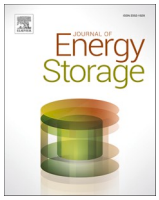
Journal of Energy Storage 62 (2023) 106903

Contents lists available at ScienceDirect

Journal of Energy Storage

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



Research papers

A deep attention-assisted and memory-augmented temporal convolutional network based model for rapid lithium-ion battery remaining useful life predictions with limited data

Zicheng Fei [a], Zijun Zhang [a,*], Fangfang Yang [b], Kwok-Leung Tsui [c]

[a] School of Data Science, City University of Hong Kong, Hong Kong SAR
[b] School of Intelligent Systems Engineering, Sun Yat-sen University, China
[c] Grado Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, United States

ARTICLE INFO

Keywords:
Lithium-ion batteries
Remaining useful life
Attention mechanism
Temporal convolutional network
Memory augmented module

ABSTRACT

Most of existing data-driven studies on lithium-ion battery remaining useful life (RUL) prediction consider a large scope of cyclic data over the entire battery life. Yet, applications of these models can be hindered due to restricted availability of such data in reality. This paper thus aims to study the battery RUL prediction from a new angle, predicting RUL via data collected from a limited number of incomplete cycles, i.e., 10 cycles, at any ageing stage. An advanced deep learning framework, the attention-assisted temporal convolutional memory-augmented network (ATCMN), is developed to realize an accurate and rapid battery RUL prediction under such challenging problem setting. To build an informative input based on limited data, a three-dimensional tensor input structure is first designed to integrate 10-cycle raw battery data including the time, capacity, and temperature dimensions obtained from the partial discharge process. To process such high dimensional input, the ATCMN first develops an attention module to automate weighting different battery parameters, time steps, and ageing cycles in the input. A temporal convolution module coordinating the dilated causal convolution and point-wise convolution is next developed to learn latent spatial-temporal feature representation from the weighted input. A memory-augmented module is further developed to enhance the latent feature representation through a reconstruction based on the historical information. Finally, the ATCMN employs a prediction module to derive nonlinear mappings from learned latent features to the battery RULs. A comprehensive computational study is conducted to verify the effectiveness of the ATCMN. Results report the higher accuracy and faster prediction of the ATCMN via benchmarking against a set of the state-of-the-art methods in the considered RUL predictions. Experimental results also show that the proposed ATCMN possesses better generalizability to different battery chemistries and operational conditions.

1. Introduction

Nowadays, lithium-ion batteries (LIBs) have been widely adopted as one of main means for the energy storage and power supply in many applications, such as portable electronics, electrical vehicles, renewable energy plants, etc., because of their high energy density, long cycle life, and low self-discharge rate [1]. Due to the recent advancement of data science principles, rapid LIB remaining useful life (RUL) predictions via data collected from a limited number of operational cycles bring a new horizon for accelerating the prognostics and diagnosis in the battery production and operation. For instance, the success of such technique can enable manufacturer to quickly estimate the battery RUL by operating just a few diagnostic cycles, thereby verifying the manufacturing process quality rapidly [2]. In addition, a rapid RUL prediction for electric vehicles and energy storage systems can help detect abnormal cells in advance so that unexpected failures of devices can be avoided [3,4]. Moreover, a rapid RUL prediction can enable a quick classification of retired LIBs into different secondary applications to improve the efficiency of the battery reutilization. Therefore, it is meaningful to develop advanced methods for accurate and rapid battery RUL predictions.

In the literature, methods for the battery RUL prediction can be divided into model-based methods and data-driven methods. Model-based methods have been mainly studied based on three types of

* Corresponding author.
E-mail address: zijzhang@cityu.edu.hk (Z. Zhang).

https://doi.org/10.1016/j.est.2023.106903
Received 29 June 2022; Received in revised form 17 January 2023; Accepted 13 February 2023
Available online 24 February 2023

models, the electrochemical models, equivalent circuit models, and empirical/semi-empirical models.

1) **Electrochemical model-based methods**. They established partial differential equations to approximate the complex electrochemical process in a battery, such as considering the solid-electrolyte interphase growth [5] and loss of active materials [6]. Due to the high computational complexity, these methods cannot appropriately serve online applications.
2) **Equivalent circuit model-based methods**. They utilized circuit components with empirical nonlinear parameters to describe battery degradation modes, such as the RC equivalent circuit models [7] and fractional-order equivalent circuit models [8]. However, several model parameters, such as the ohmic resistance, are difficult to measure, which makes them unsuitable for online applications.
3) **Empirical/semi-empirical model-based methods**. They first built mathematical models to characterize battery ageing patterns, such as the empirical exponential model [9], empirical ensemble model [10], semi-empirical Coulombic efficiency model [11], etc. Next, filtering techniques, such as the particle filter [12] and Kalman filter [13], were utilized to update model parameters for online RUL predictions. Although these methods can be operated online, the prediction accuracy relies heavily on the underlying battery degradation model.

In contrast, data-driven methods are mechanism-free and nonparametric, which possess a higher flexibility than model-based methods in modeling. Data-driven RUL prediction studies can be grouped into three main types in terms of the inference principle, the capacity degradation curve-based methods, handcrafted feature-based methods, and end-to-end deep learning methods.

1) **Capacity degradation curve-based methods**. Given the capacity data of first hundreds of cycles, these studies utilized autoregressive models to predict the future capacity degradation pattern of LIBs and obtain the RUL by extrapolating the predicted capacity to the end of life (EoL). Different data-driven models have been applied to predict the future capacity fading patterns. Reported ones include the relevance vector machine (RVM) [14], Gaussian process regression (GPR) [15], deep neural network based models, such as the recurrent neural network (RNN) and its famous variant - long short-term memory (LSTM) [16,17], as well as the hybrid network integrating the convolutional neural network (CNN) and LSTM [18].
2) **Handcrafted feature-based methods**. Based on various battery health features, these studies employed classical machine learning and recent deep learning techniques to build data-driven models for direct battery RUL predictions or estimating the battery state-of-health (SOH) to extrapolate the RUL. Many existing studies empirically developed features highly correlated to the battery SOH or RUL based on raw battery attributes. Presented handcrafted features can be briefly categorized into three groups: a) *features from original voltage/current/temperature curves*, e.g., the mean voltage falloff within a specific discharge time interval [19] and average surface temperature [20]; b) *features from processed voltage curves*, mainly the incremental capacity (IC) and differential voltage (DV) curves. For example, the position, height, and area of the IC peak [3] as well as the interval between two inflection points in the DV curve [21]; c) *features based on statistical metrics*, e.g., the variance of the difference between discharge capacity-voltage curves [2–4] and the sample entropy of a short voltage sequence [22]. Based on handcrafted features, data-driven models, such as the elastic net [2], support vector machine (SVM) [23], GPR [4], autoencoder [20], and CNN combined with LSTM [24], were next employed to develop mappings from these features to the target SOH or RUL.
3) **End-to-end deep learning methods**. Recently, a growing volume of studies presented attempts of applying deep network based models of various architectures to automatically learn latent features from raw battery data and then predict the battery RUL, which offered an end-to-end prediction process. The CNN [25], RNN [26], and their variants [27,28] were most frequently discussed in deep learning-based battery RUL prediction studies. Hong et al. [25] studied a dilated CNN to automatically extract features from the terminal voltage, current, and temperature data during the full charge-discharge process, and mapped them to the RUL. Han et al. [26] employed a LSTM with a domain adaption layer to learn features from the raw voltage, current, and temperature data during the full charge process, and next mapped them to the battery capacity. Based on the raw charge-discharge capacity, voltage, current, and temperature data, as well as a set of human-made features, Hsu et al. [27] developed a deep CNN architecture to learn the mapping from deeply learned features to the battery RUL.

Data-driven studies have reported promising results in the battery RUL prediction; Yet, there are still several limitations for existing methods. First, in most existing studies, more than 25% data of the entire battery life were generally required for the parameter estimating and model training so that accurate RUL predictions could be achieved [29]. Such high data access requirement forms an obstacle for applying these methods to realize rapid RUL predictions via limited online data. Although a few state-of-the-art studies [2,3] presented attempts of realizing accurate battery lifetime predictions via only the first-100-cycle data, they were limited to the early-stage degradation data and were not verified with data of any ageing stage in the entire battery life. In practical applications, it is common that LIBs in electrical devices may have gone through a number of ageing cycles, and an intelligent BMS is supposed to start the battery diagnosis at any cycles and to offer an accurate RUL prediction using data of lowest possible number of cycles. Thus, it is meaningful to further develop data-driven frameworks enabling an accurate and rapid online RUL prediction via battery data collected from a limited number of cycles at any ageing stage of the entire battery life. However, such challenging development has not received sufficient discussions in the literature because the information contained in a few operational cycles is quite limited for training data-driven models of sufficient quality. Innovation is thus desired at the battery data usage paradigm and advanced data-driven model development. One possible direction is to first design a more effective input structure properly integrating the information of multiple dimensions contained in data of limited cycles obtained at any battery ageing stage and next advance the deep model architecture for effectively processing such high dimensional input to extract data patterns depicting the battery degradation in latent space. Another unclear aspect of previous studies is the generalizability of reported performance as many verifications were conducted based on data of only one type of LIBs operated under a specific condition. Studying the robustness of methods towards different battery chemistries and operation conditions is so far insufficient in the literature. Moreover, most existing studies were conducted with assuming that data of the full charge and discharge process were accessible, which could be too ideal to real applications. Therefore, it is meaningful to further investigate methods working well when only the partial charge/discharge data are available.

To fill aforementioned research gaps, this paper presents a pioneering attempt of developing a more advanced deep learning model, the attention-assisted temporal convolutional memory-augmented network (ATCMN). The developed ATCMN aims to enable rapid battery RUL predictions based on battery data of only 10 incomplete cycles obtained at any battery ageing stage. In ATCMN, a tensor-like input structure is first developed to realize an organic fusion of the battery partial discharge process data from the time, capacity, and temperature information dimensions. Next, to extract the high-level latent and representative features as well as derive the mapping of inferring the RUL, the tensor-like input sequentially goes through four modules, the attention module, temporal convolution module, memory-augmented module,

and prediction module. The attention module consisting of the parameter-wise attention and cycle-wise attention aims to automate the weighting of different battery parameters as well as time steps and ageing cycles in the input so that input elements more important to the battery RUL prediction can be emphasized. The temporal convolution module then hierarchically extracts high-level spatial-temporal features from the input via coordinating the dilated causal convolution and point-wise convolution operations. A memory-augmented module is developed to further enhance the latent feature representation with a reconstruction based on historical information. Finally, the prediction module serves learning the nonlinear mapping from learned latent features to the battery RUL. Computational experiments are conducted on three datasets containing LIBs of different materials while tested under various conditions to verify the effectiveness and robustness of the proposed ATCMN. To further verify the superiority of the proposed ATCMN, a set of state-of-the-art benchmarks are considered and compared.

