



Li-ion battery capacity prediction using improved temporal fusion transformer model



William Gomez, Fu-Kwun Wang*, Jia-Hong Chou

Department of Industrial Management, National Taiwan University of Science and Technology, Taipei City, 106335, Taiwan

ARTICLE INFO

Handling editor: A. Olabi

Keywords:

Remaining useful life prediction
Improved temporal fusion transformer
Bidirectional long short-term memory
Tree-structure parzen estimator

ABSTRACT

Lithium-ion (Li-ion) batteries have near-zero energy emissions and provide power to various devices, such as automobiles and portable equipment. The strategy predicts the capacity of Li-ion in advance and can also help arrange maintenance tasks. To improve state of health (SOH) and remaining useful life (RUL) prediction accuracy, we propose an improved temporal fusion transformer (ITFT) method based on bidirectional long short-term memory (Bi-LSTM) encoder-decoder layer replacing the long short-term memory for probabilistic online RUL prediction. A novel interpretable hyperparameter tuning method called Bayesian optimization based on tree-structure Parzen estimator (TPE) is coupled with a unique ITFT model to improve RUL prediction accuracy. Furthermore, we consider the effects of keen-onset to establish the starting point of our training. The root mean square error for four batteries using the proposed model for the test data are 0.0018, 0.0019, 0.0013, and 0.0025, respectively, which outperforms the other models, with an improvement accuracy rate above 25%. The proposed model SOH results indicate that our proposed approach outperforms some previously published methods. Our online RUL prediction demonstrates relative errors of 1.18%, 1.54%, 1.06%, 2.70%, 0.67%, 2%, 3.90%, 0%, and 3.08% for nine batteries, respectively. These results for SOH and RUL predictions emphasize the excellent performance of our proposed method.

1. Introduction

Predictive maintenance (PdM) solutions are rising in popularity in electronic manufacturing. With a great industrial revolution and change in energy production, researchers and the government face tremendous pressure to find better alternatives for energy efficiency. As fossil fuel reserves dwindle, lithium-ion (Li-ion) batteries (LIBs), a form of chemical rechargeable battery, have become increasingly important in fulfilling the growing demand for electric energy. As a result of their outstanding characteristics, such as their low maintenance requirements, high voltage efficiency, long life cycle, and market-leading energy density, lithium-ion batteries are widely used in electric vehicles (EVs), communication, smart grids, and energy storage systems [1–3]. However, a high-risk failure scenario can result from extreme operating conditions, such as gassing, electrode mechanical cracking, and internal short circuits, repeatedly charging and discharging. When Li-ion batteries are used in harsh environments or subjected to poorcharging habits, etc., their degradation will be accelerated. In addressing some of the limitation of LIBs for better battery performance system, Yang et al. [4] systematically investigated the thermal characteristics of a hybrid

solid-liquid battery to help improve the development of future battery thermal management systems (BTMS). To maintain continuous energy production and storage and ensure EVs' safety and reliability, it is crucial to accurately estimate the battery's state of health (SOH) and monitor it continuously to avoid sudden failures [3–6]. The critical function of battery management systems (BMS) relies on scientific and accurate predictions of a battery's state of health, remaining useful life (RUL), and analyzing the fault [7].

There is growing interest in this field, yet several key estimations remain broadly under research. These include determining the battery's threshold failure for cell replacement and identifying the optimal training and prediction starting points for accurate RUL prediction. A thorough review of knee point and knee onset for a starting point was conducted by Attia et al. [8], comparing six knee methods. The knee point is defined as the maximum point curvature by Ref. [9], while Diao et al. [10] defined the knee as the point at which slope changes rapidly and the inflection point. A better understanding of knee –onset or –point will help formulate more effective and robust methods for RUL prediction because domain knowledge of knee point or onset assists in determining the training start point. A smoothed degradation curve or heatmap, such as the SOH row in Fig. 3, allows for the easy visual

* Corresponding author.

E-mail address: fukwun@mail.ntust.edu.tw (F.-K. Wang).

Nomenclature		Num. layers	Number of layers
Q_i	battery's discharged capacity		
Q_o	nominal (rated) capacity		
X_a	Original actual value		
X_{min}	sample minimum value		
X_{max}	sample maximum value		
N_i	normalized data		
T_j	time-varying variable		
$X_{s,t}$	time-varying known input		
$P_{s,t}$	past input		
\overrightarrow{h}	Forward hidden sequence		
\overleftarrow{h}	Backward hidden sequence		
$W_{\overrightarrow{h}y}$	Forward Weight of Bi-LSTM		
$W_{\overleftarrow{h}y}$	Backward Weight of Bi-LSTM		
b_y	Bias of Bi-LSTM		
$Pr(g(x))$	probability of good		
$Pr(l(x))$	probability of bad		
$g(x)$	good sample		
$l(x)$	bad sample		
s	static covariate		
<i>Greek symbols</i>			
α_0	intercept		
γ	abruptness of the transition		
δ_t	Weight vector for input variable		
<i>Subscript</i>			
Num. att. heads		Number of attention heads	
		<i>Acronyms</i>	
		Adam	Adaptive moment estimation
		Ah	Ampere-hour
		Bi-LSTM	Bidirectional long short-term memory
		BMS	Battery management systems
		BO	Bayesian optimization
		C	Capacity
		CC-CV	Constant current and constant voltage
		DL	Deep learning
		EVs	electric vehicles
		EOL	End of life
		GRN	Gated residual network
		IR	Internal resistance
		ITFT	Improved temporal fusion transformer
		LIBs	Lithium-ion batteries
		LSTM	Long short-term memory network
		PdM	Predictive maintenance
		RELU	rectified linear unit
		RUL	Remaining useful life
		SOC	State of charge
		SOH	State of health
		TFT	Temporal fusion transformer
		TPE	Tree-structure Parzen estimator
		<i>Superscript</i>	
		\circ	degree
		\odot	Hadamard product

identification of Knee point. In equipment performance prediction, RUL denotes the amount of time that is still useable before a system declines to an intolerable degree, which is the battery's end of life (EOL). When historical data and current state data are present, life prediction can use to estimate how long until one (or more) failures occur [10]. In recent years, several studies developed accurate RUL prediction models for LIBs, such as mathematical-based, data-driven, and hybrid models [11–13]. With the advancement of artificial intelligence, new and promising techniques for predicting lithium-ion battery RUL have emerged. However, many of these models suffer from a lack of interpretability and simplicity. This is often due to their reliance on a black-box function approach, where the inner workings remain obscure.

Model-based and data-driven techniques are the two main categories, as demonstrated by these research accomplishments over the years [13]. The model-based approach seeks to create a mathematical representation of the lithium battery degrading process, for example, the Kalman filter [14,15]. Data-driven method, which comes in many forms like machine learning, deep learning, statistical and hybrid models, has gained tremendous popularity in RUL prediction, such as the particle filter-temporal attention mechanism-bidirectional gated recurrent unit (PF-BiGRU-TSAM) [16], autoregressive method [17], support vector regression [18,19], and other machine learning-based models [20]. Various features and complexity of time series data require proper tuning for better prediction accuracy. Deep learning (DL) and hybrid models have gained significant attention for their ability to extract complex patterns and features from large datasets to improve prediction.

Ren et al. [21] integrated an autoencoder with a deep neural network (DNN) to predict the RUL of batteries under different operating conditions with an accuracy of 93.34%. Zhang et al. [22] applied a bidirectional long short-term memory with attention (Bi-LSTM-AM) to estimate the RUL of lithium-ion batteries. Many deep learning and hybrid models on SOH and RUL prediction has been made in recent years [23–25]. Sun

et al. [26] proposed a hybrid method combining variational mode decomposition (VMD) with a long short-term memory network (LSTM) and Gaussian process regression algorithm (GPR). Ensemble empirical mode decomposition with adaptive noise (CEEMDAN) can denoise the Li-ion batteries dataset from NASA, then a neural-network-based method for RUL prediction that incorporates an LSTM network, particle swarm optimization, and an attention mechanism [27]. Tang et al. [28] proposed a hybrid deep-learning model. In the analysis of State of Health (SOH) estimation, a Convolutional Neural Network (CNN) [29] was proposed to estimate LIBs' SOH during constant current, with errors typically within $\pm 2\%$, not exceeding $\pm 3\%$ for SOH greater than 80%. Another study [30] combines model-based and data-driven methods, where suitable portions of the LIB original data and model parameters are used as health features (HFs). Amogne et al. [11] proposed a deep learning model using transfer learning to estimate SOH and RUL.

Most previous RUL prediction research on Li-ion batteries utilized offline techniques. Offline modeling generates multi-step predictions using the trained offline model based on previously known samples, and it does not require any additional adjustments. In contrast, few online methods proposed do not use many inputs like the voltage, internal resistance, cycle, and temperature profile as health indicators for prediction. The online RUL prediction method uses known available data samples to train a model and apply a rolling multi-step prediction approach to predict unknown test samples. This approach reflects the RUL prediction in real-time application. Online method performance accuracy is vital for smart battery management systems for health monitoring and decision-making. A Temporal Convolution-based Long-Short Term Memory (TCLSTM) network with an attention mechanism has been designed to study the role of PdM in facilitating flexible scheduling, efficient equipment operation, and preventing unexpected downtime [31]. Yang et al. [32] proposed a novel hybrid multi-task learning data-driven model named GBLS Booster where

CNN-Transformer algorithm is suggested to augment the GBLS model using decomposition method, and the TPE technique for hyperparameter tuning.

Recently, the temporal fusion transformer (TFT) model proposed by Lim et al. [33] has been cited and used in many literatures and used in time series analysis. No studies apply this model to battery capacity and RUL estimation. TFT model has the advantage of learning long-term dependencies using a self-attention mechanism; this means it can directly learn patterns during the training process, giving the model more advantage over many DL models in terms of interpretability, robustness, and stability.

To address the problem of interpretability and improve prediction accuracy for RUL prediction, we propose an improved temporal fusion transformer (ITFT) based on a variant of Bayesian optimization (BO) known as tree-structure Parzen estimator (TPE) algorithm for hyperparameters tuning. We put forward a unique and improved TFT model based on bidirectional long short-term memory (Bi-LSTM) encoder-decoder layer replacing the LSTM for probabilistic online RUL prediction of Li-ion batteries. A knee-onset point as the training starting point is part of the proposed approach, which incorporates fixed-length training data points, while prediction starting is determined when SOH equals 0.86. During RUL prediction, online prediction makes predictions for unknown data when SOH is unavailable at the time of prediction. The method involves an iterative process that is repeated based on a multi-step ahead until the predicted capacity value reaches the failure threshold. The algorithm stops, and the predicted value of battery RUL is finally computed.

Motivated by some limitations of DL models concerning interpretability, long-term and local dependency learning ability, and power in understanding the nonlinearity relationship between SOH and multiple health indicators, a unique improved TFT model is developed with TPE for hyperparameter tuning and quantile prediction interval for uncertainty estimation.

The following points are a summary of the study's contributions and objectives:

- (1) To the best of our knowledge, this study is the first to apply an improved TFT in exploring online RUL prediction and Li-ion battery capacity degradation trend with multiple inputs using TPE for hyperparameter tuning for better prediction accuracy and stability of RUL estimation.
- (2) This study delves into a critical area of deep learning: model interpretability, which is the significance of attention weight, hyperparameters contribution, and critical features. It contrasts with many earlier deep learning models, which often needed more comprehensive interpretability of results. As a result, it pinpoints crucial factors in RUL prediction, offers credible SOH forecast analysis, and helps BMS make decisions.
- (3) The study utilizes the concept of knee-point [10] and double Bacon Watts knee-onset [34] to identify the initial transition to the accelerated degradation phase of Li-ion battery capacity. The double knee-onset is then designated as the training starting point.
- (4) Our proposed model effectively describes battery capacity degradation, even with a limited amount of historical data.

The rest of this paper is described as follows. Section 2 provides detailed implementations and descriptions of datasets. Section 3 presents a proposed network architecture, followed by experimental results of the SOH forecast and RUL estimation in section 4. Finally, section 5 outlines the significant conclusions and future research scope.

2. Li-ion capacity degradation and knee-onset concept

A total of nine Li-ion batteries are used in this study to test the proposed model for cycle life prediction. From the Toyota data set, nine

cylindrical cells with a rated capacity of 1.1 Ah and a nominal voltage is 3.3 V are selected [35]. An Arbin LBT potentiostat with 48 channels, set to 30 °C, was used to cycle these Li-ion iron phosphate (LFP)/graphite cells called A123 Systems (APR18650M1A). All batteries were charged at 1C (constant current-constant voltage) with the second current step ending at 80 % state of charge (SOC).

