lithium-ion batteries with regeneration phenomena, Energies 9 (11) (2016) 896.
[54] Wenbin Song, Di Wu, Weiming Shen, Benoit Boulet, A remaining useful life prediction method for lithium-ion battery based on temporal transformer network, Procedia Comput. Sci. 217 (2023) 1830–1838.
[55] Damai Dai, Li Dong, Shuming Ma, Bo Zheng, Zhifang Sui, Baobao Chang, Furu Wei, StableMoE: Stable routing strategy for mixture of experts, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022, pp. 7085–7095.
[56] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, Jeff Dean, Outrageously large neural networks: The sparsely-gated mixture-of-experts layer, 2017, arXiv preprint arXiv:1701.06538.
[57] Yi Li, Kailong Liu, Aoife M. Foley, Alana Zülke, Maitane Berecibar, Elise Nanini-Maury, Joeri Van Mierlo, Harry E. Hoster, Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review, Renew. Sustain. Energy Rev. 113 (2019) 109254.
[58] Zhizheng Zhang, Wen Song, Qiqiang Li, Dual-aspect self-attention based on transformer for remaining useful life prediction, IEEE Trans. Instrum. Meas. 71 (2022) 1–11.
[59] Foad H Gandoman, Joris Jaguemont, Shovon Goutam, Rahul Gopalakrishnan, Yousef Firouz, Theodoros Kalogiannis, Noshin Omar, Joeri Van Mierlo, Concept of reliability and safety assessment of lithium-ion batteries in electric vehicles: Basics, progress, and challenges, Appl. Energy 251 (2019) 113343.
[60] Hancheng Dong, Xiaoning Jin, Yangbing Lou, Changhong Wang, Lithium-ion battery state of health monitoring and remaining useful life prediction based on support vector regression-particle filter, J. Power Sources 271 (2014) 114–123.
[61] Shuai Zheng, Kosta Ristovski, Ahmed Farahat, Chetan Gupta, Long short-term memory network for remaining useful life estimation, in: 2017 IEEE International Conference on Prognostics and Health Management, ICPHM, IEEE, 2017, pp. 88–95.
[62] Kai Goebel, Bhaskar Saha, Abhinav Saxena, Jose R. Celaya, Jon P. Christophersen, Prognostics in battery health management, IEEE Instrum. Meas. Mag. 11 (4) (2008) 33–40.
[63] Abhinav Saxena, Jose Celaya, Edward Balaban, Kai Goebel, Bhaskar Saha, Sankalita Saha, Mark Schwabacher, Metrics for evaluating performance of prognostic techniques, in: 2008 International Conference on Prognostics and Health Management, IEEE, 2008, pp. 1–17.
[64] Concetta Semeraro, Mariateresa Caggiano, Abdul-Ghani Olabi, Michele Dassisti, Battery monitoring and prognostics optimization techniques: challenges and opportunities, Energy 255 (2022) 124538.
[65] KW See, Guofa Wang, Yong Zhang, Yunpeng Wang, Lingyu Meng, Xinyu Gu, Neng Zhang, KC Lim, L Zhao, Bin Xie, Critical review and functional safety of a battery management system for large-scale lithium-ion battery pack technologies, Int. J. Coal Sci. Technol. 9 (1) (2022) 36.
[66] Cunsong Wang, Ningyun Lu, Senlin Wang, Yuehua Cheng, Bin Jiang, Dynamic long short-term memory neural-network-based indirect remaining-useful-life prognosis for satellite lithium-ion battery, Appl. Sci. 8 (11) (2018) 2078.
[67] Parag C. Pendharkar, A computational study on the performance of artificial neural networks under changing structural design and data distribution, European J. Oper. Res. 138 (1) (2002) 155–177.
[68] Jimmy Lei Ba, Jamie Ryan Kiros, Geoffrey E. Hinton, Layer normalization, 2016, arXiv preprint arXiv:1607.06450.
[69] Tim Cooijmans, Nicolas Ballas, César Laurent, Çağlar Gülçehre, Aaron Courville, Recurrent batch normalization, in: International Conference on Learning Representations, 2016.
[70] Tim Salimans, Durk P. Kingma, Weight normalization: A simple reparameterization to accelerate training of deep neural networks, Adv. Neural Inf. Process. Syst. 29 (2016).
[71] Wei He, Nicholas Williard, Michael Osterman, Michael Pecht, Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method, J. Power Sources 196 (23) (2011) 10314–10321.
[72] Yinjiao Xing, Eden WM Ma, Kwok-Leung Tsui, Michael Pecht, An ensemble model for predicting the remaining useful performance of lithium-ion batteries, Microelectron. Reliab. 53 (6) (2013) 811–820.
[73] Nick Williard, Wei He, Michael Osterman, Michael Pecht, Comparative analysis of features for determining state of health in lithium-ion batteries, Int. J. Progn. Health Manage. 4 (1) (2013).
[74] Bhaskar Saha, Kai Goebel, Battery Data Set, NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA, 2008.
[75] Bhaskar Saha, Kai Goebel, Uncertainty management for diagnostics and prognostics of batteries using Bayesian techniques, in: Proceedings of the 29th IEEE Aerospace Conference, IEEE, 2008, pp. 1–8.
[76] Zunya Shi, Abdallah Chehade, A dual-LSTM framework combining change point detection and remaining useful life prediction, Reliab. Eng. Syst. Saf. 205 (2021) 107257.