Major contributions of this research are summarized as follows:

1) *From the application perspective*, a deep learning powered framework is developed to realize the rapid battery RUL prediction with only data of 10 incomplete cycles obtained at any battery ageing stage, which is pioneering in the literature. This framework offers a great potential for facilitating the battery design optimization, online diagnosis, and recycling.
2) *From the technical perspective*, an innovative deep learning model ATCMN composed of four coordinated network modules, the attention module, temporal convolution module, memory-augmented module, and prediction module, for better feature engineering and more advanced modeling is proposed for the first time in the literature to tackle the rapid battery RUL prediction task with limited and incomplete cyclic data.
3) *From the performance improvement aspect*, the proposed framework offers the better generalization ability to different LIB materials and operation conditions. It also provides higher accuracy and faster prediction by comparing with existing state-of-the-art benchmarks.

2. Experimental Data

In this paper, three LIB datasets considering different chemical materials and test conditions are investigated to verify the performance of the proposed method. Two of them are publicly accessible while one is collected via experiments conducted in our laboratory. Detailed specifications of selected cell samples from three datasets are summarized in Table 1. Fig. 1 plots the capacity degradation curves of cell samples in each group. The battery lifetime corresponds to the cycle number with 20 % nominal capacity loss.

1) Dataset I: LFP battery

Dataset I is a public dataset provided by the Massachusetts Institute of Technology and Stanford University [2]. The dataset was collected from commercial lithium-iron phosphate (LFP)/graphite cells at an environmental temperature of 30 °C. The nominal capacity of cells is 1.1 Ah, and the upper and lower cutoff voltages are 3.6 V and 2.0 V, respectively. During cycling, the cells were charged with various fast-charging policies, denoted as "C1(S1%)-C2". That means, the cells were first charged with a constant current C1 until the SOC reached S1% and next charged with a constant current C2 until the SOC reached 80 %. Subsequently, the cells went through a 1C constant-current constant-voltage (CCCV) charging stage until reaching the upper cutoff voltage. During the discharge process, all cells were identically discharged with a constant current 4C until reaching the lower cutoff voltage. Signals of the terminal voltage, current, temperature, and capacity were continuously collected and real-time computed.

In Dataset I, the cell samples have a wide lifetime range varying from 300 cycles to 2300 cycles. Among them, six short-lived samples and six long-lived samples are chosen to evaluate and verify our proposed method. These selected samples are sorted as Group A/Group B and numbered as A1–A6/B1–B6, as listed in Table 1.

2) Dataset II: LCO battery

Dataset II is the widely utilized CS2 dataset provided by the Center for Advanced Life Cycle Engineering (CALCE) at the University of Maryland [30]. The dataset was collected from prismatic lithium cobalt oxide (LCO)/graphite cells at the room temperature 25 °C. The nominal capacity of cells is 1.1 Ah, and the upper and lower cutoff voltages are 4.2 V and 2.7 V, respectively. Three cell samples, CS2_35, CS2_36, and CS2_37, are selected, which are sorted as Group C and numbered as C1–C3, as listed in Table 1. In the charging process, these samples were charged with a 0.5C CCCV protocol. In the discharge process, they were discharged with a constant current 1C until reaching the lower cutoff voltage. Signals of the voltage, current and capacity were continuously collected and calculated.

3) Dataset III: NMC battery

Dataset III is collected from our own experiment platform as shown in Fig. 2. Cycling tests were conducted on three cylindrical Li(NiMnCo)O2 (NMC)/carbon cells at an ambient temperature of 25°C, using Arbin BT2000 tester. The nominal capacity of cells is 1.35 Ah, and the upper

Table 1
Specifications of different groups of battery samples.

		Cell no.	Lifetime	Charge policy	Discharge rate	Nominal capacity	Upper/lower cutoff voltage
Dataset I: LFP battery	Group A	A1	495	'4.8C(80%)-4.8C'	4C	1.1 Ah	3.6/2 V
		A2	461	'4.8C(80 %)-4.8C'			
		A3	468	'4.9C(27 %)-4.75C'			
		A4	509	'4.9C(61 %)-4.5C'			
		A5	471	'4.8C(80 %)-4.8C'			
		A6	498	'4.9C(69 %)-4.25C'			
	Group B	B1	1009	'5C(67 %)-4C'			
		B2	1063	'5.3C(54 %)-4C'			
		B3	1048	'5.6C(19 %)-4.6C'			
		B4	1039	'5.3C(54 %)-4C'			
		B5	1078	'4.8C(80 %)-4.8C'			
		B6	1028	'5.6C(19 %)-4.6C'			
Dataset II: LCO battery	Group C	C1	596	0.5C	1C	1.1 Ah	4.2/2.7 V
		C2	538				
		C3	609				
Dataset III: NMC battery	Group D	D1	761	1C	1C	1.35 Ah	4.0/2.5 V
		D2	874				
		D3	733				

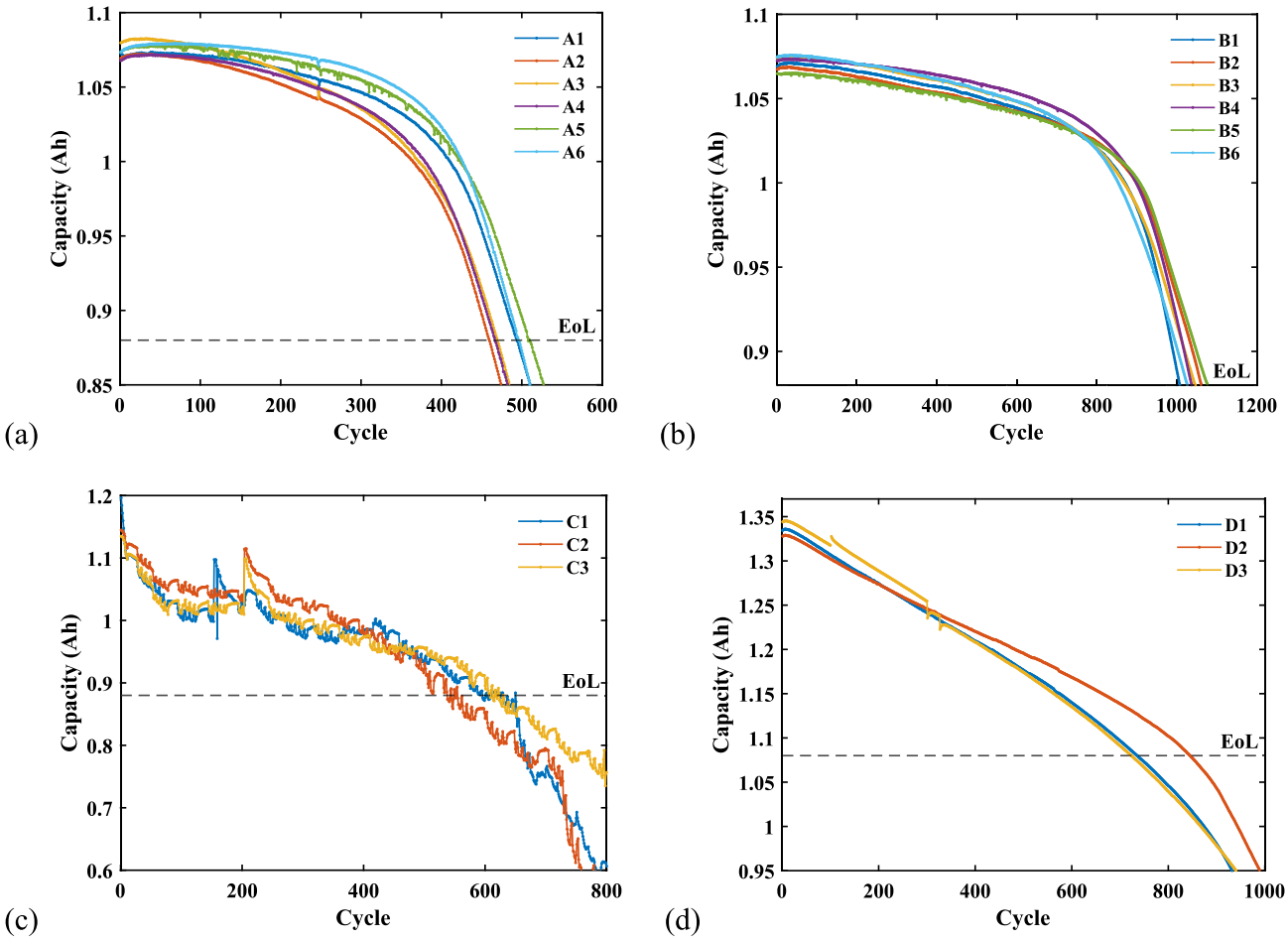


Fig. 1. Capacity degradation curves of cell samples: (a) Group A; (b) Group B; (c) Group C; (d) Group D.

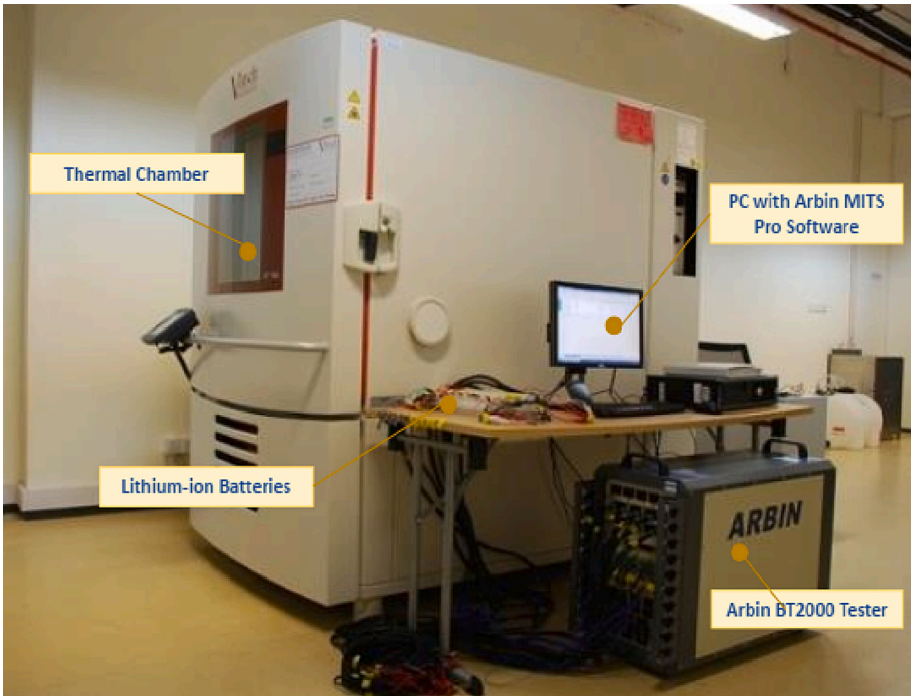


Fig. 2. Battery experimental platform.

and lower cutoff voltages are 4.0 V and 2.5 V, respectively. In each test cycle, the cells first underwent a 1C CCCV charge process and then a 1C constant current discharge process. Signals of the voltage, current and capacity were continuously collected and calculated. These samples are sorted as Group D and numbered as D1-D3, as listed in Table 1.