The degradation of Li-ion batteries uses various performance characteristics, such as a decrease in maximal capacity, power output reductions, and an increase in internal resistance. A battery's overall aging degree can be quantified using SOH. We define SOH as the ratio between actual and nominal capacity. It is given by:

$$SOH_i = \frac{Q_i}{Q_o} \times 100 \quad (1)$$

where SOH_i is the SOH of the battery at the i th cycle, Q_i is the battery's discharged capacity at the i th cycle, and Q_o is the nominal (rated) capacity (1.1Ah).

During fast charging, all batteries charge using a one-step or two-step procedure with a method of C1(Q1)-C2(80%)-1C (100%)-4C(0%). The first step uses constant current, while the second switches currents at a specific SOC. In the second current step, the cells charge to 80% SOC after charging to 1C (constant current-constant voltage-(CC-CV)). When the cells reach 80% saturation, they charge galvanostatically at 1C to 3.6 V and then potentostatically at 3.6 V with a current cutoff of C/50. Charge rates (C-rates) for the first and second steps are C1 and C2, respectively, while SOC at the time of current switching is Q1. The C-rate refers to the rate at which a battery is charged or discharged relative to its capacity. For example, a C-rate of 1C means that the current being applied will charge or discharge the battery in 1 h. A battery is tested at 80% of its original capacity in a 30 °C environmental chamber under different fast charging conditions. Fig. 1 shows the degradation curves of nine batteries used for evaluation. Similar deterioration paths are observed in batteries tested at the same discharge rate, showing that the available capacity gradually degrades.

This study defines RUL as the number of cycles between the current number (prediction starting point) and the number at the end of life. It is defined as:

$$RUL_{cycle} = T_{EOL} - T_{sp} \quad (2)$$

where RUL_{cycle} is the remaining useful life at the prediction starting point (T_{sp}), T_{EOL} is the end of life, and T_{sp} is the prediction starting cycle number for RUL prediction. For example, Fig. 2 shows RUL estimation under a failure threshold of 80%.

We used nine battery cells under different charging policy to validate the prediction performance, stability and robustness of our proposed model. Technical specification and operating parameters of the batteries are briefly detailed in Tables 1 and 2 respectively, details can be found in Ref. [35]. Fig. 4 illustrates the 3D plot depicting the State of Health (SOH) and temperature of battery #1 over time (cycle). The analysis reveals that temperatures exceeding 30 °C have a minimal impact on battery capacity, while higher temperature charging rates lead to capacity reduction. This suggests that a slow charging approach at low multiples is advisable when designing battery charging strategies under low temperature conditions.

Randomly selected batteries from Toyota dataset with nine different characteristic parameters and applied multivariate analysis approach with normalized capacity as the SOH, which is the dependent variable and cycle, average temperature (Tavg) and internal resistance (IR) as independent variables, $SOH = f(\text{cycle}, \text{average_temperature}, \text{IR})$. We use a public Toyota Li-on battery dataset of 124 graphite/LiFePO₄ batteries for analysis.

Using the Pearson correlation coefficient, we conducted a correlation analysis between SOH and the independent variables. All parameters were negatively correlated, as shown in Table 3.

Chou et al. [36] applied the Bacon-Watts method to determine the

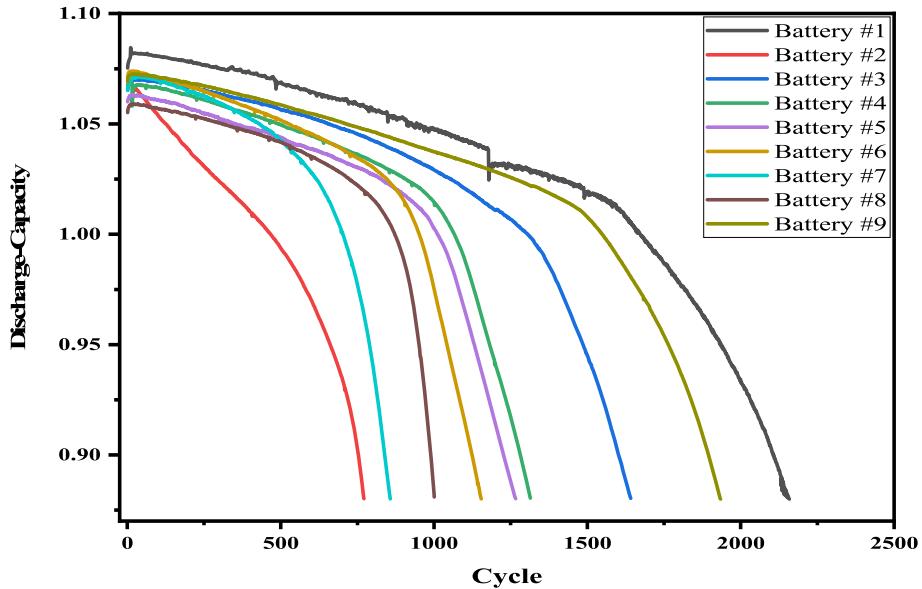


Fig. 1. Capacity deterioration curves of batteries.

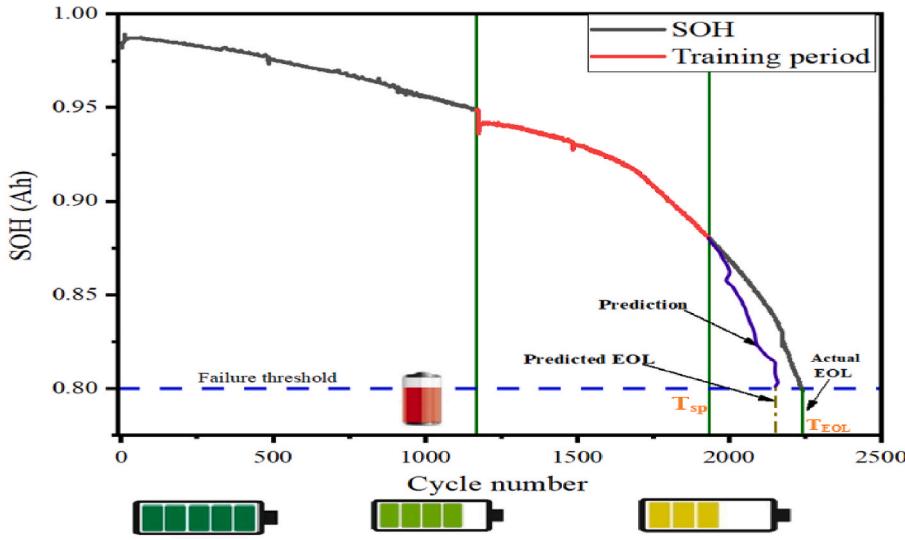


Fig. 2. Definition of RUL estimation.

knee-onset as the training starting point for RUL online prediction. In this study, we applied the double Bacon-Watts model which is a modification of the Bacon-Watts model. Two straight-line relationships find the left and right of some unknown transition point x_1 . The Bacon-Watts model is defined as follows:

$$SOH = \alpha_0 + \alpha_1(x - x_1) + \alpha_2(x - x_1)\tanh[(x - x_1)/\gamma] + Z \quad (3)$$

where x is the cycle number, α_0 is a type of intercept of the leftmost segment (at $x = x_1$), α_1 and α_2 control the slopes of the intersecting lines, γ is the abruptness of the transition with a small value, and Z is the residuals following a normal distribution with mean zero and variance constant.

The double Bacon-Watts model is as follows:

$$SOH = \alpha_0 + \alpha_1(x - x_0) + \alpha_2(x - x_0)\tanh[(x - x_0)/\gamma] + \alpha_3(x - x_2)\tanh[(x - x_2)/\gamma] + Z \quad (4)$$

where x_0 and x_2 represent the change points, all the other terms are in the same vein as in Eq. (3). The knee-point detection is not to inform

end-user in advance that their capacity is diminishing due to nonlinearity but simply that accelerated SOH degradation is underway. At the knee onset, capacity decay ceases to be approximated as a linear function, marking accelerated degradation onset [9]. Knee onset is preferred in the application when estimating RUL prediction. Fig. 5 depicts the knee point, Bacon Watts, and double Bacon Watts knee onset for batteries # 1 and #4. The double Bacon-Watts knee onset is the training starting point for the RUL prediction in the study, the first vertical line (green) in Fig. 2. The knee-onset of the double Bacon-Watt model occurs at cycle #1549. This indicates that the accelerated degradation of State of Health (SOH) begins at this cycle. The double Bacon-Watts knee-onset points for all batteries #1 to #9 are 1549, 539, 947, 858, 817, 753, 544, 725, and 1451 respectively.

3. Proposed method

When battery capacity recovers, a local mutation point occurs during the overall degradation process of LIBs. The correlation between battery capacity and health indicators is nonlinear. To address the above-stated

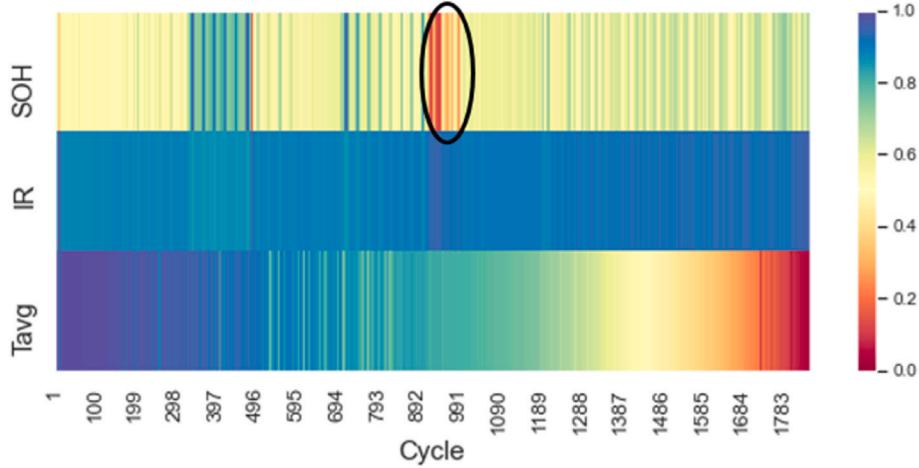


Fig. 3. Heatmap of Li-ion battery #1.

Table 1
Li-ion battery technical specifications.

Specifications	Value
Nominal capacity	1.1Ah
Nominal voltage	3.3 V
Electrolyte	carbonate based
Anode material	Iron phosphate (LFP)/graphite
Cathode material	LiFePO ₄
Ambient temperature	30 °C

Table 2
The operating parameters of battery cells.

Battery	Charging policy	Capacity rated	Cell barcode	Cycle life
#1	3.6C(80%)-3.6C	1.1Ah	EL150800460486	2158
#2	3.7C(31%)-5.9C	1.1Ah	EL150800737366	772
#3	4.8C(80%)-4.8C	1.1Ah	EL150800737377	1642
#4	5.3C(54%)-4C	1.1Ah	EL150800737378	1315
#5	5.6C(19%)-4.6C	1.1Ah	EL150800737294	1267
#6	5.6C(36%)-4.3C	1.1Ah	EL150800737280	1115
#7	5.9C(15%)-4.6C	1.1Ah	EL150800737320	858
#8	5.9C(60%)-3.1C	1.1Ah	EL150800737285	1002
#9	5C(67%)-4C	1.1Ah	EL150800737334	1935

limitations, We proposed a novel unique improved TFT model by replacing the LSTM encoder-decoder layers with Bi-LSTM encoder-decoder layers. We used a flexible and powerful hyperparameter search method BO-TPE to obtain optimal hyperparameters for better RUL prediction accuracy. Fig. 6 illustrates the detail procedure framework for RUL cycle life prediction. It should be noted that the data after the RUL prediction starting point for all batteries represent unknown data in the research problem.

3.1. Data preprocessing

The data noornalization process reduce the noise impact of the collected data for better forecasting, and this also solves the problem of inconsistency in training speeds and predictability and speed up the optimization algorithm's convergence during hyperparameter optimization. The normalization formula is as follows:

$$N_i = \frac{X_a - X_{min}}{X_{max} - X_{min}} \quad (5)$$

where the normalized data is N_i , the overall sample minimum value is X_{min} , the overall sample maximum value is X_{max} and X_i denotes the value that is being normalized.