3. Problem Formulation and Methodology

3.1. Tensor data structure construction

In the online battery management system (BMS), real-time battery signals, such as the charging and discharging time (t), voltage (V), current (I), temperature (T), and capacity (Q) are continuously collected and computed. In this section, we analyze the temporal pattern of these signals and develop a novel data structure to effectively fuse the physical and temporal information in these signals. We will focus on the discharge process to allow tolerating the impact of largely varied charging conditions on the battery RUL in studied datasets.

Among these raw battery signals, V varies within a fixed range $[V_l, V_u]$, where V_l and V_u denote the lower and upper cutoff voltage. Considering V as the anchor, Fig. 3 plots curves of t, Q and T over V in different cycles, which are collected from a cell sample in Dataset I. As shown in Fig. 3, the spatial shape and location of t, Q, and T curves are varying as the battery ages. Specifically, curves of signals t and Q tend to move towards lower left, which is physically caused by the capacity loss due to the usage of LIBs. While the curve of T tends to move to upper right. This rising trend of T is mainly resulted from the increase of internal resistance as LIBs degrade. These variations of t, Q, and T can effectively reflect the battery degradation and RUL decline pattern. Therefore, to fully utilize such information, a three-dimensional tensor-like data structure $\boldsymbol{X}_i$ is developed as shown in Fig. 4. For the i-th cycle, $\boldsymbol{X}_i$ is expressed as (1).

$$\begin{aligned}\boldsymbol{X}_i &= \begin{bmatrix} \boldsymbol{T}(\boldsymbol{V})_{i-W+1}, \ldots, \boldsymbol{T}(\boldsymbol{V})_{i-1}, \boldsymbol{T}(\boldsymbol{V})_i \\ \boldsymbol{Q}(\boldsymbol{V})_{i-W+1}, \ldots, \boldsymbol{Q}(\boldsymbol{V})_{i-1}, \boldsymbol{Q}(\boldsymbol{V})_i \\ \boldsymbol{t}(\boldsymbol{V})_{i-W+1}, \ldots, \boldsymbol{t}(\boldsymbol{V})_{i-1}, \boldsymbol{t}(\boldsymbol{V})_i \end{bmatrix} \in R^{H\times W\times D}, \\ \boldsymbol{T}(\boldsymbol{V})_i &= \left(T_{i,1}, T_{i,2}, \ldots, T_{i,D}\right) \in R^D, \\ \boldsymbol{Q}(\boldsymbol{V})_i &= \left(Q_{i,1}, Q_{i,2}, \ldots, Q_{i,D}\right) \in R^D, \\ \boldsymbol{t}(\boldsymbol{V})_i &= \left(t_{i,1}, t_{i,2}, \ldots, t_{i,D}\right) \in R^D \end{aligned} \tag{1}$$

where H, W, D denote the height, width, and depth of the tensor. The data structure $\boldsymbol{X}_i$ integrates the physical, cycle-wise, and temporal information of the raw battery data into the height, width, and depth dimensions, respectively. Given that the incomplete discharging is quite common in real applications, in this study, a partial voltage range of $[V_l^p, V_u^p]$ is considered for each type of studied LIBs, which is determined through a series of comparative experiments. The partial voltage range $[V_l^p, V_u^p]$ is [2.5 V, 3.2 V], [3.1 V, 3.8 V], and [2.9 V, 3.6 V] for LFP, LCO, and NMC batteries, respectively. It should be noted that the relative t and Q are considered, which means that the discharge t and Q increase from zero at the upper voltage limit, $t_u^p = Q_u^p = 0$.

As shown in Fig. 4(a), the two-dimensional tabular forms of t, Q, and T signals are stacked, which result in a height dimension $H = 3$. The

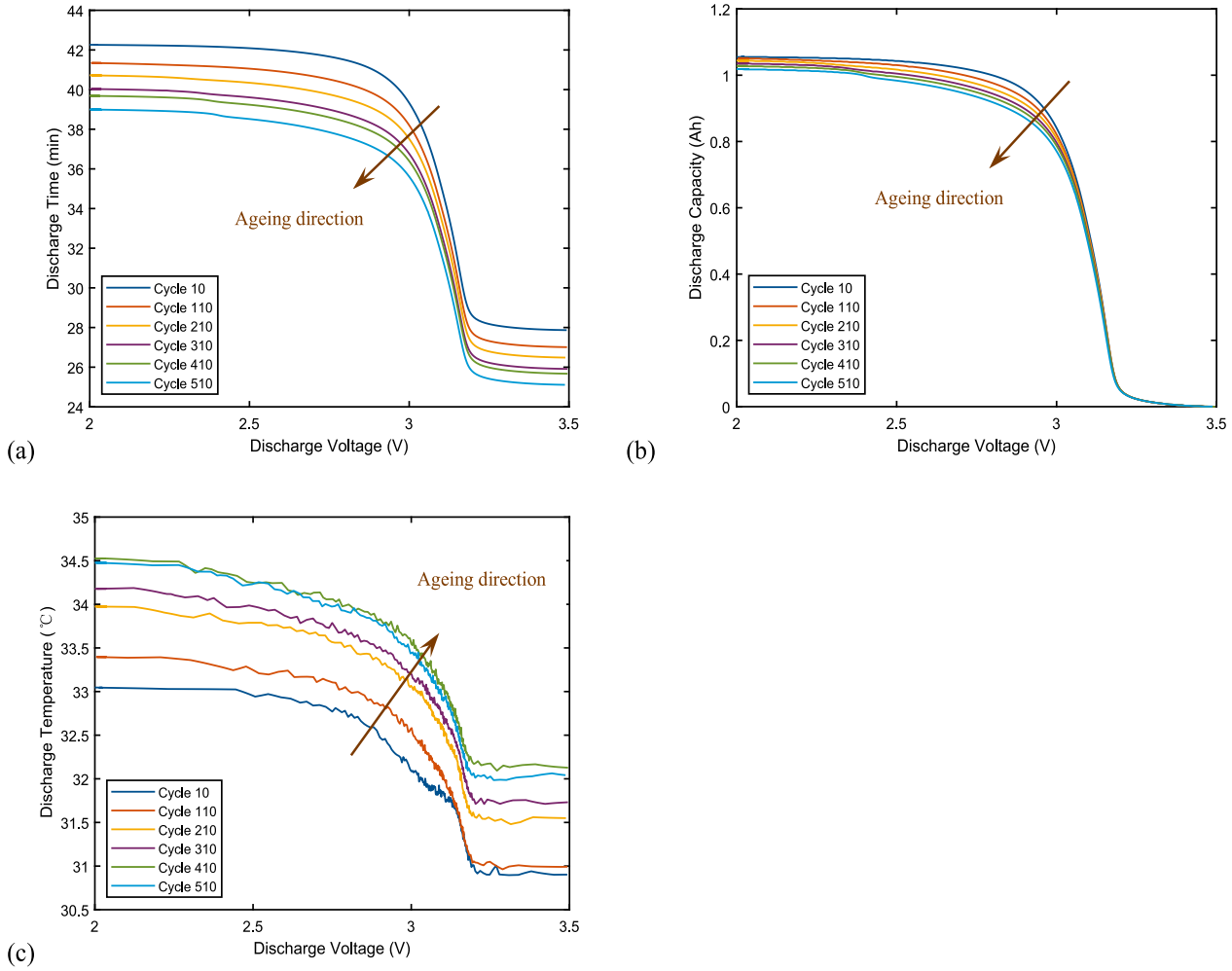


Fig. 3. Curves of (a) discharge time; (b) capacity; (c) temperature over discharge voltage at different ageing cycles, a cell sample in Dataset I.

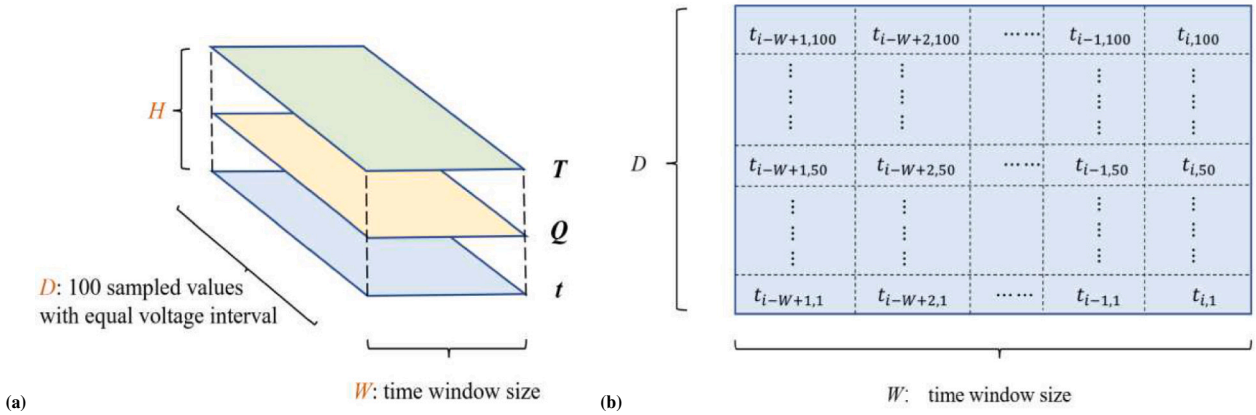


Fig. 4. (a) The input tensor structure; (b) Detail of the t map in the tensor.

width W of the tensor represents the size of the time window, which is set as 10 cycles after a number of preliminary trials. That means, for the i-th cycle, $\boldsymbol{X}_i = (\boldsymbol{x}_{i-9}, \boldsymbol{x}_{i-8}, \ldots, \boldsymbol{x}_{i-1}, \boldsymbol{x}_i)$, $\boldsymbol{x}_i = (\boldsymbol{t}_i, \boldsymbol{Q}_i, \boldsymbol{T}_i) \in R^{3\times D}$. The depth D of the tensor corresponds to the number of sampled values with an equal voltage interval in each cycle. Here, the value of D is set as 100. Fig. 4(b) shows the detail of the t map in the tensor. For example, $t_{i-W+1,\ 50}$ denotes the 50-th sampled t value at cycle $i - W + 1$.