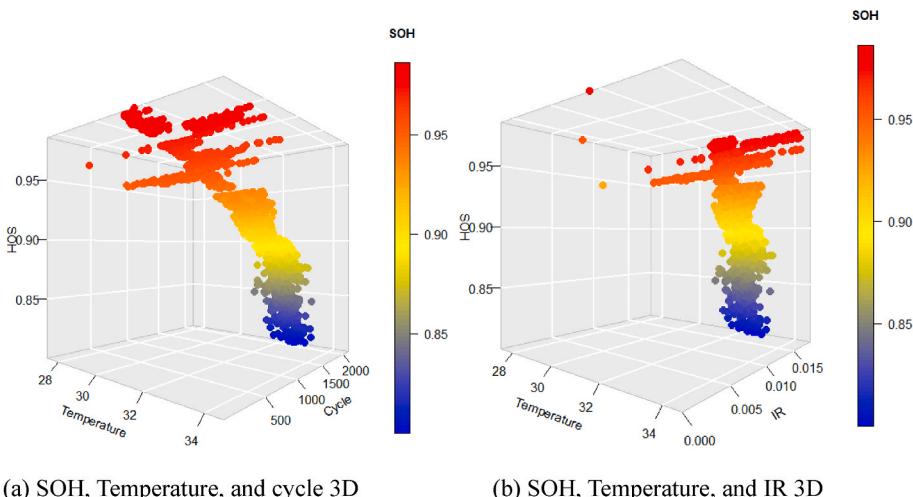


Fig. 4. A 3D plot of battery #1.

Table 3

Correlation analysis results of the health indicator (SOH).

Battery	#1	#2	#3	#4	#5	#6	#7	#8	#9
Cycle	-0.92	-0.97	-0.89	-0.87	-0.87	-0.84	-0.87	-0.82	-0.90
IR	-0.22	-0.94	-0.46	-0.92	-0.88	-0.89	-0.81	-0.89	-0.63
Temperature	-0.34	-0.84	-0.08	-0.45	-0.75	-0.67	-0.27	-0.52	-0.61

Note: IR = internal resistance.

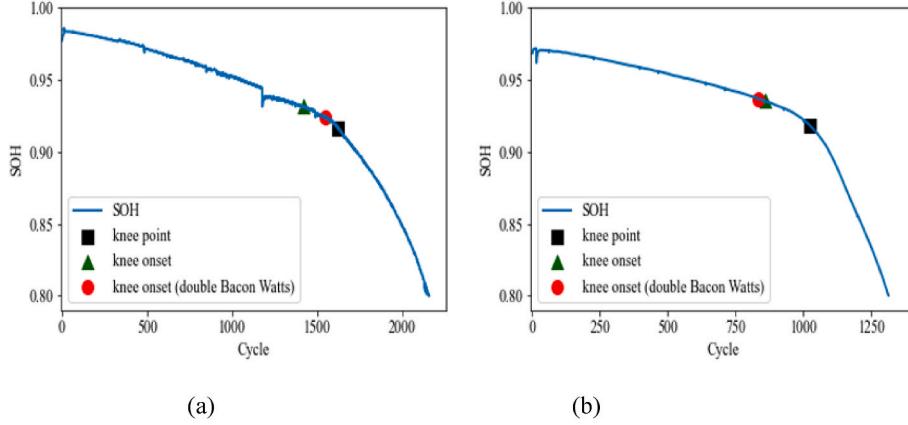


Fig. 5. Knee-onset estimation plots for (a) battery #1; (b) battery #4.

3.2. Improved temporal fusion transformer

In multi-horizon forecasting, the TFT is an attention-based DNN architecture that is fast and interpretable [33]. TFT utilizes specialized building blocks to select relevant features and remove unused components, enabling high performance over various tasks. The attention-based DNN integrates LSTM into its architecture to forecast time series over multiple horizons. Using Bi-LSTM and interpretive layers of self-attention, the algorithm can perform local and long-term processing better. Our ITFT model replaces the LSTM with a Bi-LSTM to learn long-term dependence more efficiently. TFT and many improved TFTs offer improved performance with greater interpretability over other DL-based models [33,37]. Several modules with different functions construct features for different types of input data in the TFT model, such as the gating method, variable selection networks, temporal processing methods consisting of the Bi-LSTM and multi-headed attention layers, and the static covariate encoders. Using gating architecture allows ignoring any unnecessary components, allowing networks to select variables at each time step and specify inputs. The sequence-to-sequence or Bi-LSTM layer handles local processing, while the interpretable multi-head attention block handles long-term dependency. In contrast to traditional DNNs, which overfit and predict features with irrelevant targets at each time step, variable selection networks identify important input features at each time step, thus improving the model's resiliency to changing inputs. Encoding static features facilitates modeling temporal dynamics by improving the prediction of the target.

In our work, we create a static variable call battery id (e.g. Battery_1). We applied a quantile prediction approach for prediction interval estimation. The task of function φ in Eq. (6) is to concurrently forecast the SOH at each time index point until SOH reaches a failure threshold 0.80. $[y_{s,t}, \dots, y_{s,t+\rho}, \dots, y_{s,t+T_0-1}]$, where s is the static covariate, and T_j represents the time-varying variables. Fig. 7 illustrates the framework of the proposed ITFT model. Two categories include the time-related covariates $\omega_{s,t} = [X_{s,t}, P_{s,t}]$, where $X_{s,t}$ represents time-varying known inputs such as the time index (cycle), and $P_{s,t}$ denotes the past inputs only known at the current time but unknown in the future (example: SOH, average temperature, and internal resistance). First, the

time-related input features $P_{s,t-T_0}, \dots, P_{s,t-1}$ and $X_{s,t-T_0}, \dots, X_{s,t+T_0-1}$ were fed into two input feature interpretation modules without shared parameters. The transformed inputs, $\tilde{P}_{s,t-T_0:t-1}$ and $\tilde{X}_{s,t-T_0:t+T_0-1}$ are then fed into multi-layer Bi-LSTM encoders and decoders, respectively. The outputs of the encoders and decoders of the last layer integrate into the multi-step fusion module, which assigns weights to the information at each time point based on its importance in learning the following temporal dependencies. The predicted SOH quantile values obtain by using minimizing the quantile loss function. In Fig. 7, the modules with the same color have the same parameters, which is important to note (see the details of Bi-LSTM in Fig. 8 and GRN in Fig. 9).

$$\hat{y}_{s,t+\gamma} = \varphi(\gamma, s, \mathbf{P}_{s,t-T_0:t-1}, \mathbf{X}_{s,t-T_0:t+\gamma}; \theta) \quad \gamma \in \{1, 2, \dots, T_o - 1\} \quad (6)$$

The expression $\hat{y}_{s,t+\gamma}$ represents the forecasted s -th sample quantile of the γ -step-ahead prediction at time t , where φ denotes the prediction model. We used “Adam” as the optimizer and rectified linear unit (RELU) for activation function. A weight vector δ_t indicates how important input variables are to the output. Apart from the Bi-LSTM encoder decoder layer all other modules of TFT and their mathematical expressions are from Lim et al. [33] and fully detailed there for reference. The hyperparameter configuration directly impacts the performance of a machine learning or deep learning model.

3.2.1. Bi-LSTM

We use Bi-LSTM to improve the learning efficiency of long-term dependencies memory. As an improvement on the traditional LSTM model, this model involves forward and backward LSTMs, which focus on left-to-right and right-to-left sequence data [38]. Bi-LSTM has two hidden layers; the first is an LSTM forward layer, and the second is an LSTM backward layer. A Bi-LSTM computes the forward hidden sequence \vec{h}_t , the backward hidden sequence \overleftarrow{h}_t and the output sequence y_t by iterating the backward layer and then updating the output layer.

$$\vec{h}_t = \text{LSTM}(x_t, h_{t-1}^-) \quad (7)$$

$$\overleftarrow{h}_t = \text{LSTM}(x_t, h_{t-1}^+) \quad (8)$$

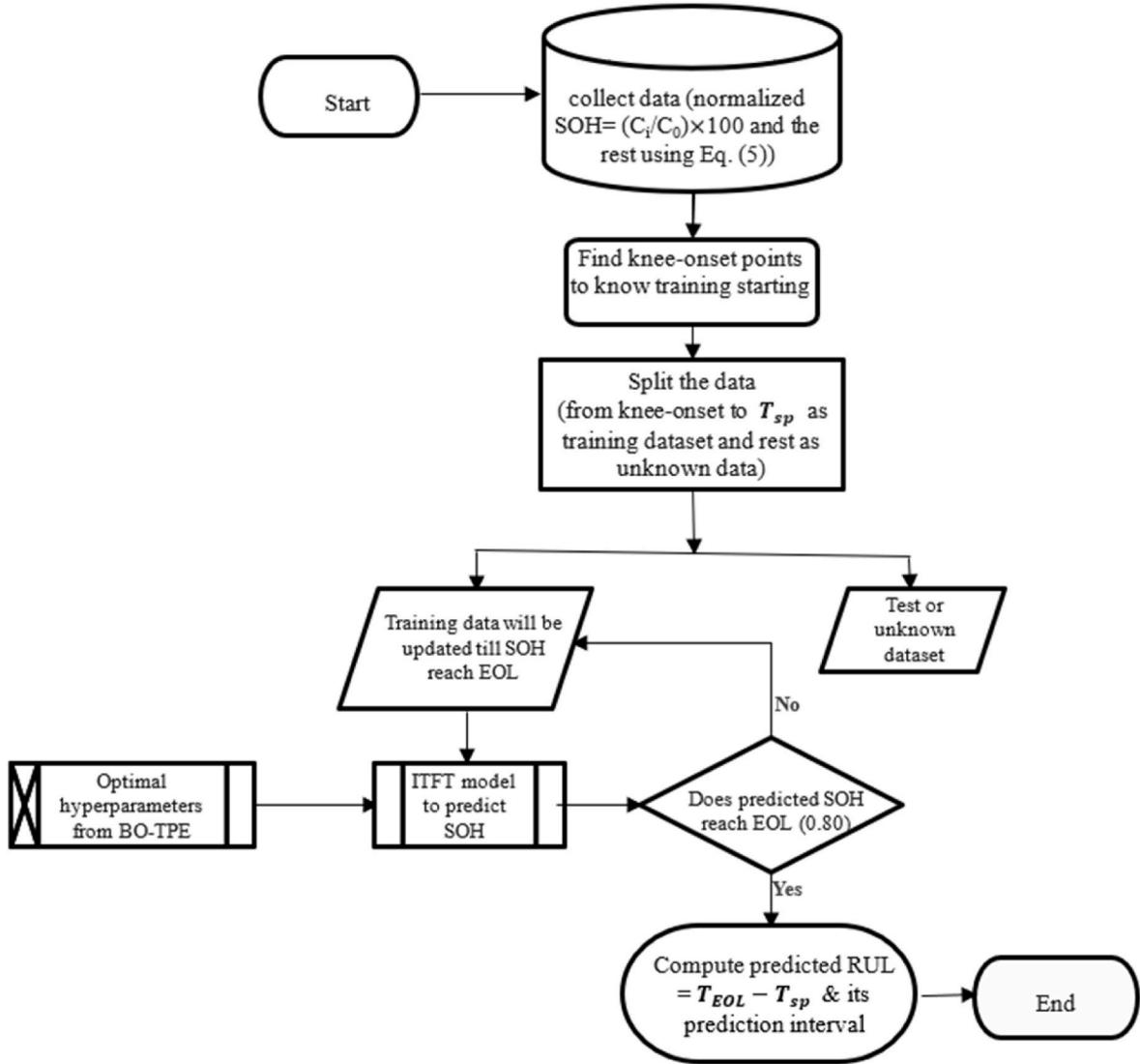


Fig. 6. Proposed method for online RUL prediction.

$$y_t = W_{\overrightarrow{h}_y} \overrightarrow{h}_t + W_{\overleftarrow{h}_y} \overleftarrow{h}_t + b_y \quad (9)$$

The LSTM updating procedure is in Eqs. (7)–(9), where $W_{\overrightarrow{h}_y}$ and $W_{\overleftarrow{h}_y}$ are represent the weight parameters of the forward LSTM cell and the backward LSTM cell and b_y as the bias of the output.