3.2. Problem formulation and framework development

Given a training dataset $Data = \{(\boldsymbol{t}_i, \boldsymbol{Q}_i, \boldsymbol{T}_i, y_i)\}_{i=1}^{M}$, the goal of this paper is to learn the nonlinear mapping $f(\cdot)$ from raw battery signals $(\boldsymbol{t}_i, \boldsymbol{Q}_i, \boldsymbol{T}_i)$ to the RUL y_i, where M is the number of training samples. In Section 3.1, to fully fuse the local spatial and temporal information in $(\boldsymbol{t}, \boldsymbol{Q}, \boldsymbol{T})$, we reorganize them into a tensor data structure $\boldsymbol{X}_i$. Therefore, the task is transformed into learning the nonlinear mapping $F(\cdot)$ from $\boldsymbol{X}_i$ to the RUL y_i as described in (2). Fig. 5 demonstrates the graphical expression of our problem formulation.

$$y_i = F(\boldsymbol{X}_i) = F\begin{pmatrix} \boldsymbol{T}_{i-W+1}, \ldots, \boldsymbol{T}_{i-1}, \boldsymbol{T}_i \\ \boldsymbol{Q}_{i-W+1}, \ldots, \boldsymbol{Q}_{i-1}, \boldsymbol{Q}_i \\ \boldsymbol{t}_{i-W+1}, \ldots, \boldsymbol{t}_{i-1}, \boldsymbol{t}_i \end{pmatrix}, W = 10 \tag{2}$$

To effectively process the spatial and temporal information in $\boldsymbol{X}_i$, we develop a deep ATCMN framework to learn the mapping $F(\cdot)$. The flowchart of the proposed battery RUL prediction framework is illustrated in Fig. 6.

3.3. ATCMN-based battery RUL prediction model development

Fig. 7 demonstrates the overall architecture of the proposed ATCMN, which is composed of four network-based modules, the attention module, temporal convolution module, memory-augmented module, and prediction module. The attention module is developed to weight the importance of different time steps, battery parameters, and different cycles in the input tensor to the RUL prediction. The temporal convolution module is developed to capture the spatial and temporal patterns from the attention weighted input as well as to extract high-level latent and representative features. The memory-augmented module is developed to further enhance the latent spatial-temporal features via a reconstruction based on their historical information. Finally, the prediction module is employed to derive the mapping from deeply extracted features to the target battery RUL. Details of each module are sequentially elaborated next.

3.3.1. Attention module

Different battery parameters together with their values at different time steps and ageing cycles in the developed tensor-like input could contribute differently to the battery RUL prediction. To allow the ATCMN model to focus more on important elements in the input, a two-channel attention mechanism customized from the generic one discussed in [31] is employed to automatically assign weights on different elements in the tensor-like input. The two-channel attention mechanism conducts the parameter-wise and cycle-wise attention on the input sequentially.

The parameter-wise attention channel is designed to increase the

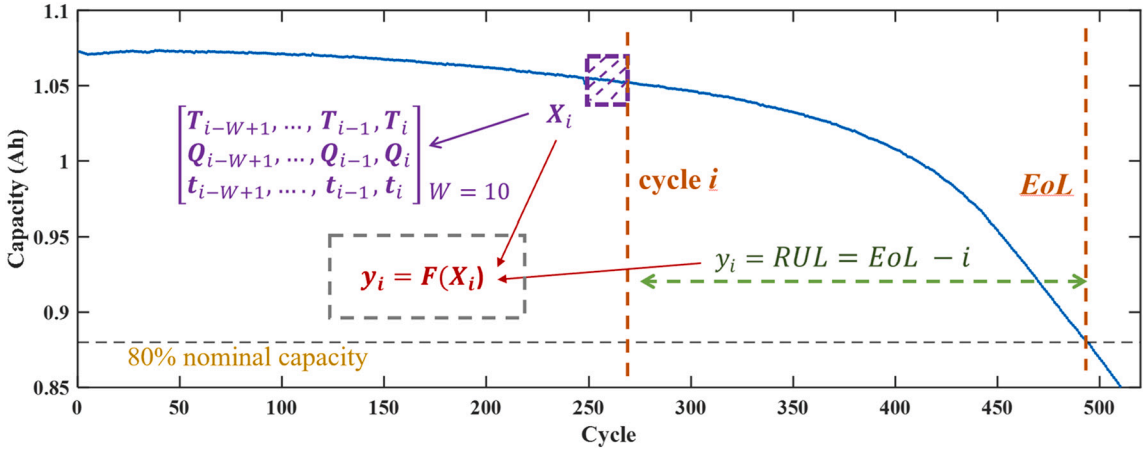


Fig. 5. Graphical expressions of the proposed problem formulation.

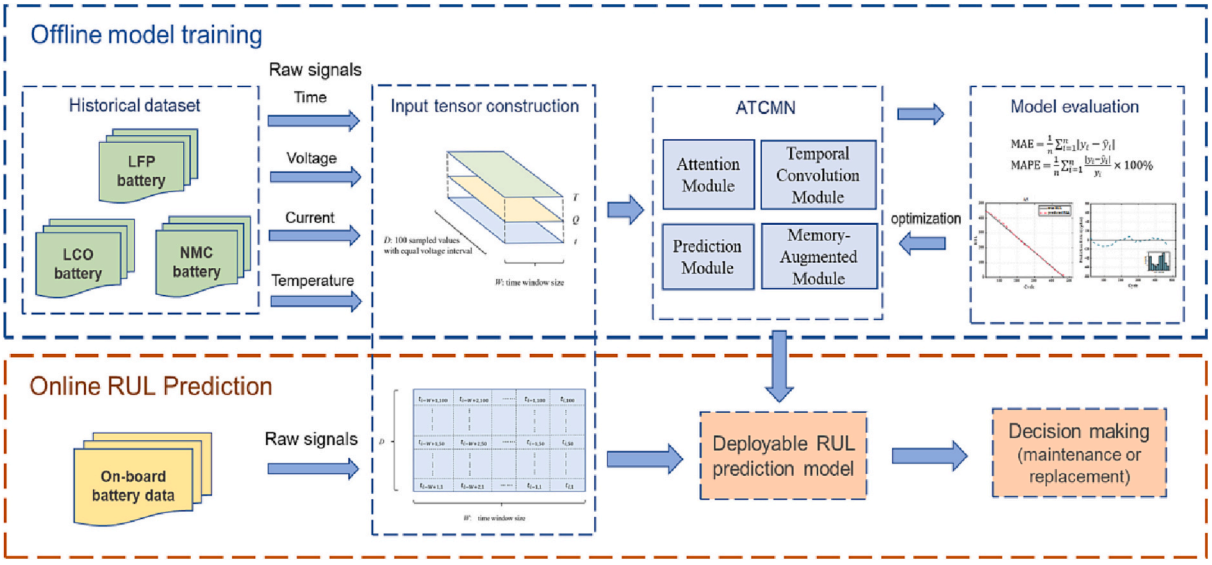


Fig. 6. The flowchart of the proposed framework for rapid online RUL predictions of LIBs.

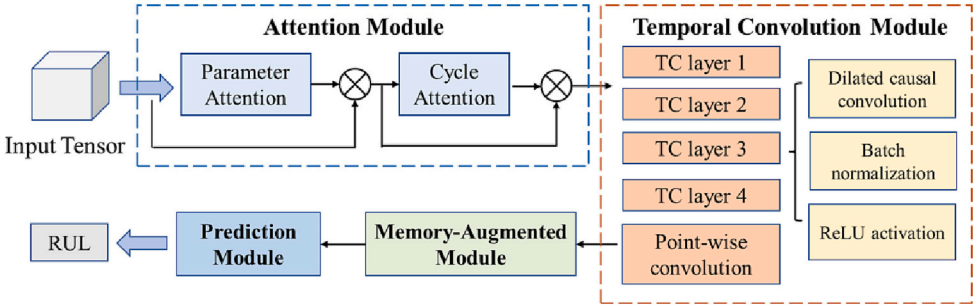


Fig. 7. Overall architecture of the proposed ATCMN.

model attention on more important battery parameters. Fig. 8(a) illustrates the structure of the parameter-wise attention. As shown in Fig. 8 (a), this process first performs both max pooling and average pooling operations to compress the input tensor over the width and depth dimensions (cycle dimension and time step dimension) and generates two one-dimensional vectors $M \in R^H$ and $A \in R^H$ as parameter-wise descriptors. The M and A are next fed into a shared multi-layer perceptron (MLP) with two fully connected layers. Outputs of the MLP are merged together and go through a sigmoid activation operation to generate the parameter attention map $\boldsymbol{PA} \in R^{1\times1\times H}$. The computation of $\boldsymbol{PA}$ is expressed in (3).

$$\begin{aligned}\mathbf{PA}(\mathbf{X}) &= \sigma(\mathrm{MLP}(\mathrm{MaxPooling}(\mathbf{X})) + \mathrm{MLP}(\mathrm{AvgPooling}(\mathbf{X}))\\ &= \sigma((W_2(W_1(M)+B_1)+B_2)+(W_2(W_1(A)+B_1)+B_2))\end{aligned} \tag{3}$$

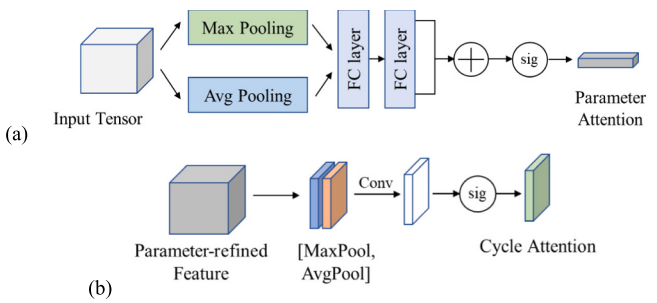


Fig. 8. (a) Parameter-wise attention; (b) cycle-wise attention.

where σ denotes the sigmoid operation, please refer to (4). The (W_1, W_2, B_1, B_2) are trainable parameters of the MLP. The number of hidden nodes is {20, 3} in the MLP.

$$\sigma(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x} \tag{4}$$

Finally, an element-wise multiplication is conducted between the input tensor $\boldsymbol{X}$ and the parameter attention map $\boldsymbol{PA}$ to generate the refined feature map $\boldsymbol{X}'$ with weighted battery parameters as depicted in (5).

$$\boldsymbol{X}' = \boldsymbol{PA} \otimes \boldsymbol{X} \tag{5}$$

After the parameter-wise attention, the cycle-wise attention channel processes the refined feature map $\boldsymbol{X}'$ and assigns importance weights to different time steps in different cycles. Fig. 8(b) shows the structure of the cycle-wise attention. As shown in Fig. 8(b), this process conducts max pooling and average pooling along the height dimension (battery parameter dimension) and generates two two-dimensional feature maps $F_{max} \in R^{D\times W}$ and $F_{avg} \in R^{D\times W}$, which are then concatenated to form a cycle/time information descriptor $F \in R^{D\times W\times 2}$. The F next goes through a convolution and a sigmoid operation, and finally outputs the cycle attention map $\boldsymbol{CA} \in R^{D\times W\times 1}$. The $\boldsymbol{CA}$ suggests which cycles and time steps are more beneficial to the battery RUL prediction. The calculation of $\boldsymbol{CA}$ is depicted in (6).

$$\begin{aligned}\mathbf{CA}(\mathbf{X}') &= \sigma(\mathrm{Conv}[\mathrm{MaxPooling}(\mathbf{X}');\mathrm{AvgPooling}(\mathbf{X}')])\\ &= \sigma\left(Conv\left[F_{max};F_{avg}\right]\right)\end{aligned} \tag{6}$$

where $Conv(\cdot)$ denotes a convolution operation with a filter size of 5×5. The cycle attention map $\boldsymbol{CA}$ is finally multiplied with the parameter-refined map $\boldsymbol{X}'$ to further emphasize contributions of more important cycles and time steps. The generated feature map $\boldsymbol{X}''$ as depicted in (7) is the input of the next module for further latent feature learning.

$$\boldsymbol{X}'' = \boldsymbol{CA} \otimes \boldsymbol{X}' \tag{7}$$

3.3.2. Temporal convolution module

In ATCMN, a TCN module is next developed to further extract latent features based on the attention-weighted tensor $\boldsymbol{X}''$. Through introducing causal and dilated convolution structures into the standard CNN, the TCN can better capture both the local spatial and temporal patterns from tensor inputs [32]. Fig. 9 illustrates the process of dilated causal convolution. In causal convolution, the output at time t is obtained by convolving entries of time t and earlier in the previous layer, which can be denoted as $Cau(\cdot)$ in (8):

$$y_i = Cau\left(\boldsymbol{X}_i''\right) = Cau\left(\boldsymbol{x}_{i-W+1}'', \boldsymbol{x}_{i-W+2}'' \ldots \boldsymbol{x}_{i-1}'', \boldsymbol{x}_i''\right) \tag{8}$$

Different from traditional convolutions, it does not consider the information of the future time. This is also consistent with the practical situation of our online RUL prediction task.

In dilated convolution, a fixed interval is introduced between every two adjacent entries, which can effectively expand the receptive field. As illustrated in Fig. 9, a dilation factor $d = 2$ means that the input entries with an interval of 2 are convolved together to generate the output of the next layer. The dilated causal convolution $Dil(\cdot)$ can be denoted as (9):

$$Dil\left(\boldsymbol{X}_i''; w, b\right) = \sum_{k=1}^{K} f_k \cdot \boldsymbol{x}_{i-(k-1)\cdot d}'' \tag{9}$$

where f is a convolutional kernel with length K, f_k is the k-th element in the kernel, (w, b) are weights and bias of the kernel that need to be learned. Here, d is growing exponentially with the number of layers, enabling a larger receptive field of the TCN.

After the dilated causal convolution, a batch normalization $BN(\cdot)$ as depicted in (10) is applied to standardize the convolution outputs.

$$BN(\boldsymbol{h}; \gamma, \beta) = \gamma \frac{\boldsymbol{h} - \widehat{E}(\boldsymbol{h})}{\sqrt{\widehat{Var}(\boldsymbol{h}) + \boldsymbol{\epsilon}}} + \beta \tag{10}$$

where $\boldsymbol{h} = Dil(\boldsymbol{X}_i'')$ is the convolutional output, $\widehat{E}(\boldsymbol{h})$ and $\widehat{Var}(\boldsymbol{h})$ are the mean and variance of $\boldsymbol{h}$, ϵ is a regularization parameter, while γ and β are two parameters that need to be learned. A ReLU activation function $R(\boldsymbol{p}) = \max(0, \boldsymbol{p})$ is next applied to tackle the gradient vanishing, where $\boldsymbol{p}$ is the normalized result.

The TCN module in the proposed ATCMN contains four temporal convolutional (TC) layers. Based on a large number of computational trials, we utilize {8, 16, 32, 64} convolution kernels with the kernel size $K = 2$ in each TC layer. The dilation factors $d = \{1, 2, 4, 8\}$ are considered. To maintain the full temporal information as the network goes deeper, we conduct zero padding in each TC layer to keep subsequent layers the same length W as the input layer.

Finally, inspired by the depth-wise separable convolution in the Xception network [33], we add a point-wise convolution operation at the end of the TCN module to further aggregate the learned high-level latent features over multiple channels. The point-wise convolution is expressed in (11).

$$\boldsymbol{U} = f_p \odot \boldsymbol{P} \tag{11}$$

where $\boldsymbol{P}$ is the feature map learned by TC layers. $\boldsymbol{U}$ denotes the output of the point-wise convolution, which is a one-channel aggregated feature map. f_p denotes a $1 \times 1 \times C$ filter, and C is the number of channels in $\boldsymbol{P}$.

3.3.3. Memory-augmented module

To better incorporate the historical degradation information into the RUL prediction, a memory-augmented module is next adopted to further process and reconstruct the latent feature representation $\boldsymbol{U}$ learned by the TCN module. A typical memory network [34] consists of four components, *In*, *Gen*, *Out*, and *Res*. The *In* denotes the input feature $\boldsymbol{U}$ learned from the previous module. The *Gen* denotes a generalization process for the memory unit updates during the model training. The *Out* denotes the output feature reconstructed by the memory network. The *Res* denotes the final prediction module that maps the output feature to the target RUL.

Fig. 10 illustrates the structure of the memory-augmented module utilized in the proposed ATCMN. Using the input feature $\boldsymbol{U}$ as a query, this module first compares it with a set of historical memory units to attain the most relevant units through the memory addressing operator and finally outputs the reconstructed feature representation $\boldsymbol{Z}$. The feature reconstruction process is depicted in (12).

$$\begin{aligned} \mathbf{Z} &= \sum\nolimits_{i=1}^{N} \mathrm{w}_i \mathbf{m}_i \\ w_i &= \frac{exp(d(\boldsymbol{U}, \boldsymbol{m}_i))}{\sum_{j=1}^{N} exp(d(\boldsymbol{U}, \boldsymbol{m}_j))} \\ dis(\boldsymbol{U}, \boldsymbol{m}_i) &= \frac{\boldsymbol{U}\boldsymbol{m}_i^T}{\|\boldsymbol{U}\|\|\boldsymbol{m}_i\|} \end{aligned} \tag{12}$$

where $\boldsymbol{m}_i$ is the i-th memory unit. N is a hyperparameter determining the number of memory units, which is set to 2000 in this study. $dis(\cdot, \cdot)$ denotes a cosine similarity measurement.

3.3.4. Prediction module

A prediction module is finally employed to learn the nonlinear mapping between the feature representation $\boldsymbol{Z}$ learned from previous modules and the battery RUL y. This module contains two fully connected (FC) layers. The mapping process can be mathematically

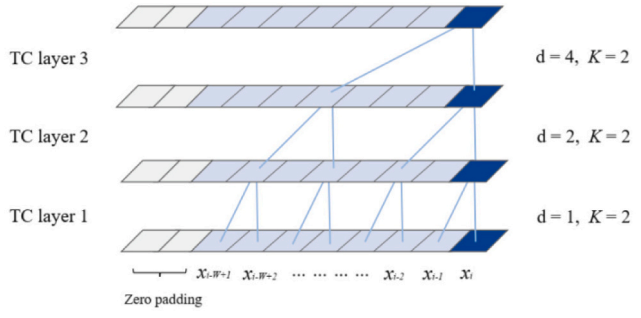


Fig. 9. Illustration of dilated causal convolutions.

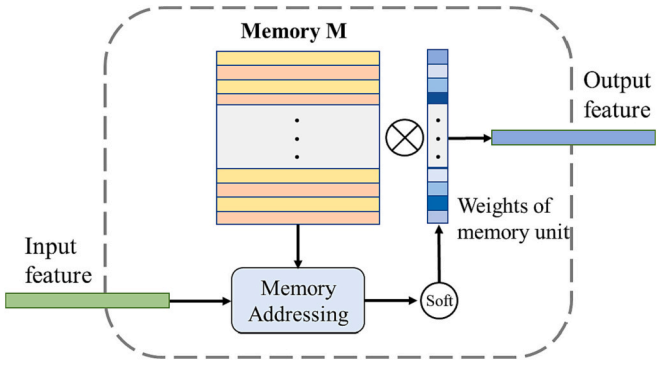


Fig. 10. Structure of the memory-augmented module.

expressed as (13).

$$\widehat{y} = g(\boldsymbol{Z}; \boldsymbol{W}, \boldsymbol{B}) = Dr(\boldsymbol{W} \times \boldsymbol{Z} + \boldsymbol{B}) \tag{13}$$

where $\widehat{y}$ is the predicted RUL, $(\boldsymbol{W}, \boldsymbol{B})$ are parameters of two FC layers that need to be learned, and $Dr(\cdot)$ is a drop out operation preventing the model from overfitting. In two FC layers, we utilize {50, 50} nodes and set the drop out probability to 0.2.

The whole ATCMN is trained via the loss function defined in (14).

$$Loss(\boldsymbol{X}, y; \theta) = \frac{1}{M} \sum_{m=1}^{M} (F_\theta(\boldsymbol{X}) - y)^2 - \alpha w log(w) \tag{14}$$

where M is the number of training samples, F denotes the mapping of the developed ATCMN, θ denotes the network parameters to be learned, and w is the weight of memory units. α is a hyperparameter, which is set to 0.0001 in this study.

Pseudo codes of the offline model development and online model implementation are offered in Algorithms 1 and 2, respectively.

4. Results and Discussion

In this section, computational experiments are conducted to evaluate the proposed method. In Sections 4.1 and 4.2, the performance of different battery parameters and different time window sizes are analyzed, respectively. In Section 4.3, the robustness of our framework is verified. In Section 4.4, the superiority of our method is validated by comparing with a set of benchmarks. In Section 4.5, an ablation study is conducted to explore the effect of handcrafted features and deeply learned features. Computational experiments are conducted on a server with three NVIDIA GeForce RTX 2080 GPUs, and a 64-bit ubuntu operating system. In Group A–D, the cyclic data of cell A1–A4, B1–B4, C1–C2, and D1–D2 are used as training samples in each group, while those of cell A5-A6, B5-B6, C3 and D3 are utilized as testing samples in each group. The MAE (mean absolute error) and MAPE (mean absolute percentage error) described in (15) and (16) are employed to evaluate the method performance:

$$\text{MAE} = \frac{1}{M} \sum_{i=1}^{M} |y_i - \widehat{y}_i| \tag{15}$$

$$\text{MAPE} = \frac{1}{M} \sum_{i=1}^{M} \frac{|y_i - \widehat{y}_i|}{y_i} \times 100\% \tag{16}$$

where M denotes the number of training cycles, y_i and $\widehat{y}_i$ denote the actual and predicted RUL, respectively. The MAE measures the absolute error between the true and predicted RUL, while the MAPE describes the relative error rate.

4.1. Effect of input battery parameters

To evaluate the value of considering various battery parameters on the RUL prediction, multiple combinations of raw battery parameters (t, Q and T) are considered as inputs. Experiments are first conducted on Group A and B of Dataset I since Datasets II and III do not provide records of the T parameter. For a fair comparison, the time window size is fixed as 10 cycles. Results are demonstrated in Table 2 and Fig. 11.