3.2.2. Gated residual network (GRN)

In online prediction, the relationship between independent variables and the targets is unknown, making it difficult to ascertain which are essential. GRN gives flexibility to the model to use non-linear processing where it is necessary only. The network can alter its depth and complexity using the gating mechanisms to perform effectively on various datasets and tasks.

$$\text{GRN}_{\omega}(a, c) = \text{LayerNorm}(a + \text{GLU}_{\omega}(\eta_1)) \quad (10)$$

$$\eta_1 = W_{1,\omega}\eta_2 + b_{1,\omega} \quad (11)$$

$$\eta_2 = \text{ELU}(W_{2,\omega}a + W_{3,\omega}c + b_{2,\omega}) \quad (12)$$

Intermediate layers are $\eta_1 \in R^{d_{\text{model}}}$ and $\eta_2 \in R^{d_{\text{model}}}$, the standard normalization layer called LayerNorm [39] and the gated linear units (GLUs) control the model's unwanted components. ELU is the

exponential linear unit [40]. GLU mathematical formula is as follows:

$$\text{GLU}_{\omega}(\mu) = \sigma(W_{4,\omega}\mu + b_{4,\omega}) \odot (W_{5,\omega}\mu + b_{5,\omega}) \quad (13)$$

where $\mu \in R^{d_{\text{model}}}$ and $\sigma(\cdot)$ is the sigmoid activation function. The weight and the biases of the model are $W(\cdot) \in R^{d_{\text{model}} \times d_{\text{model}}}$ and $b(\cdot) \in R^{d_{\text{model}}}$, respectively. d_{model} is the hidden state size across ITFT, and \odot is the Hadamard product. The outputs of GLU can be all close to 0, GLU helps the model control the nonlinear impact of the GRN's contribution to the original input a as much as it can.

3.2.3. Variable selection network

As hyperparameter tuning in many DL models help to improve prediction accuracy, the same is for feature selection. Knowing relevant inputs and their impact on output is helpful for modeling and decision-making. ITFT applied variable selection networks to select important variables for prediction and remove noisy inputs that could negatively affect performance [33]. Because not all inputs, especially in battery dataset have a significant impact on SOH degradation. Given the flattened vector of all past inputs at period t can be represented as $[\cdot]_t = [\xi_t^{(1)^T}, \dots, \xi_t^{(m_x)^T}]^T$ and ξ_t^j denote the transformed input of j th variable. Weights of feature selection V_{xt} in Eq. (14) are obtain by inputting $[\cdot]_t$

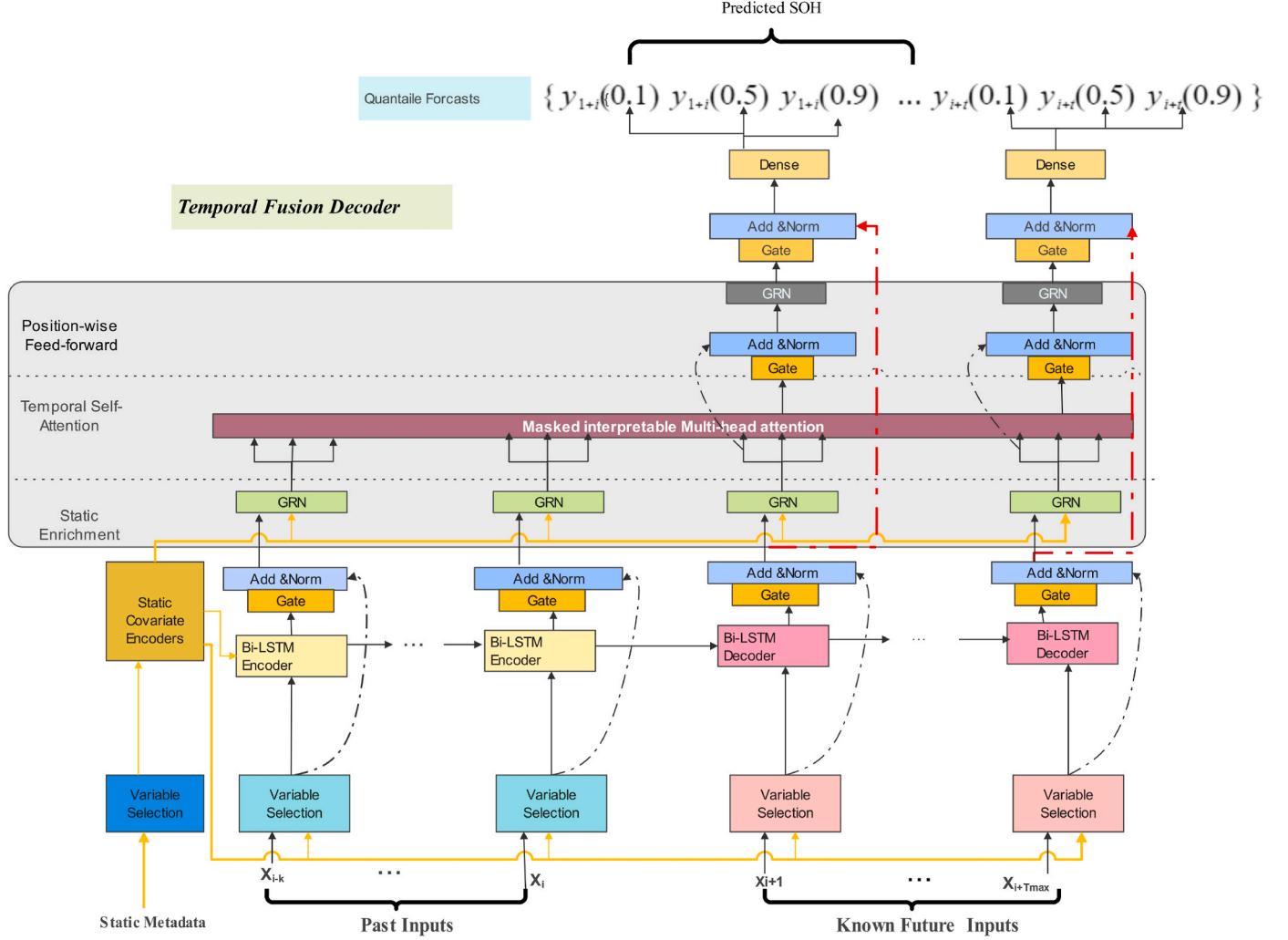


Fig. 7. The architecture of the ITFT.

and C_s into GRN follows by softmax layer. Its GRN nonlinearly calculate each ξ_t^j is $v_{xt}^{(j)}$ is the j th observation of the vector V_{xt} .

$$V_{xt} = (GRN_{V_x}([\square]_t C_s)) \quad (14)$$

$$\tilde{\xi}_t^{(j)} = GRN_{\tilde{\xi}^{(j)}}(\xi_t^j) \quad (15)$$

$$\tilde{\xi}_t = \sum_{j=1}^{m_x} v_{xt}^{(j)} \tilde{\xi}_t^{(j)} \quad (16)$$

3.2.4. Interpretable multi-head attention module

A multi-timestep fusion module with an interpretable multi-head attention mechanism [33] is to learn long-term associations in battery capacity RUL prediction across several time steps. A sequence-to-sequence layer Bi-LSTM handles local processing, while a unique interpretable multi-head attention block handles long-term dependencies. The concept of "multi-head attention" was first put forth in Ref. [41] to help improve the learning capacity of the standard attention mechanism, and it was refined in Ref. [33] to share values across multiple heads and use additive aggregation, as seen in Eqs. 17–20. Based on the relative importance of various time steps, it can balance the outputs of all encoders and decoders at the time steps before the present forecast time step. The input matrix X multiplies the linear transformation matrices W_Q , W_K , and W_V to produce Q , K , and V , using a linear transformation by self-attention mechanism. As a result, each head

interpretable multi-headed attention shares the matrix V . In this work, given h_n is the number of heads and $W * h$ changes for each head, interpretable multi-head attention must be carried out h_n times. The mathematical expressions of interpretable multi-head attention are as follows:

$$Attention(Q, K, V) = A(Q, K)V \quad (17)$$

$$A(Q, K) = softmax \left(\frac{QK^T}{\sqrt{n}} \right) \quad (18)$$

$$Interpretable\ Multi - Head(Q, K, V) = \tilde{H} \tilde{W}_H \quad (19)$$

$$\tilde{H} = A(Q, K) VW_V = \frac{1}{h_m} \sum_{h=1}^{h_m} A(QW_Q^h, KW_K^h) VW_V \quad (20)$$

where n is the dimension of the vector corresponding to K , $A(\cdot)$ is the normalized function and $h = 1, \dots, h_m$. The weight matrices for the h th head of Q and K are W_Q^h and W_K^h , respectively. The weight matrix of the final linear mapping is \tilde{W}_H , and W_V is the weight matrix of V shared by all heads.

3.2.5. Loss functions and quantile forecasts

Similar to TFT [33], ITFT generates prediction intervals based on point predictions. The linear transformation produced by the temporal fusion decoder from the output is used to forecast quantiles as in Eq. (22)

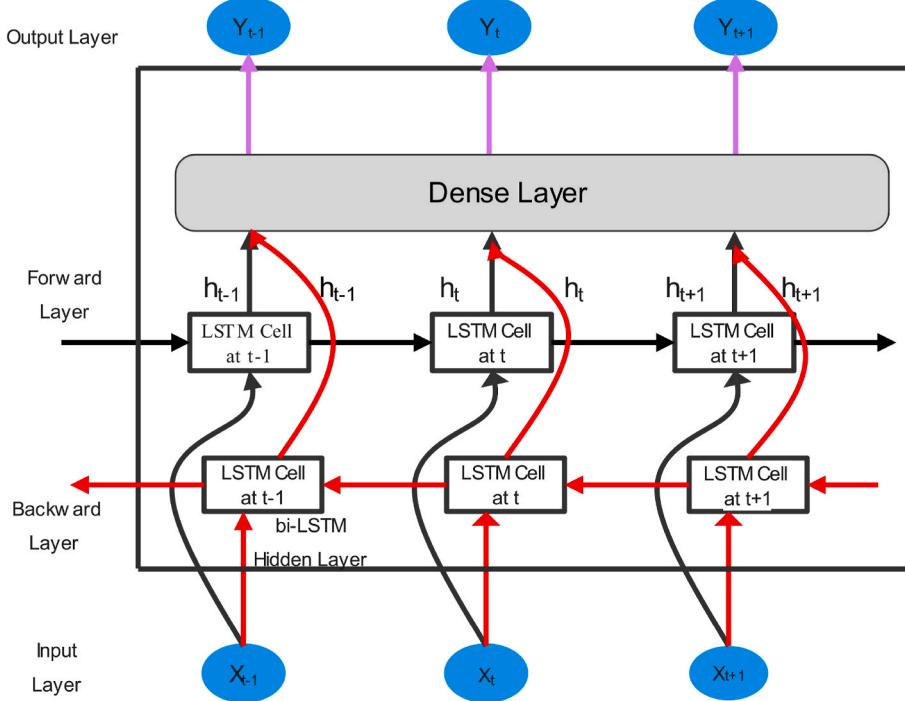


Fig. 8. Structure of the Bi-LSTM model.

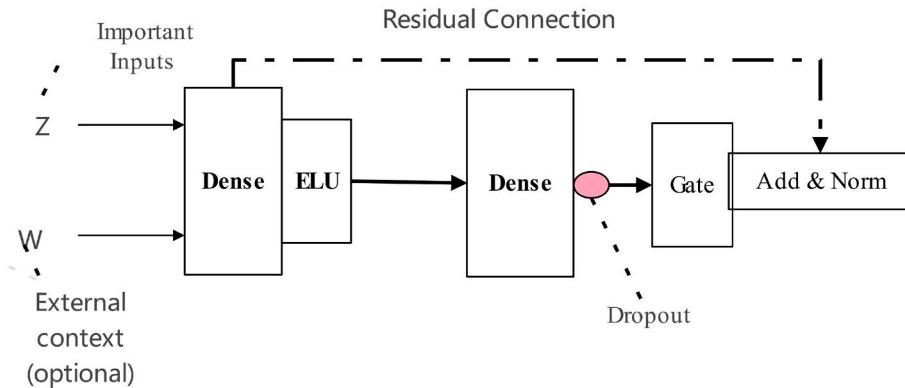


Fig. 9. Gated residual network architecture.

and predicts different percentiles (such as the 10th, 50th, and 90th) simultaneously at each time step. We apply the percentiles mentioned above. The joint quantile loss function utilized in the standard quantile prediction is the average value of all quantile losses under a set of probabilities as expressed in Eq. (23).

$$\hat{y}(q, t, \tau) = W_q \tilde{\psi}(t, \tau) + b_q \quad (21)$$

$$L(\Omega, W) = \sum_{y_t \in \Omega} \sum_{q \in Q} \sum_{\tau=1}^{\tau_{max}} \frac{QL(y_t, \hat{y}(q, t - \tau, \tau), q)}{M_{\tau_{max}}} \quad (22)$$

$$QL(y, \hat{y}, q) = q(y - \hat{y}) + q(1 - q)(\hat{y} - y) + q\text{-Risk} = \frac{\sum_{y_t \in \Omega} \sum_{\tau=1}^{\tau_{max}} QL(\hat{y}(q, t - \tau, \tau), q)}{\sum_{y_t \in \Omega} \sum_{\tau=1}^{\tau_{max}} |y_t|} \quad (23)$$

where W_q and b_q are linear coefficients for the given quantile q . The parameter Ω is the domain of training data containing M samples. Q are

the output quantiles ($Q = (0.1, 0.5, 0.9)$) for this paper, W denotes the weights of ITFT, and $(.)_+$ is equal to $\max(0, .)$. We assess the normalized quantile losses over the whole predicting horizon for out-of-sample testing or online future prediction.

3.2.6. Bayesian optimization based on tree-structured parzen estimator (BO-TPE)

Having good machine learning to achieve better prediction accuracy involves finding the model's ideal hyperparameters. In this paper, to automatically optimize the hyperparameters of the proposed model of both offline and online prediction, we implement a Tree-structured Parzen Estimator, a variant of BO. Table 4 shows the hyperparameters importance rank to model performance and Table 5 shows the optimal hyperparameters. Hyperparameter importance helps us better understand the relative importance of each hyperparameter on the results of the trial as shown in Fig. 10 for batteries #1, #2, and #3. Among many parameters of can greatly affect online prediction, lookback window or maximum encoder length and maximum prediction step deeply affect multi-step ahead rolling prediction performance. Due to their effects on

Table 4

Hyperparameters importance to model performance.

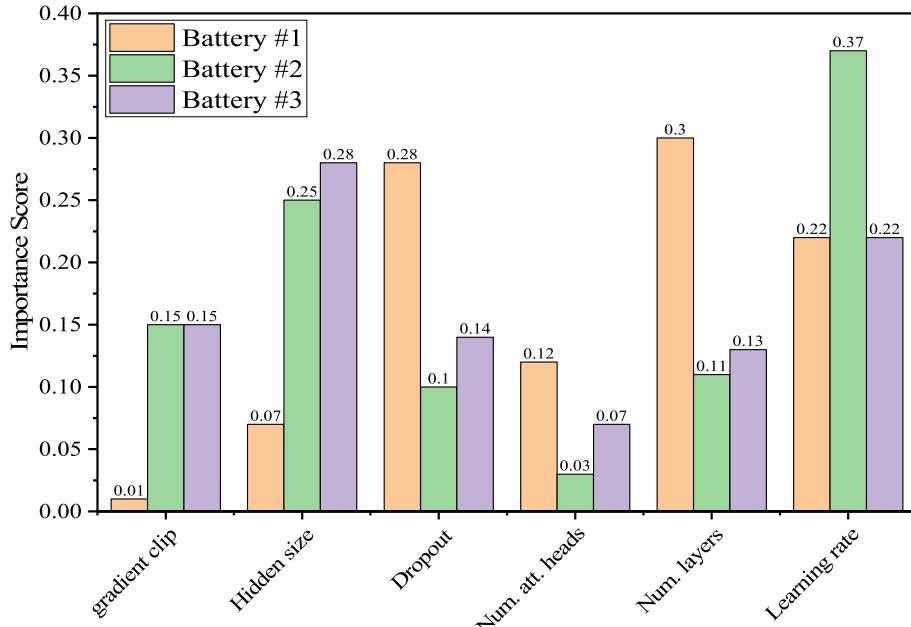
Hyperparameter	#1	#2	#3	#4	#5	#6	#7	#8	#9
Gradient clip	0.01	0.15	0.15	0.11	0.05	0.14	0.36	0.16	0.13
Hidden size	0.07	0.25	0.28	0.60	0.11	0.19	0.16	0.36	0.30
Dropout	0.28	0.10	0.14	0.14	0.67	0.40	0.08	0.11	0.10
Num. att. heads	0.12	0.03	0.07	0.01	0.10	0.02	0.03	0.05	0.01
Num. layers	0.30	0.11	0.13	0.05	0.04	0.01	0.02	0.10	0.09
Learning rate	0.22	0.37	0.22	0.09	0.02	0.25	0.35	0.21	0.37

Table 5

Hyperparameters used in the model for SOH and RUL prediction.

Hyperparameter	#1	#2	#3	#4	#5	#6	#7	#8	#9
Gradient clip	0.2	0.03	0.5	0.07	0.3	0.2	0.02	0.8	0.1
Hidden size	94	162	64	205	145	53	57	85	41
Dropout	0.4	0.2	0.3	0.1	0.1	0.4	0.2	0.4	0.3
Hidden cont. size	87	73	36	129	110	51	56	36	32
Num. att. heads	2	3	1	3	2	2	2	1	4
Num. layers	4	4	2	2	2	4	4	2	1
Learning rate	0.008	0.002	0.02	0.008	0.002	0.003	0.05	0.02	0.02
Batch size	32	32	8	32	44	16	32	32	32

Note: Hidden cont. size = Hidden continuous size, Num. att. Heads = number of attention heads, Num. layers = number of layers.

**Fig. 10.** Hyperparameter importance ranking score during model tuning for batteries #1, #2 and #3.

the multistep-ahead rolling prediction bias cumulation problem and the ITFT capture of the time series relationship, the maximum encoder and maximum prediction steps are sensitive to performance. From [Table 4](#) and [Fig. 10](#), we can observe that hidden size and learning rate, on average, provide a greater weight to the model performance during model training.

We selected seven hyperparameters for tuning. Another typical BO surrogate model is the Tree-structured Parzen estimator (TPE) [42]. TPE is run on the Optuna library, TPE efficiently explores huge areas and eliminate inconsistent trials to yield quicker results by the use of an early stopping role to avoid overfitting problem for model fine-tuning in every trial or iteration. TPE has recently been proposed to address the limitations of conventional BO approaches in dealing with categorical and conditional parameters, thus improving hyperparameter selection. To serve as the generative models for all domain variables, BO-TPE generates two density functions, $l(x)$ as bad and $g(x)$ as good samples. The

observation results are split into good and bad outcomes by a pre-determined percentile y^* in order to apply TPE. The two sets of data are then modeled using straightforward Parzen windows as follows:

$$P(x|y) = \begin{cases} Pr(g(x)), & \text{if } y < y^* \\ Pr(l(x)), & \text{if } y \geq y^* \end{cases} \quad (24)$$

where $Pr(g(x))$ and $Pr(l(x))$ are probabilities of good and bad groups respectively. y is the validation loss and a predefined threshold as y^* . optimization process ends if the evaluation metric proposed is minimize or to total trial or iteration is finish. [Fig. 11](#) shows the graphical display of optimal hyperparameters with the top ten best trials for battery #2.

4. Analysis results

We propose both an offline SOH estimation and online RUL life

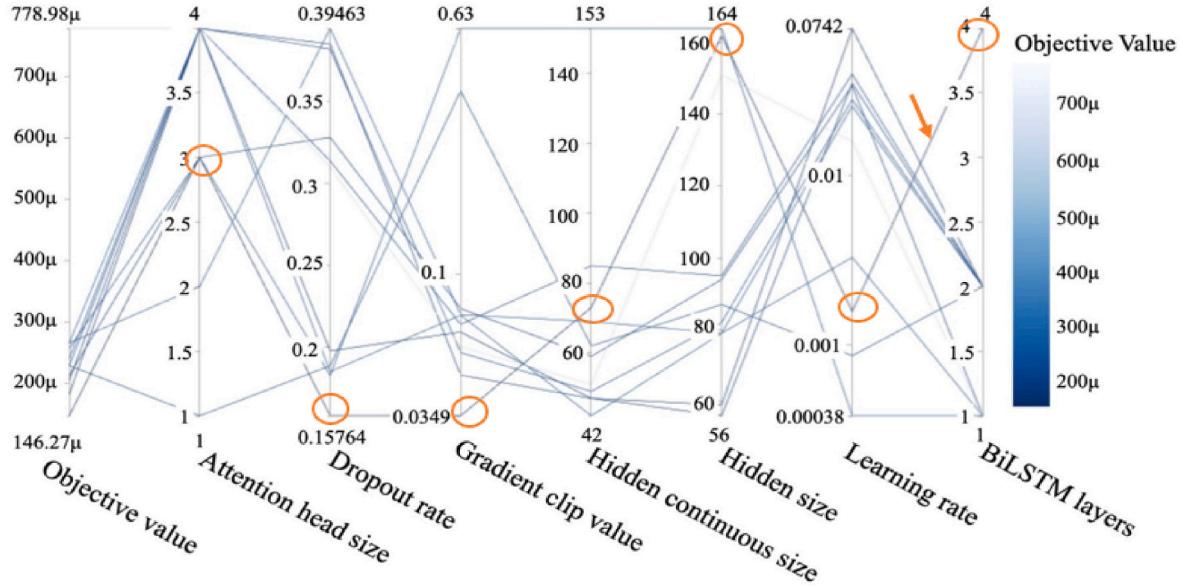


Fig. 11. Parallel coordinate plot during model tuning of hyperparameters for ITFT.

prediction for the nine batteries under different charging conditions to evaluate our model's stability, feasibility, and robustness in handling different types and data characteristics. Python was used to create all of the analyses for this study. With the PyTorch library [43], the system operates in the Spyder 3.7.3 environment.

We adopt the double Bacon Watts models to determine the starting point for training dataset and decide on the prediction starting point at $SOH = 0.86$; all from batteries #1 to #9 are said to reach EOL when their SOH falls to 80% of rated capacity. In this study, the training length represents the number of cycles at the knee-onset to the number of cycles at the RUL prediction's starting point.

4.1. Evaluation metrics

We assessed the performance of our proposed model using a number of evaluation measures to determine its precision and efficacy in predicting capacity flow and the RUL of the batteries. The root mean square error (RMSE), mean absolute error (MAE), absolute error (AE) and relative error (RE) are used to evaluate the offline and online future prediction performance of our proposed method in this study. These formulas are given as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (SOH_i - \widehat{SOH}_i)^2} \quad (25)$$

$$AE = |RUL_a - \widehat{RUL}_p| \quad (26)$$

$$RE(\%) = \frac{|RUL_a - \widehat{RUL}_p|}{RUL_a} \times 100 \quad (27)$$

where i denotes the cycle number, and SOH_i and \widehat{SOH}_i stand for the observed SOH and predicted SOH, respectively. \widehat{RUL}_p represents the predicted RUL value and RUL_a denotes the actual RUL value.

To facilitate a more detailed comparison of the model's performance, we introduce the improvement rate (IR), which provides a more precise basis for evaluating the relative performance of the two models. The IRs of model A and model B, concerning RMSE, can be calculated using the following formula:

$$IR_{RMSE} = \frac{RMSE_A - RMSE_B}{RMSE_A} \times 100 \quad (28)$$

where $RMSE_A$ represents the RMSE of any competing model, and $RMSE_B$ is the RMSE of our proposed model.

4.2. Model performance for known SOH prediction

This subsection deals with the SOH estimation analysis. TFT (original with LSTM layer), DeepAR, and Bi-LSTM-AM were compared with the proposed model to evaluate prediction performance. Table 5 shows the hyperparameters used for SOH and RUL prediction.

Table 6 shows the prediction accuracy of the four models regarding RMSE for four batteries with different charging policies. Results demonstrate that the proposed model for the test data has the lower RMSE for batteries #2, #8, and #9, while Bi-LSTM-AM has the lowest RMSE for battery #1. On average, the proposed model outperforms the other models, followed by the Bi-LSTM-AM model. Fig. 12 also shows the SOH prediction performance and their prediction errors for test data of four batteries ranging from the prediction starting at $SOH = 0.86$ to the EOL of each battery. Fig. 12 shows that the proposed model gives a better performance. From Fig. 12 (right-side) we can see closer predicted SOH values to actual SOH as compare to the competing model. The error plots in Fig. 12 (left-side) depicts closer errors to zero illustrating better prediction.

Table 6
Prediction performance for the test data using different models.