As shown in Table 2, for all testing cells, the best prediction performance (marked as bold) is achieved when the t, Q and T parameters are all considered. Besides, compared with considering only t in the input, better performance is achieved when an additional parameter is combined. Adding the Q parameter could bring more accuracy improvement than adding the T parameter, which implies that the Q parameter has a higher benefit on the RUL prediction than the T parameter. In addition, it is observable that, when Q and T parameters are considered in inputs, results of all tested cells are worse than those of considering only t in the input. This indicates the superiority of the t parameter than other two on the RUL prediction. In Fig. 11, it is observable that the proposed method can well capture the actual battery RUL attenuation trajectories based on information of only 10 partial cycles.

4.2. Effect of time window size

Different lengths of time window contain different amount of battery temporal information and will affect the prediction performance. To

Algorithm 1
Offline model development.

1: **Input**: Battery parameters (t, Q, T) and RUL y in historical datasets, time window size W, number of iterations N
2: Construct the tensor input:

$$\boldsymbol{X}_{\mathrm{i}} = \text{TensorInput}\begin{pmatrix} \boldsymbol{T}_{\mathrm{i-W+1}}, \ldots, \boldsymbol{T}_{\mathrm{i-1}}, \boldsymbol{T}_{\mathrm{i}} \\ \boldsymbol{Q}_{\mathrm{i-W+1}}, \ldots, \boldsymbol{Q}_{\mathrm{i-1}}, \boldsymbol{Q}_{\mathrm{i}} \\ \boldsymbol{t}_{\mathrm{i-W+1}}, \ldots, \boldsymbol{t}_{\mathrm{i-1}}, \boldsymbol{t}_{\mathrm{i}} \end{pmatrix}$$

3: Initialize the network parameters θ
4: **for** epoch $j = 1$ to N

$$\theta = \arg\min \text{Loss}(\boldsymbol{X}, \mathrm{y}; \theta)$$

Update parameters θ using back propogation and gradient-based optimizer

end for

5: **Output**: The fine-trained model F_θ

Algorithm 2
Online model implementation.

1: **Input**: Online battery parameters (t^*, Q^*, T^*)
2: Construct the tensor input:

$$\boldsymbol{X}_i^* = \text{TensorInput}\begin{pmatrix} \boldsymbol{T}^*_{\text{i}-\text{W}+1}, \ldots, \boldsymbol{T}^*_{\text{i}-1}, \boldsymbol{T}^*_{\text{i}} \\ \boldsymbol{Q}^*_{\text{i}-\text{W}+1}, \ldots, \boldsymbol{Q}^*_{\text{i}-1}, \boldsymbol{Q}^*_{\text{i}} \\ \boldsymbol{t}^*_{\text{i}-\text{W}+1}, \ldots, \boldsymbol{t}^*_{\text{i}-1}, \boldsymbol{t}^*_{\text{i}} \end{pmatrix}$$

3: Predict the battery RUL:

$$\boldsymbol{y}_i^* = \text{F}_\theta(\boldsymbol{X}_i^*)$$

4: **Output**: The predicted RUL $\boldsymbol{y}_i^*$

Table 2
RUL prediction results of different input parameter combinations.

Parameter combination			Group A				Group B			
			Cell A5		Cell A6		Cell B5		Cell B6	
t	*Q*	*T*	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
√			23	11.2 %	17	15.1 %	37	16.7 %	13	10.3 %
√		√	18	10.5 %	14	13.9 %	34	12.9 %	22	15.4 %
√	√		10	8.1 %	12	11.9 %	28	12.7 %	12	8.1 %
	√	√	27	14.6 %	21	17.5 %	39	21.3 %	31	21.7 %
√	√	√	**4**	**7.4 %**	**8**	**7.5 %**	**21**	**10.8 %**	**11**	**7.7 %**

evaluate the effect of the time window size on RUL prediction, comparative experiments are conducted using window sizes of {1, 5, 10, 15, 20, 30} cycles. Three parameters *t*, *Q* and *T* are all employed as inputs. Results on are summarized in Table 3.

As shown in Table 3, error metrics of the lowest values are marked as bold. For all test cells, the best RUL prediction is achieved when the time window size is set to 10 or 15 cycles. In general, a window size of 10 cycles is a good trade-off that simultaneously guarantee the accuracy and efficiency of the rapid RUL prediction. Smaller time window sizes, such as 1 or 5 cycles, or much larger time window sizes, such as 30cycles, present worse performance for RUL predictions. The reason may be of two folds. A too small time-window size cannot provide sufficient historical information for an accurate RUL prediction. While, with a too large time-window, the current ATCMN framework cannot effectively capture the full temporal information from such long sequence data. Therefore, the optimal time window size is determined as 10 cycles in this study.

4.3. Verification of the method robustness

To verify the robustness of the proposed method, experiments are further carried out on datasets of Group C and D, which have different battery materials and test conditions with Group A and B. In these two battery groups, the *T* parameter is not recorded; therefore, only *t* and *Q* parameters are employed as inputs. In each group, cell C1–C2, D1–D2 are used as offline training samples, while cell C3 and D3 are utilized as online testing samples. Results are demonstrated in Table 4 and Fig. 12.

As shown in Table 4, the proposed framework presents a strong generalizability to various LIB materials and operation conditions. For cell C3, the MAE and MAPE are 23 cycles and 17.1 %, respectively. While these metrics for cell D3 are 13 cycles and 6.8 %, respectively. As presented in Fig. 12, the proposed method can basically capture the actual RUL decay trend for cells C3 and D3. There is still room for improving the prediction accuracy if more LCO and NMC cell samples are available for the model training.

4.4. Comparison with benchmarks

To verify the effectiveness of the proposed ATCMN method, we benchmark it against three state-of-the-art deep learning frameworks reported in recent studies. In [24], a parallel CNN and LSTM deep neural network was developed to process a set of handcrafted features and predict the battery RUL. In [27], a deep CNN-based framework was proposed to learn battery degradation patterns from the raw battery charge-discharge data and a set of handcrafted features as well as predict the battery RUL. The 'Full RUL DNN' in [27] is considered as a benchmark in our experiments, because it serves for the same prediction task as our study, which is to predict the battery RUL with data obtained at any ageing stage. In [28], a hybrid parallel residual CNN (HPR CNN) framework was proposed to process the raw voltage, current, and temperature data, as well as output the battery RUL. For a fair comparison, we design inputs (including handcrafted features/raw battery parameters) for these deep learning frameworks based on 10-cycle battery data collected from the partial charge/discharge process. We also train the deep learning model with only 10-cycle data, although some benchmarking model is trained with a larger scope of data. One representative testing cell is picked out from each battery group, and comparison results are summarized in Table 5.

As shown in Table 5, the proposed ATCMN framework demonstrates a higher accuracy than the considered benchmarking deep learning models. Prediction errors of the ATCMN framework are the lowest for all testing groups, especially in Group C and D, which contains different

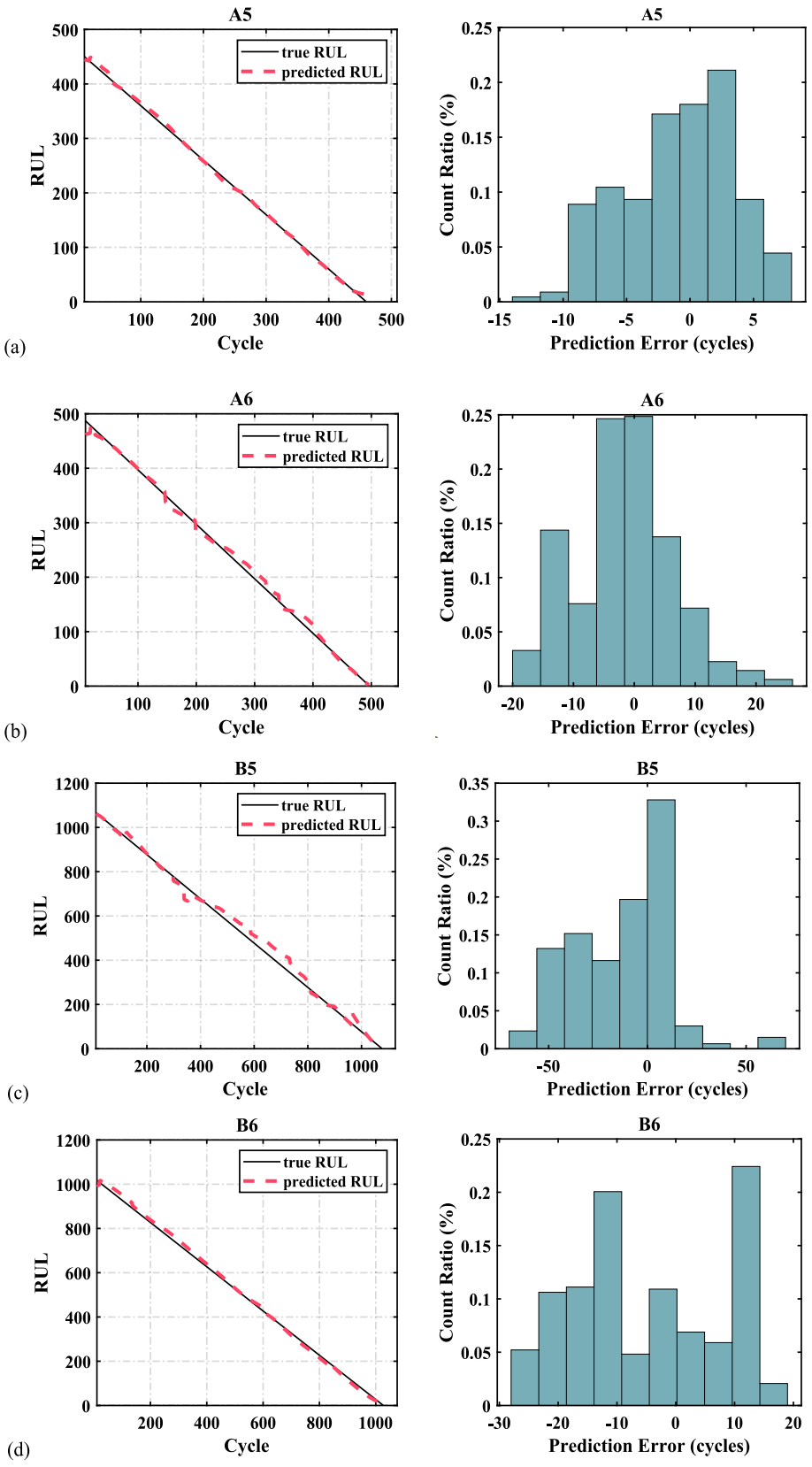


Fig. 11. RUL prediction results of testing cells in Group A and B using the three-parameter input.

Table 3
RUL prediction results of different time window sizes.