Battery	Training size	Test size	Models			
			TFT	DeepAR	Bi-LSTM-AM	ITFT
RMSE/ IR_{RMSE}						
#1	484	255	0.0050 64.71%	0.0101 82.18%	0.0016 -11.11%	0.0018
#2	353	130	0.0053 65.05%	0.0045 47.93%	0.0040 40.81%	0.0023
#8	220	57	0.0050 74%	0.0039 66.67%	0.0049 73.45%	-
#9	467	195	0.0061 59.02%	0.0034 26.47%	0.0057 56.14%	0.0025
R^2						
#1	484	255	0.939	0.754	0.988	0.979
#2	353	130	0.922	0.945	0.985	0.985
#8	220	57	0.938	0.964	0.942	0.989
#9	467	195	0.910	0.972	0.922	0.983

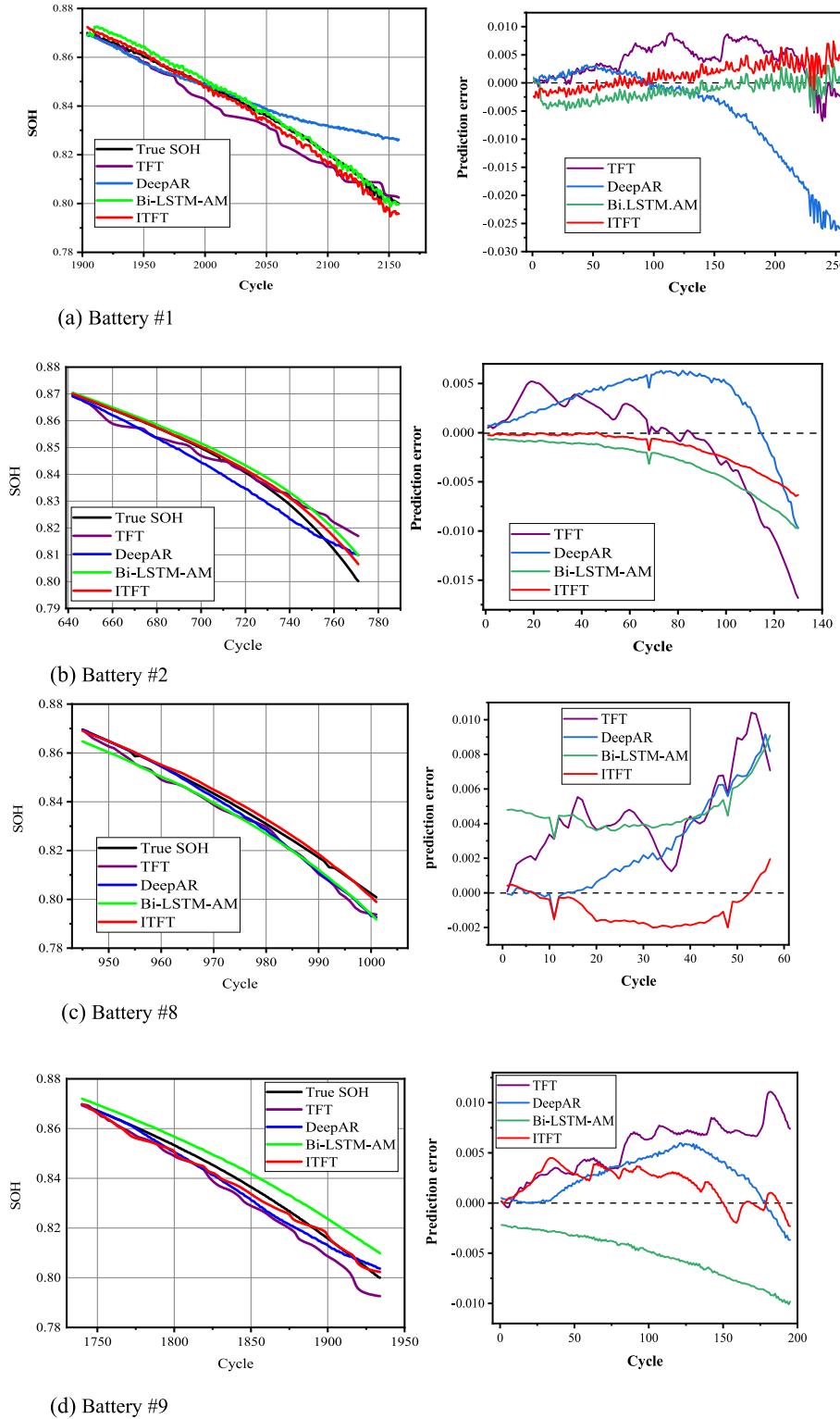


Fig. 12. SOH prediction results (left side), and the error details (right side) of the proposed method with three other models for test data.

Except for BiLSTM-AM in battery #1, the proposed model exhibited RMSE values of 64.71% and 82.18%, which are superior to those of TFT and DeepAR, respectively. For battery #2, our proposed model exhibited 65.05%, 47.93%, and 40.81% improvements in RMSE against TFT, DeepAR, and BiLSTM-AM, respectively. The same trend of improvement can also be seen for battery #8 and #9. These prediction results demonstrate that our proposed method excels in RMSE, as evidenced by the IR results. The proposed model exhibits strong generalization and

robustness, attributed to its utilization of diverse battery datasets. In experiments with batteries used for SOH estimation, our proposed method consistently achieves maximum errors below 1% and demonstrates improved accuracy compared to competing models. From Tables 6 and 7, the proposed model has shown superiority over three competing models and some published SOH prediction models based on RMSE. Fig. 13 displays the ECDF based on absolute error. In Fig. 13 (b-d), the error line reaches the threshold level earlier compared to the

Table 7The RMSE and IR_{RMSE} of existing published models with the proposed model.

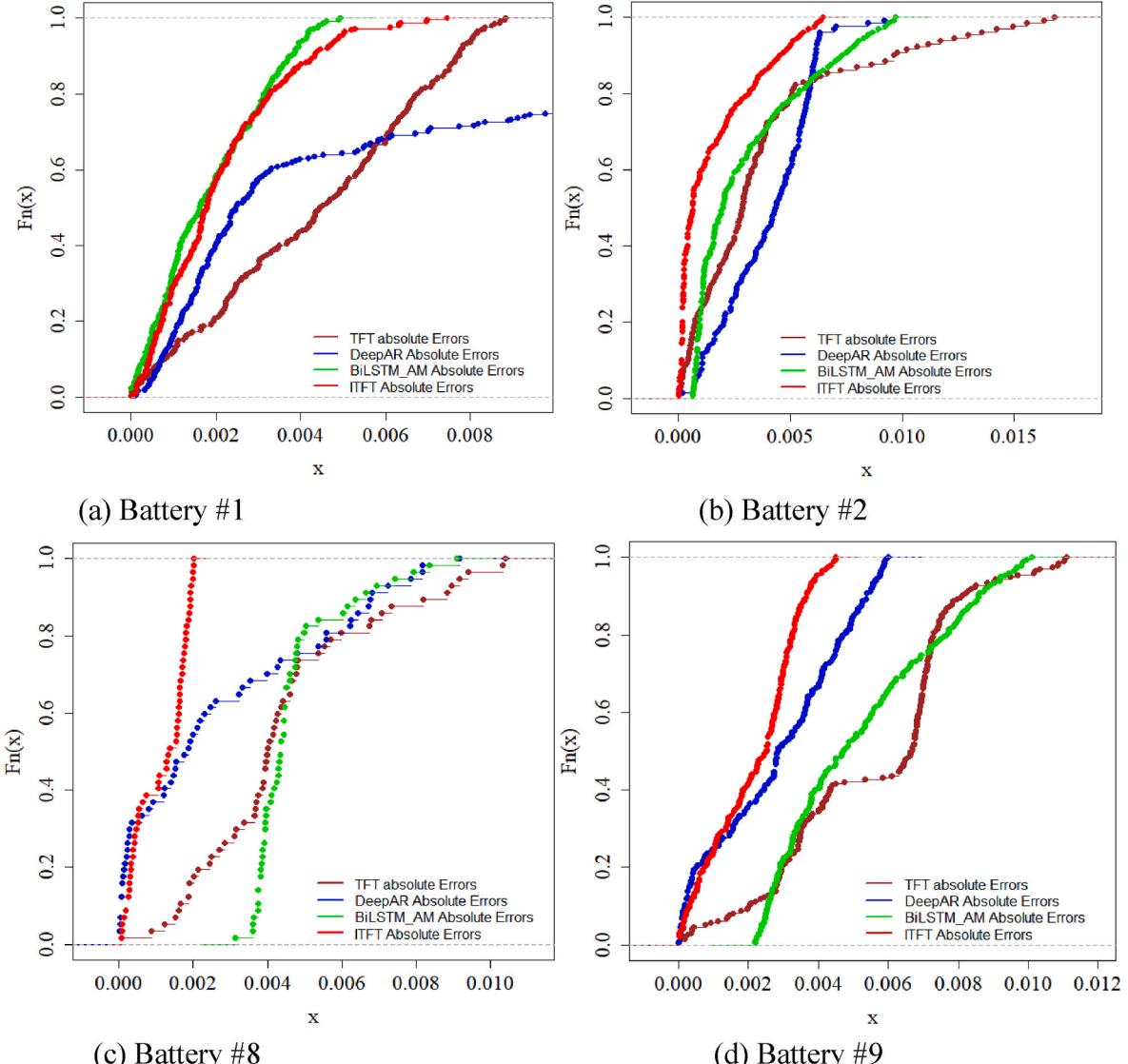
Model	Battery #1		Battery #8	
	RMSE	IR _{RMSE}	RMSE	IR _{RMSE}
Bi-LSTM-AM-TR [11]	0.0070	75.71%	–	–
CEEMD + H + GPR + LSTM + Error [25]	0.0020	15%	–	–
LRP-driven LSTM [44]	0.00634	73.19%	–	–
LSTM-LAM [45]	–	–	0.005	74%
Proposed model (ITFT)	0.0017	–	0.0013	–

competing models, indicating that the proposed ITFT model better captures random deviations. As a result, the error is smaller, and the prediction accuracy is higher.

To further verify the superiority and effectiveness of our proposed model, we compared our proposed algorithm with recent related literature using the same battery cells with a similar estimation approach. **Table 7** presents a comparison between our proposed method and four existing techniques—Bi-LSTM-AM-TR [11], CEEMD + H + GPR + LSTM + Error [25], LRP-driven LSTM [44], and LSTM-LAM [45] in terms of prediction performance for batteries #9 and #8. **Table 7** demonstrates

that our proposed ITFT method surpasses several previously published approaches in predicting SOH. For battery #1, the RMSE improvement accuracy values of the proposed ITFT were 75.71%, 15% and 73.19% better than those of Bi-LSTM-AM-TR [11], CEEMD + H + GPR + LSTM + Error [25], and LRP-driven LSTM [44], respectively.

A Kolmogorov-Smirnov predictive accuracy (KSPA) test [46,47] is analyzed to test the models' absolute errors' statistically significant difference. The strength of the KSPA test lies in its ability not only to differentiate the forecasted distributions of two models but also to ascertain whether the model exhibiting the lowest error also has a reduced random error compared to the alternative model. The test is a complementary statistical method for differentiating the predictive accuracy of two distinct sets of forecasts. For battery #1, the ITFT vs. Bi-LSTM-AM comparison yielded p-values of 0.9392 and 0.2987 for one-sided and two-sided tests, indicating no significant difference in their prediction performance. However, in all four experiments, the one-sided and two-sided p-values from the KSPA test for all competing models versus the proposed model were less than 0.01* based on a P-value of 0.05. It confirms significant statistical differences between the prediction model proposed in this paper and the three comparison models. The empirical cumulative distribution function (CDF) is a critical element in the KSPA test. The proposed ITFT model effectively

**Fig. 13.** SOH prediction of empirically determined cumulative distribution functions (c.d.f.) of errors for four batteries.

captures random deviations, resulting in smaller errors and higher prediction accuracy. The model's overall advantages and enhanced prediction accuracy are statistically significant, which is crucial for the performance of the recommendation algorithm.

4.3. Online RUL future prediction

Precise long-term prediction of RUL and robust uncertainty management are of significant importance for lithium-ion batteries. All batteries are considered to have reached their EOL when their capacity drops to 80%. This section focuses on future capacity prediction, to evaluate our proposed model's operability in real time application and prediction, utilizing a multi-step-ahead rolling forecast approach. The maximum prediction length (look forward window) and maximum

encoder length (lookback window size) are fixed at 15% and 16% of the training length respectively, and epoch set at 800, while the online look forward (future prediction time step) is 10. Using the relative error and absolute error, the performance of RUL prediction is obtained when the predicted SOH value reaches the EOL. Our study used double Bacon Watts method to determine the training starting point for RUL prediction. For example, Fig. 14(a) shows future prediction results for battery #1 with the Training starting point at 1549 to 1903 based on double Bacon Watts (left); the same applies for all batteries as shown in Fig. 14. The green dotted line in Fig. 14 represents the failure threshold (80%).

Double Bacon Watts knee-onset identification for the training starting point is very vital in SOH estimation, most especially RUL prediction. The knee onset helps to give new direction about the situation of the capacity fade, assisting to achieve a good RUL prediction. Table 8 shows

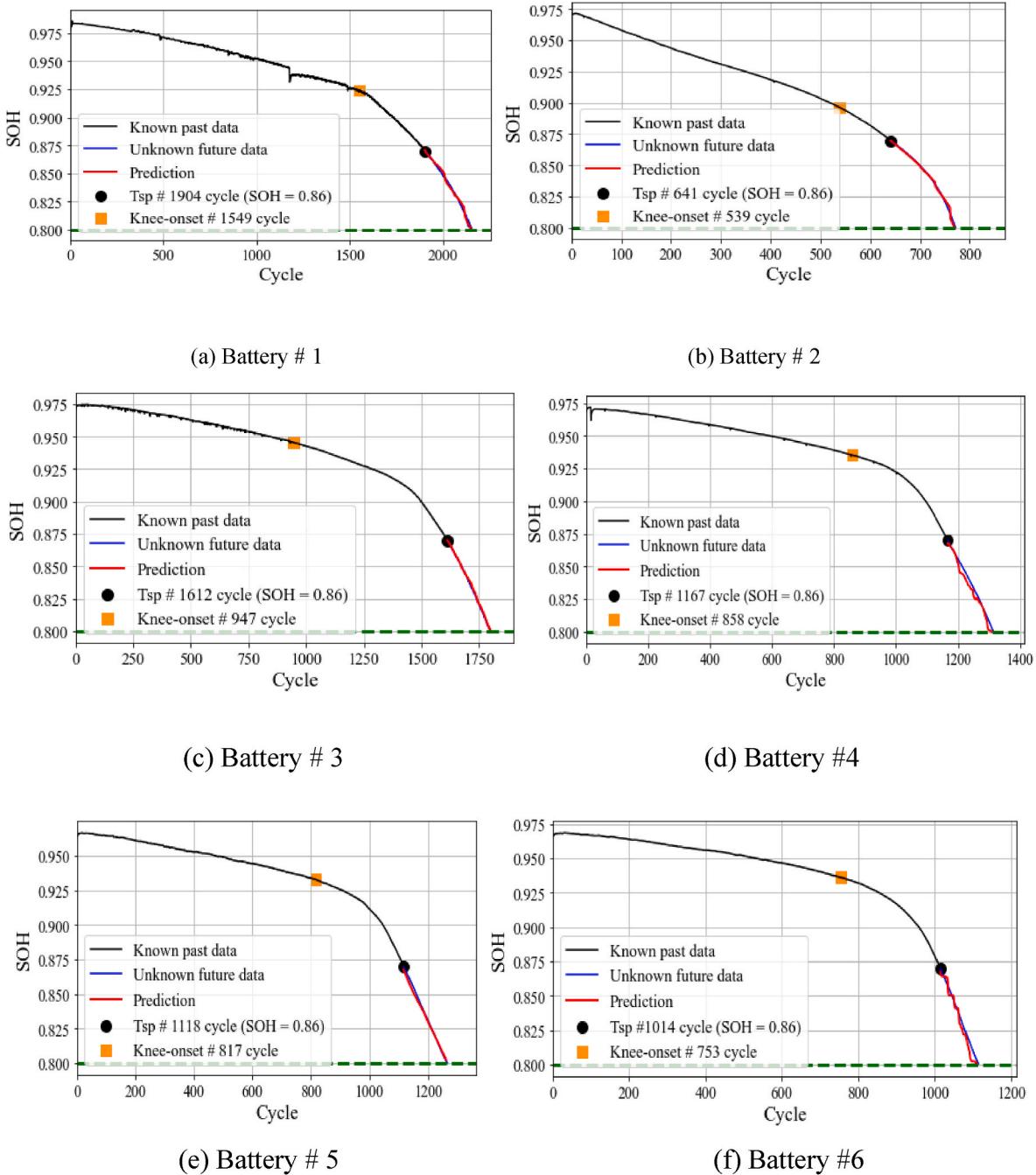


Fig. 14. RUL prediction results of the proposed model for nine Li-ion batteries with different charging policy.

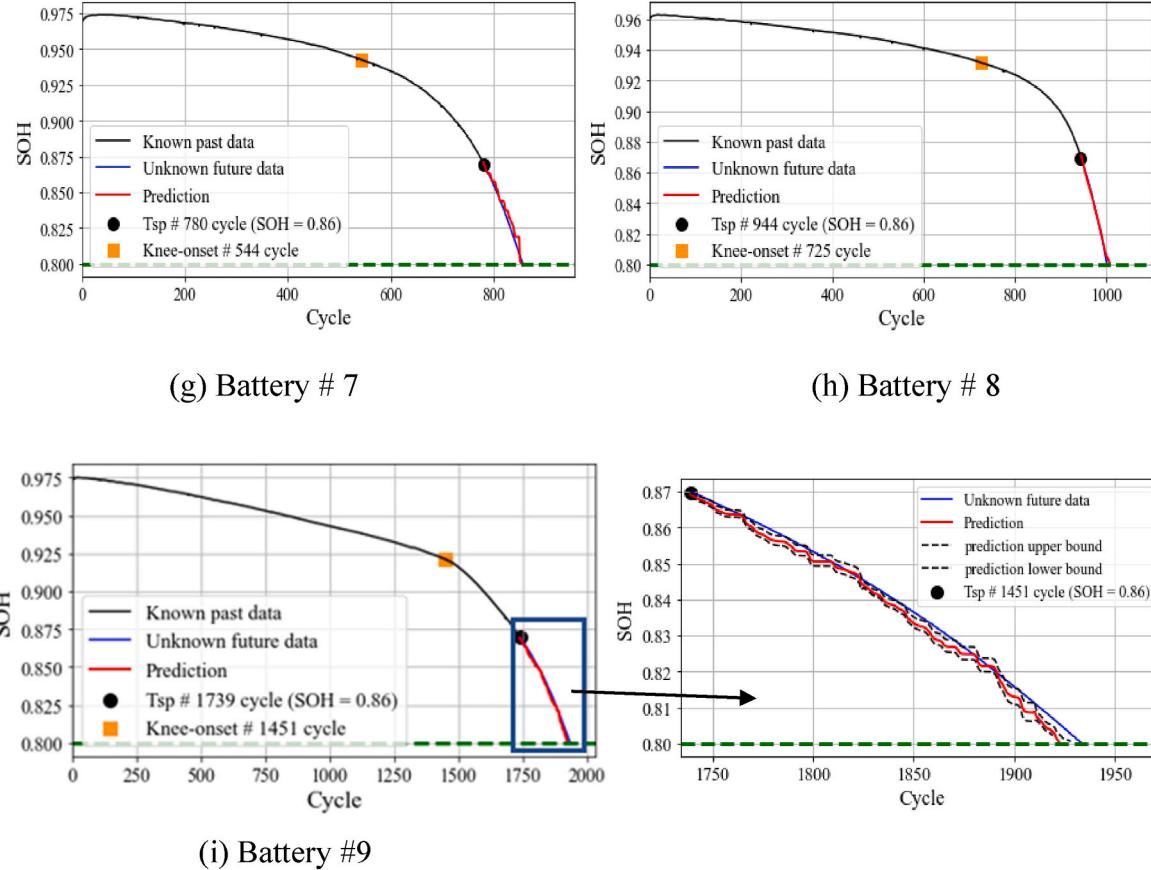


Fig. 14. (continued).

95% RUL prediction results based on each battery's training starting point (knee-onset). In contrast to many conventional data-driven RUL prediction methods, which typically provide only point RUL estimates, our proposed ITFT model captures and describes the uncertainty associated with RUL in the battery degradation process, as shown in Table 8. In many cases, it is difficult to pinpoint the exact mechanism of rapid change in degradation or isolate it experimentally. Therefore, providing a suitable training starting point within the rapid degradation trend helps with the exact data needed. Early warning RUL predictions are achieved based on the prediction results in Fig. 14, this can also be seen in Table 8.

To verify the robustness of the proposed method nine batteries with different charging policy are used. The quantile regression method generates 95% prediction intervals in Table 8 and also demonstrated in Fig. 14(i). Our proposed online RUL prediction model, on average, achieves a relative error of 1.79%. The results indicate that the proposed ITFT model can be effectively developed for predicting the RUL.

To gain a clearer understanding, we examine the process of how the

proposed model selects and evaluates time steps for its decision-making. Attention mechanisms and variable selection weights help one get a better glimpse into how deep learning models work on the original data stream. Therefore, this paper did an in-depth interpretability analysis during optimization and model training—hyperparameters' importance weight during model tuning to shown in Table 4 of section 3.2.6. Table 9 details the variable importance weight analysis results of batteries #1 to #6. For values greater than 0.1, indicated by *. Table 9 shows that static created variable battery ID and future known input time index (cycle) play a significant role in prediction performance. Variable importance analysis aids in comprehending the time-dependent relationships within a dataset. Fig. 15 shows mean attention weight (10%, 50%, and 90%) trends during online RUL prediction of four batteries in a heatmap. We can observe that Fig. 15(a) depicts a smooth persistent pattern over time index. Smooth persistent temporal patterns of attention weights distributed equally across all time indexes can indicate good RUL

Table 8
Online RUL prediction results using the proposed model.

Battery	Training length	Actual RUL	Predicted RUL	Prediction interval	AE	RE (%)
#1	1549–1903	255	252	[247–254]	3	1.18
#2	539–640	130	128	[124–129]	2	1.54
#3	947–1611	188	186	[183, 192]	2	1.06
#4	858–1166	148	144	[142–145]	4	2.70
#5	817–1117	149	148	[147, 149]	1	0.67
#6	753–1013	100	98	[95–100]	2	2.00
#7	544–779	77	74	[70–75]	3	3.90
#8	725–943	57	57	[54–58]	0	0.00
#9	1451–1738	195	189	[186–191]	6	3.08

Table 9
A mean variable importance weights for batteries # 1, # 2, # 3, #4, #5, and #6.

	#1	#2	#3	#4	#5	#6
Static covariates						
Encoder length	0.894*	0.130*	0.135*	0.119*	0.006	0.123*
Battery ID	0.106*	0.870*	0.865*	0.881*	0.994*	0.877*
Past inputs						
Temperature	0.478*	0.103*	0.171*	0.170*	0.030	0.141*
IR	0.124*	0.028	0.017	0.052	0.508*	0.774*
Time index (Cycle)	0.229*	0.731*	0.494*	0.588*	0.157*	0.042
SOH	0.169*	0.138*	0.317*	0.190*	0.305*	0.0.043
Future known inputs						
Time index (Cycle)	1*	1*	1*	1*	1*	1*

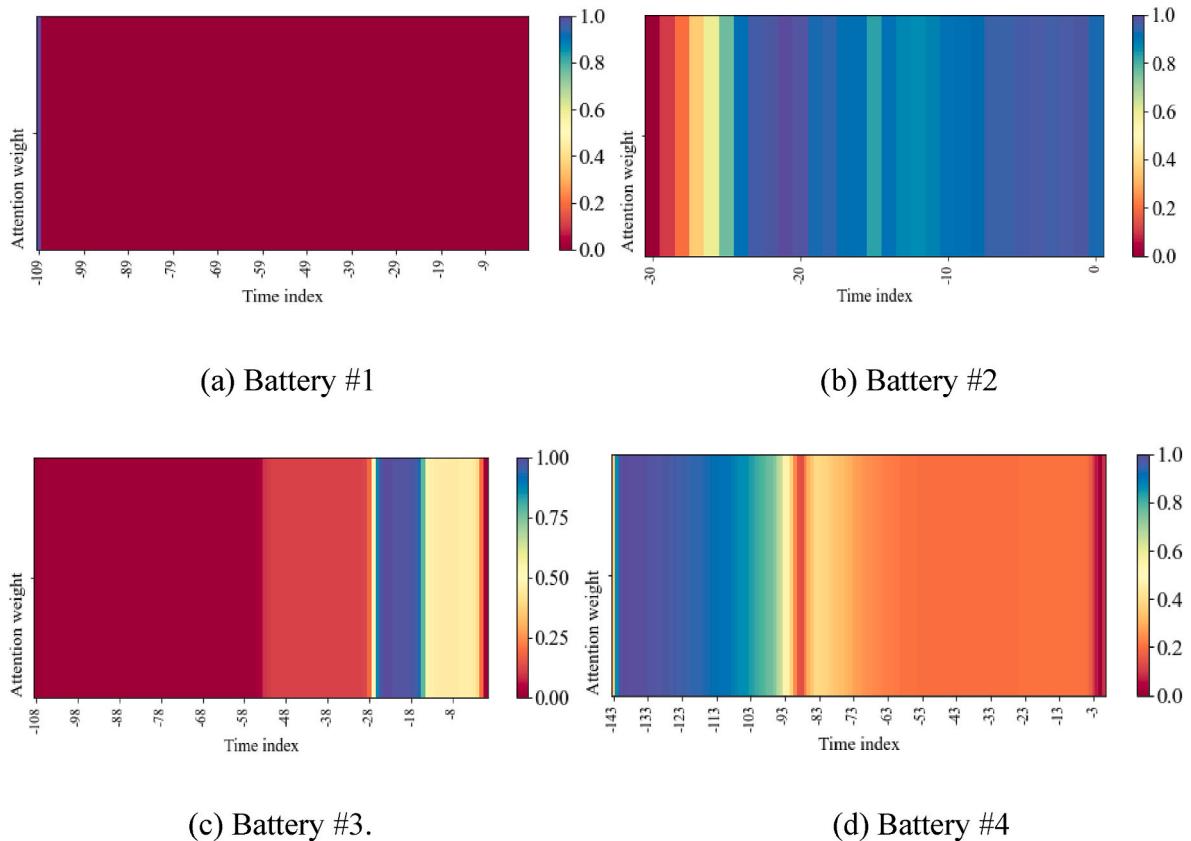


Fig. 15. Average attention weights temporal patterns in the heatmap for four batteries.

prediction results. This signifies a robust capture of local dependencies, while also considering distant values when required. The "Time index (Cycle)" stands out for its notable influence on temporal features, as evidenced in [Table 9](#) under past features.