Time window size	Group A				Group B			
	Cell A5		Cell A6		Cell B5		Cell B6	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
1	14	12.7 %	19	14.6 %	38	19.3 %	22	18.5 %
5	11	11.8 %	11	8.3 %	30	15.2 %	21	17.8 %
10	**4**	7.4 %	**8**	**7.5 %**	**21**	**10.8 %**	11	**7.7 %**
15	**4**	**7.1 %**	15	9.1 %	37	18.8 %	**10**	7.9 %
20	9	10.0 %	17	13.1 %	33	17.1 %	16	9.2 %
30	12	12.6 %	25	18.7 %	40	19.7 %	27	18.8 %

Table 4
RUL prediction results on Group C and D.

Cell no.	Specifications		Performance	
	Material	Charge/discharge rate	MAE	MAPE
C3	LiCoO2	0.5/1C	23	17.1 %
D3	Li(*NMC*)O2	1/1C	13	6.8 %

types of LIBs. This also indicates the better generalization ability of the proposed ATCMN method.

Results of previous comparative studies have demonstrated the superiority of the ATCMN in predicting battery RUL using data obtained at any ageing stage. Meanwhile, it is worth noting that some latest studies could realize accurate early-stage battery lifetime predictions via data obtained within the first 100 cycles. To further examine the performance of the ATCMN from the perspective of early-stage battery lifetime prediction, we compare it with three state-of-the-art methods reported in recent studies. The first benchmark has a single handcrafted feature and corresponds to the "variance" model presented in [2]. The feature was extracted based on the first-100-cycle degradation data and was input into an elastic net (EN) model to make early battery lifetime predictions. The second benchmark adopts 18 handcrafted features, corresponding to the "full" model presented in [2]. The whole set of features was extracted using the first-100-cycle degradation data and was fed into an EN model to predict the battery lifetime. The third benchmark has four handcrafted features obtained based on the first-80-cycle degradation data [35]. The features were input into a general regression neural network (GRNN) to early predict the battery lifetime. These three benchmarking methods were all proposed based on Dataset I; therefore, comparative experiments are conducted on Group A and B from Dataset I. Two metrics, the absolute error (AE) and absolute percentage error (APE) between the true and predicted battery lifetime, are adopted to evaluate the method performance. Results are reported in Table 6.

As shown in Table 6, in the early-stage battery lifetime prediction task, the proposed framework outperforms considered benchmarks with a higher accuracy and much earlier prediction point. By using just the first-10-cycle information, our method achieves APEs of 0.6 %, 4.2 %, 0.6 % and 1.5 % on testing cells A5, A6, B5 and B6, respectively. These metrics are lower than all three benchmarks, which utilize first-80-cycle

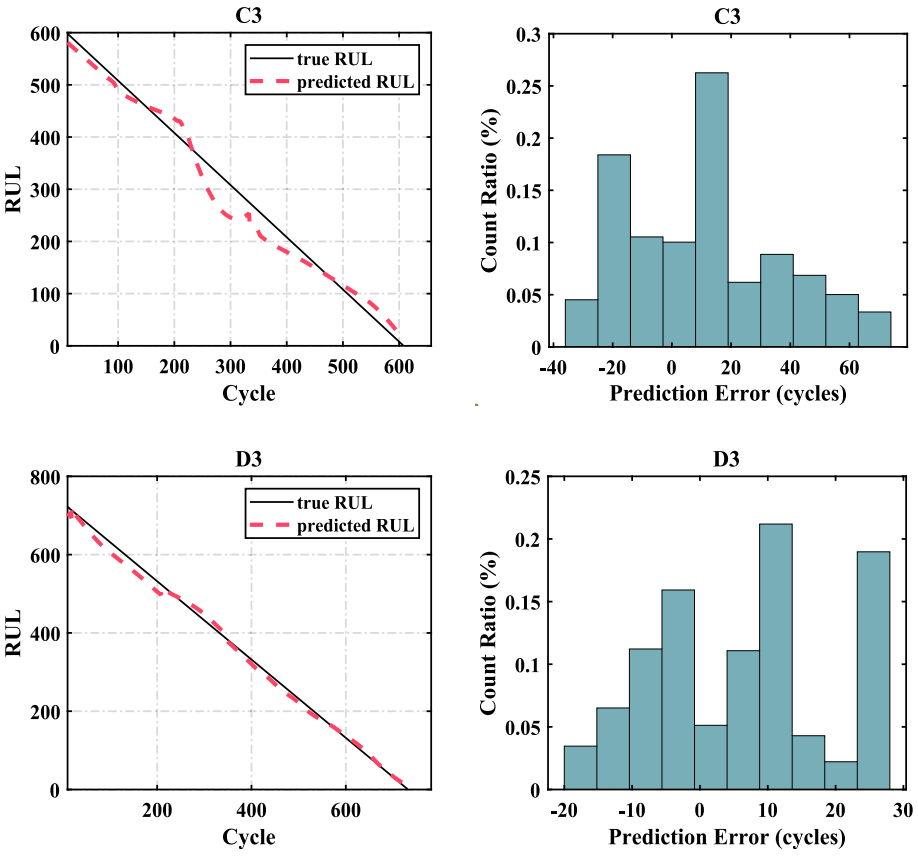


Fig. 12. RUL prediction results on Group C and D.

Table 5
RUL prediction results of different deep learning frameworks.

Model	Group A		Group B		Group C		Group D	
	Cell A5		Cell B6		Cell C3		Cell D3	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
ATCMN	**4**	**7.4 %**	**11**	**7.7 %**	**23**	**17.1 %**	**13**	**6.8 %**
CNN + LSTM [24]	28	19.7 %	31	22.9 %	62	34.3 %	41	26.4 %
Deep CNN [27]	12	11.6 %	18	13.4 %	38	21.3 %	26	18.9 %
HPR CNN [28]	16	13.2 %	24	17.7 %	47	25.2 %	33	23.8 %

Numbers with bold font represent best prediction results.

Table 6
Early-stage battery lifetime prediction results of different benchmarks.

Benchmarks (early prediction position)	Cell no.	True lifetime	Predicted lifetime	AE (cycles)	APE (%)
1 feature + EN [2] (Cycle 100)	A5	471	486	15	3.2 %
	A6	498	655	157	24.0 %
	B5	1078	1256	178	14.2 %
	B6	1028	1065	37	3.5 %
18 features + EN [2] (Cycle 100)	A5	471	480	9	1.9 %
	A6	498	605	107	17.7 %
	B5	1078	1160	82	7.1 %
	B6	1028	1088	60	5.5 %
4 features + GRNN [35] (Cycle 80)	A5	471	510	39	7.6 %
	A6	498	673	175	26.0 %
	B5	1078	1118	40	3.6 %
	B6	1028	992	36	3.6 %
Proposed method (**Cycle 10**)	A5	471	468	**3**	**0.6 %**
	A6	498	477	**21**	**4.2 %**
	B5	1078	1072	**6**	**0.6 %**
	B6	1028	1013	**15**	**1.5 %**

Numbers with bold present best AE and APE attained.

and first-100-cycle battery data. Apart from this early-stage battery lifetime prediction, the proposed method can also make accurate RUL predictions at any ageing stage along the whole life trajectory of LIBs with only 10-cycle historical information.

4.5. Ablation study: deeply learned features vs. handcrafted features

In this section, we conduct an ablation study to examine two hypotheses: 1) Will latent features learned from the proposed ATCMN beat the commonly used handcrafted features? 2) Will a combination of learned latent features and handcrafted features better serve the LIB RUL prediction task? Based on the proposed method, a new framework jointly considering learned latent features and handcrafted features is developed as shown in Fig. 13.

As shown in Fig. 13, in this new framework, a LASSO model is applied to predict the battery RUL, which is a commonly used regularized model that could force many feature coefficients to become smaller or exactly zero through adding a l_1-norm penalty on the loss function [36]. To examine the performance of two types of features, we set three

Table 7
Ablation study: handcrafted features.

No.	Feature description
1	Average height of IC peak from cycle $i - W + 1$ to i
2	Average voltage of IC peak from cycle $i - W + 1$ to i
3	Average discharging time from cycle $i - W + 1$ to i
4	Variance of $\Delta Q_{i-(i-W+1)}(V)$
5	Mean of $\Delta Q_{i-(i-W+1)}(V)$
6	Mean of minimum temperature from cycle $i - W + 1$ to i
7	Mean of average temperature from cycle $i - W + 1$ to i
8	Mean of maximum temperature from cycle $i - W + 1$ to i
9	Internal resistance, difference between cycle $i - W + 1$ and i
10	Average internal resistance from cycle $i - W + 1$ to i

experimental groups of model inputs: 1) learned latent features $\boldsymbol{Z}$ from the proposed ATCMN; 2) ten handcrafted features $\boldsymbol{H}$ inspired by [3], as listed in Table 7; 3) combination of two types of features $\boldsymbol{Z} + \boldsymbol{H}$. Experiments are conducted on Group A and B of Dataset I, and results are summarized in Table 8.

Based on Table 8, answers for the two hypotheses can be obtained:

1) Learned latent features outperform handcrafted features in this rapid battery RUL prediction task. When considering $\boldsymbol{Z}$ as the input, prediction errors of all test cells are lower than those of considering $\boldsymbol{H}$ as the input. Moreover, through analyzing the feature selection result of LASSO, we find that the selected features almost entirely belong to learned latent features, which implies the great value of learned latent features.
2) The combination of learned latent features and handcrafted features cannot bring improvement to the prediction. For example, when both types of features are utilized, the MAE and MAPE of cell A5 are 20 cycles and 14.1 %, respectively. Improvement cannot be obtained by comparing with results of using only $\boldsymbol{Z}$ as the input. This observation is consistent for all test cells.

5. Conclusions

In this paper, a deep ATCMN framework was developed for the rapid RUL prediction of LIBs. The ATCMN realized accurate predictions via data of just 10 incomplete cycles obtained at any ageing stage of LIBs. To apply the ATCMN, the battery time, capacity, and temperature data collected during the incomplete discharge process were reorganized as a tensor structure to serve as the input of the ATCMN. Next, the ATCMN

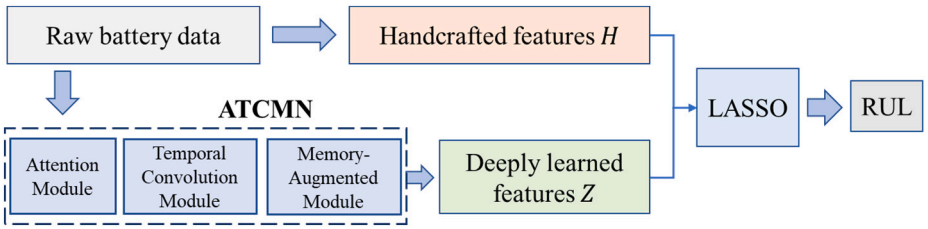


Fig. 13. Ablation study: RUL prediction framework integrating handcrafted features and learned latent features.