5. Conclusion

Our study validates our model's effectiveness and robustness using nine Toyota Li-ion batteries with different charging policies. We experiment with two BMS functionalities: SOH estimation and RUL prediction. In SOH estimation, our model shows over 25% improvement compared to counterparts, with maximum error below 1%. Additionally, it outperforms some published methods. Online RUL prediction deals with unknown data after the prediction starting point. For RUL prediction, our method provides early warning results, making it suitable for battery predictive maintenance. Many existing deep learning-based studies lack model interpretability. Our proposed model addresses this issue by assigning different contribution weights at different times, enhancing the understanding of capacity degradation trends. By employing a multistep-ahead rolling prediction method, our model improves future online capacity prediction, aiding in battery replacement and maintenance strategies. The upgraded Sequence-to-Sequence Layer significantly enhances the model's learning efficiency for long-term dependency memory. It's important to note that our approach is currently limited to a single battery. To enhance practicality, our future work will involve designing a battery RUL prediction method for different battery types and packs. Additionally, investigating the impact of different starting points and training durations on model performance in online RUL predictions is a valuable avenue for exploration.

Funding

This research was funded by the National Science and Technology

Council, Taiwan, grant number NSTC-112-2221-E011-114-MY3.

CRediT authorship contribution statement

William Gomez: Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Fu-Kwun Wang:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Formal analysis, Conceptualization. **Jia-Hong Chou:** Visualization, Validation, Software, Methodology, Investigation, Formal analysis.

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

The authors do not have permission to share data.

References

- [1] Strange C, Li S, Gilchrist R, dos Reis G. Elbows of internal resistance rise curves in li-ion cells. *Energies* 2021;14(4):1206.
- [2] Semeraro C, Caggiano M, Olabi A-G, Dassisti M. Battery monitoring and prognostics optimization techniques: challenges and opportunities. *Energy* 2022; 255:124538.
- [3] Chen M, Ma G, Liu W, Zeng N, Luo X. An overview of data-driven battery health estimation technology for battery management system. *Neurocomputing* 2023;532: 152–69.
- [4] Yang R, Xie Y, Li K, Tran M-K, Fowler M, Panchal S, et al. Comparative study on the thermal characteristics of solid-state lithium-ion batteries. *IEEE Trans Transp Electrific* 2024;10(1):1541–57. <https://doi.org/10.1109/TTE.2023.3289997>.

- [5] Hasib SA, Islam S, Chakraborty RK, Ryan MJ, Saha DK, Ahamed MH, et al. A comprehensive review of available battery datasets, rul prediction approaches, and advanced battery management. *IEEE Access* 2021;9:86166–93.
- [6] Gou B, Xu Y, Feng X. An ensemble learning-based data-driven method for online state-of-health estimation of lithium-ion batteries. *IEEE Trans Transp Electrif* 2021;7(2):422–36.
- [7] Ali Zafar, Nengroo Hussain, Park Kim. Online remaining useful life prediction for lithium-ion batteries using partial discharge data features. *Energies* 2019;12(22):4366.
- [8] Attia PM, Bills A, Brosa Planella F, Dechent P, dos Reis G, Dubarry M, et al. Review—“knees” in lithium-ion battery aging trajectories. *J Electrochim Soc* 2022;169:060517.
- [9] Fermín-Cuetos P, McTurk E, Allerhand M, Medina-López E, Anjos MF, Sylvester J, et al. Identification and machine learning prediction of knee-point and knee-onset in capacity degradation curves of lithium-ion cells. *Energy AI* 2020;1:100006.
- [10] Diao W, Saxena S, Han B, Pecht M. Algorithm to determine the knee point on capacity fade curves of lithium-ion cells. *Energies* 2019;12(15):2910.
- [11] Amogne ZE, Wang F-K, Chou J-H. Transfer learning based on transferability measures for state of health prediction of lithium-ion batteries. *Batteries* 2023;9(5):280.
- [12] Jiang H, Wang H, Su Y, Kang Q, Meng X, Yan L, et al. Multiple health indicators assisting data-driven prediction of the later service life for lithium-ion batteries. *J Power Sources* 2022;542:231818.
- [13] Yang K, Tang Y, Zhang S, Zhang Z. A deep learning approach to state of charge estimation of lithium-ion batteries based on dual-stage attention mechanism. *Energy* 2022;244:123233.
- [14] Xue Z, Zhang Y, Cheng C, Ma G. Remaining useful life prediction of lithium-ion batteries with adaptive unscented kalman filter and optimized support vector regression. *Neurocomputing* 2020;376:95–102.
- [15] Wu M, Qin L, Wu G, Huang Y, Shi C. State of charge estimation of power lithium-ion battery based on a variable forgetting factor adaptive kalman filter. *J Energy Storage* 2021;41:102841.
- [16] Zhang J, Huang C, Chow M-Y, Li X, Tian J, Luo H, et al. A data-model interactive remaining useful life prediction approach of lithium-ion batteries based on PF-BiGRU-TSAM. *IEEE Trans Ind Inf* 2024;20(2):1144–54.
- [17] Hasan MM, Ali Pourmousavi S, Jahanbani Ardakani A, Saha TK. A data-driven approach to estimate battery cell temperature using a nonlinear autoregressive exogenous neural network model. *J Energy Storage* 2020;32:101879.
- [18] Li Q, Li D, Zhao K, Wang L, Wang K. State of health estimation of lithium-ion battery based on improved ant lion optimization and support vector regression. *J Energy Storage* 2022;50:104215.
- [19] Sharma P, Bora BJ. A review of modern machine learning techniques in the prediction of remaining useful life of lithium-ion batteries. *Batteries* 2023;9(1):13.
- [20] Cao R, Zhang X, Yang H. Prediction of the heat generation rate of lithium-ion batteries based on three machine learning algorithms. *Batteries* 2023;9(3):165.
- [21] Ren L, Zhao L, Hong S, Zhao S, Wang H, Zhang L. Remaining useful life prediction for lithium-ion battery: a deep learning approach. *IEEE Access* 2018;6:50587–98. <https://doi.org/10.1109/ACCESS.2018.2858856>.
- [22] Zhang Z, Zhang W, Yang K, Zhang S. Remaining useful life prediction of lithium-ion batteries based on attention mechanism and bidirectional long short-term memory network. *Measurement* 2022;204:112093.
- [23] Wang Z, Ma Q, Guo Y. Remaining useful life prediction of lithium-ion batteries based on deep learning and soft sensing. *Actuators* 2021;10(9):234.
- [24] Wang W, Zhang L, Yu H, Yang X, Zhang T, Chen S, et al. Early prediction of the health conditions for battery cathodes assisted by the fusion of feature signal analysis and deep-learning techniques. *Batteries* 2022;8(10):151.
- [25] Li H, Fu L, Zhang Y. A novel hybrid data-driven method based on uncertainty quantification to predict the remaining useful life of lithium battery. *J Energy Storage* 2022;52:104984.
- [26] Sun C, Qu A, Zhang J, Shi Q, Jia Z. Remaining useful life prediction for lithium-ion batteries based on improved variational mode decomposition and machine learning algorithm. *Energies* 2022;16(1):313.
- [27] Qu J, Liu F, Ma Y, Fan J. A neural-network-based method for rul prediction and soh monitoring of lithium-ion battery. *IEEE Access* 2019;7:87178–91.
- [28] Tang Y, Yang K, Zheng H, Zhang S, Zhang Z. Early prediction of lithium-ion battery lifetime via a hybrid deep learning model. *Measurement* 2022;199:111530.
- [29] Chen J, Manivannan M, Duque J, Kollmeyer P, Panchal S, Gross O, et al. A convolutional neural network for estimation of lithium-ion battery state-of-health during constant current operation, vol. 2023. Detroit, MI, USA: *IEEE Trans. Transp. Electrif. Conf. Expo*; 2023. p. 1–6.
- [30] Chen L, Xie S, Lopes AM, Li H, Bao X, Zhang C, et al. A new SOH estimation method for Lithium-ion batteries based on model-data-fusion. *Energy* 2024;286:129597.
- [31] Hsu C-Y, Lu Y-W, Yan J-H. Temporal convolution-based long-short term memory network with attention mechanism for remaining useful life prediction. *IEEE Trans Semicond Manuf* 2022;35(2):220–8.
- [32] Yang P, Yang HD, Meng XB, Song CR, He TL, Cai JY, et al. Joint evaluation and prediction of SOH and RUL for lithium batteries based on a GBLS booster multi-task model. *J Energy Storage* 2024;75:109741.
- [33] Lim B, Arik SÖ, Loeff N, Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int J Forecast* 2021;37(4):1748–64.
- [34] Bacon DW, Watts DG. Estimating the transition between two intersecting straight lines. *Biometrika* 1971;58(3):525–34.
- [35] Severson KA, Attia PM, Jin N, Perkins N, Jiang B, Yang Z, et al. Data-driven prediction of battery cycle life before capacity degradation. *Nat Energy* 2019;4:383–91.
- [36] Chou J-H, Wang F-K, Lo S-C. Predicting future capacity of lithium-ion batteries using transfer learning method. *J Energy Storage* 2023;71:108120.
- [37] Li D, Tan Y, Zhang Y, Miao S, He S. Probabilistic forecasting method for mid-term hourly load time series based on an improved temporal fusion transformer model. *Int J Electr Power Energy Syst* 2023;146:108743.
- [38] Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Network* 2005;18(5–6):602–10.
- [39] Ba JL, Kiros JR, Hinton GE. Layer normalization. 2016, 06450. arXiv:1607.
- [40] Clevert D, Unterthiner T, Hochreiter S. Fast and accurate deepnetwork learning by exponential linear units (ELUs). *ICLR (Poster)*; 2016.
- [41] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. *Proc. Adv. Neural Inf. Process. Syst.* 2017;30:6000–10.
- [42] Bergstra J, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. *Proc. Adv. Neural Inf. Process. Syst.* 2011;24:2546–54.
- [43] Paszke A, Gross S, Massa F, et al. PyTorch: an imperative style, high-performance deep learning library. *Proc. Adv. Neural Inf. Process. Syst.* 2019;32:8024–35.
- [44] Wang F, Zhao Z, Zhai Z, Shang Z, Yan R, Chen X. Explainability-driven model improvement for SOH estimation of lithium-ion battery. *Reliab Eng Syst Saf* 2023;232:109046.
- [45] Lin M, Wu J, Meng J, Wang W, Wu J. State of health estimation with attentional long short-term memory network for lithium-ion batteries. *Energy* 2023;268:126706.
- [46] Fan G-F, Zhang L-Z, Yu M, Hong W-C, Dong S-Q. Applications of random forest in multivariable response surface for short-term load forecasting. *Int J Electr Power Energy Syst* 2022;139:10807.
- [47] Niu Y, Su Y, Li S, Wan S, Cao X. Deep adversarial autoencoder recommendation algorithm based on group influence. *Inf Fusion* 2023;100:101903.