Table 8
RUL prediction results of using different feature inputs.

Features	Group A				Group B			
	Cell A5		Cell A6		Cell B5		Cell B6	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
Learned latent features	**17**	**12.4 %**	**21**	**17.6 %**	**45**	**25.7 %**	**26**	**17.4 %**
Handcrafted features	48	24.9 %	55	26.8 %	74	32.8 %	59	30.7 %
Handcrafted + Latent features	20	14.1 %	22	17.8 %	46	26.0 %	29	17.6 %

Numbers with bold present the lowest value of MAE and MAPE produced by considering different types of features based on 4 tested battery cells.

composed of four network-based modules, the attention module, temporal convolution module, memory-augmented module, and prediction module, was sequentially conducted to engineer high-level spatial and temporal latent features as well as develop the mapping from latent features to the RUL.

In computational studies, three datasets containing LIBs of different materials and tested with different conditions were utilized to study the accuracy and generalizability of the ATCMN on battery RUL predictions. Results indicated that 10-cycle was the optimal window size setting that simultaneously guaranteed the accuracy and efficiency of rapid battery RUL predictions. Results also demonstrated that the ATCMN could provide higher accuracy, faster and earlier prediction, as well as better generalizability than a set of state-of-the-art deep learning frameworks and early battery lifetime prediction methods. Finally, results of an ablation study indicated that features learned by the ATCMN brought higher value than commonly used handcrafted features in this task.

The proposed framework offers a promising potential to accelerating the design optimization, online diagnosis, and recycling of LIBs. However, there still exist a few limitations on this method, and we plan to further improve it from the following two directions.

1) The proposed method could work well on different types of LIBs; yet, the model training is separately conducted for each type of LIBs. This is mainly limited by the inherent degradation discrepancy between different battery chemistries and different operation conditions, as well as the limited number of training batteries of different types. To tackle this issue, we plan to collect data of more types of battery degradation samples from both our laboratory experiments (experimental data) and our industrial partner (practical operational data). Based on sufficient training samples, we would like to explore more innovative deep learning model structures, such as the Vision Transformer-based deep model, to develop a more generalized RUL prediction framework which could realize the model training and accurate RUL prediction for different types of LIBs simultaneously.
2) The proposed method is currently verified with data only collected at the laboratory setting. The method effectiveness still needs further evaluation based on real-world operational data. In the future, we plan to collaborate with our industrial partner to obtain more practical battery data and further improve and verify the proposed method.

CRediT authorship contribution statement

Zicheng Fei: Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft. **Zijun Zhang:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration. **Fangfang Yang:** Data curation, Formal analysis. **Kwok-Leung Tsui:** Resources, Writing – review & editing, Supervision, Funding acquisition.

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

Two datasets utilized in this research are publicly available. Authors have included links of accessing these data in the manuscript. Dataset I: https://data.matr.io/1/projects/5c48dd2bc625d700019f3204. Dataset II: https://web.calce.umd.edu/batteries/data.htm.

Acknowledgments

This research was supported in part by Hong Kong Research Grants Council General Research Fund project with No. 11204419, in part by National Natural Science Foundation of China Young Scientist Fund project with No. 52007160, and in part by CityU HK Strategic Research Grant with No. 7005692.

References

[1] S. Zhang, B. Zhai, X. Guo, K. Wang, N. Peng, 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 26 (2019), 100951, 2019/12/01/.
[2] K.A. Severson, P.M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M.H. Chen, M. Aykol, P.K. Herring, D. Fraggedakis, Data-driven prediction of battery cycle life before capacity degradation, Nat. Energy 4 (5) (2019) 383–391.
[3] Z. Fei, F. Yang, K.-L. Tsui, L. Li, Z. Zhang, Early prediction of battery lifetime via a machine learning based framework, Energy 225 (2021).
[4] Z. Fei, Z. Zhang, F. Yang, K.-L. Tsui, L. Li, Early-stage lifetime prediction for lithium-ion batteries: a deep learning framework jointly considering machine-learned and handcrafted data features, J. Energy Storage 52 (2022).
[5] M.B. Pinson, M.Z. Bazant, Theory of SEI formation in rechargeable batteries: capacity fade, accelerated aging and lifetime prediction, J. Electrochem. Soc. 160 (2) (2012), A243.
[6] S. Koohi-Fayegh, M.A. Rosen, A review of energy storage types, applications and recent developments, J. Energy Storage 27 (2020), 101047.
[7] A. Farmann, D.U. Sauer, Comparative study of reduced order equivalent circuit models for on-board state-of-available-power prediction of lithium-ion batteries in electric vehicles, Appl. Energy 225 (2018) 1102–1122, 2018/09/01/.
[8] Y. Wang, M. Li, Z. Chen, Experimental study of fractional-order models for lithium-ion battery and ultra-capacitor: modeling, system identification, and validation, Appl. Energy 278 (2020), 115736, 2020/11/15/.
[9] Y. Chang, H. Fang, Y. Zhang, A new hybrid method for the prediction of the remaining useful life of a lithium-ion battery, Appl. Energy 206 (2017) 1564–1578, 2017/11/15/.
[10] Y. Xing, E.W.M. Ma, K.-L. Tsui, M. Pecht, An ensemble model for predicting the remaining useful performance of lithium-ion batteries, Microelectron. Reliab. 53 (6) (2013) 811–820, 2013/06/01/.
[11] F. Yang, X. Song, G. Dong, K.-L. Tsui, A coulombic efficiency-based model for prognostics and health estimation of lithium-ion batteries, Energy 171 (2019) 1173–1182.
[12] L. Chen, Y. Ding, B. Liu, S. Wu, Y. Wang, H. Pan, Remaining useful life prediction of lithium-ion battery using a novel particle filter framework with grey neural network, Energy 244 (2022) 122581.
[13] X. Lai, Y. Huang, X. Han, H. Gu, Y. Zheng, A novel method for state of energy estimation of lithium-ion batteries using particle filter and extended Kalman filter, J. Energy Storage 43 (2021), 103269.
[14] Z. Chen, N. Shi, Y. Ji, M. Niu, Y. Wang, Lithium-ion batteries remaining useful life prediction based on BLS-RVM, Energy 234 (2021), 121269.
[15] J. Wang, Z. Deng, T. Yu, A. Yoshida, L. Xu, G. Guan, A. Abudula, State of health estimation based on modified Gaussian process regression for lithium-ion batteries, J. Energy Storage 51 (2022/07/01/, 2022.), 104512.
[16] X. Li, L. Zhang, Z. Wang, P. Dong, Remaining useful life prediction for lithium-ion batteries based on a hybrid model combining the long short-term memory and Elman neural networks, J. Energy Storage 21 (2019) 510–518.

[17] S. Kim, Y.Y. Choi, K.J. Kim, J.-I. Choi, Forecasting state-of-health of lithium-ion batteries using variational long short-term memory with transfer learning, J. Energy Storage 41 (2021).
[18] G. Ma, Y. Zhang, C. Cheng, B. Zhou, P. Hu, Y. Yuan, Remaining useful life prediction of lithium-ion batteries based on false nearest neighbors and a hybrid neural network, Appl. Energy 253 (2019), 113626.
[19] Y. Zhou, M. Huang, Y. Chen, Y. Tao, A novel health indicator for on-line lithium-ion batteries remaining useful life prediction, J. Power Sources 321 (2016) 1–10.
[20] F. Xu, F. Yang, Z. Fei, Z. Huang, K.-L. Tsui, Life prediction of lithium-ion batteries based on stacked denoising autoencoders, Reliab. Eng. Syst. Saf. 208 (2021).
[21] L. Wang, C. Pan, L. Liu, Y. Cheng, X. Zhao, On-board state of health estimation of LiFePO4 battery pack through differential voltage analysis, Appl. Energy 168 (2016) 465–472, 2016/04/15/.
[22] X. Hu, J. Jiang, D. Cao, B. Egardt, Battery health prognosis for electric vehicles using sample entropy and sparse Bayesian predictive modeling, IEEE Trans. Ind. Electron. 63 (4) (2015) 2645–2656.
[23] X. Li, C. Yuan, Z. Wang, State of health estimation for Li-ion battery via partial incremental capacity analysis based on support vector regression, Energy 203 (2020), 117852, 2020/07/15/.
[24] L. Ren, J. Dong, X. Wang, Z. Meng, L. Zhao, M.J. Deen, A Data-Driven Auto-CNN-LSTM Prediction Model for Lithium-Ion Battery Remaining Useful Life, IEEE Transactions on Industrial Informatics 17 (5) (May 2021) 3478–3487.
[25] J. Hong, D. Lee, E.-R. Jeong, Y. Yi, Towards the swift prediction of the remaining useful life of lithium-ion batteries with end-to-end deep learning, Appl. Energy 278 (2020).
[26] T. Han, Z. Wang, H. Meng, End-to-end capacity estimation of Lithium-ion batteries with an enhanced long short-term memory network considering domain adaptation, J. Power Sources 520 (2022).
[27] C.-W. Hsu, R. Xiong, N.-Y. Chen, J. Li, N.-T. Tsou, Deep neural network battery life and voltage prediction by using data of one cycle only, Appl. Energy 306 (2022).
[28] Q. Zhang, L. Yang, W. Guo, J. Qiang, C. Peng, Q. Li, Z. Deng, A deep learning method for lithium-ion battery remaining useful life prediction based on sparse segment data via cloud computing system, Energy 241 (2022).
[29] X. Hu, L. Xu, X. Lin, M. Pecht, Battery lifetime prognostics, Joule 4 (2) (2020) 310–346.
[30] W. He, N. Williard, M. Osterman, M. 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, 2011/12/01/.
[31] S. Woo J. Park J.-Y. Lee I. S. Kweon n.d. "Cbam: Convolutional Block Attention Module." pp. 3-19.
[32] J. Bi, X. Zhang, H. Yuan, J. Zhang, M. Zhou, A Hybrid Prediction Method for Realistic Network Traffic With Temporal Convolutional Network and LSTM, IEEE Transactions on Automation Science and Engineering 19 (3) (July 2022) 1869–1879.
[33] F. Chollet n.d. "Xception: Deep Learning With Depthwise Separable Convolutions." pp. 1251-1258.
[34] D. Gong L. Liu V. Le B. Saha M. R. Mansour S. Venkatesh A. V. D. Hengel, n.d. "Memorizing Normality to Detect Anomaly: Memory-augmented Deep Autoencoder for Unsupervised Anomaly Detection." pp. 1705-1714.
[35] Y. Zhang, Z. Peng, Y. Guan, L. Wu, Prognostics of battery cycle life in the early-cycle stage based on hybrid model, Energy 221 (2021).
[36] J. Li, K. Cheng, S. Wang, F. Morstatter, R.P. Trevino, J. Tang, H. Liu, Feature selection: a data perspective, ACM Comput. Surv. 50 (6) (2017), 94.