Domain generalization prognosis method for Lithium-ion battery state of health with Transformer and Multi-kernel MMD

Yafei Zhu[1], Tianyi Guo[2], Xiang Li[3](✉), Yewei Zhang[2], Wei Zhang[2]

1 International Engineering College, Shenyang Aerospace University, Shenyang 110136, China.

2 School of Aerospace Engineering, Shenyang Aerospace University, Shenyang 110136, China.

3 Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, Xi'an 710049, China

*Corresponding author: lixiang@xjtu.edu.cn

Received Month X, XXXX | Accepted Month X, XXXX | Posted Online Month X, XXXX

In recent years, a number of intelligent algorithm have been proposed for forecasting the lithium-ion battery state of health (SOH). Due to the varying specifications and operating conditions of batteries, it is difficult to anticipate the health condition of lithium battery as it begins to deteriorate. There are still few studies on health state prediction models for different types of batteries. In this paper, 40 battery data from 5 public datasets are selected to carry out research, and a model architecture consisting of Denoising Autoencoder and Transformer is designed. One or two types of battery packs are identified as the source domain, and multiple types of battery packs are identified as the target domain. By employing Maximum Mean Discrepancy (MMD) on the Transformer architecture, the source and target domains were evaluated and found to converge as training continued. Finally, 29 transfer learning combination tasks were constructed. Results show that the model built with two kinds of batteries as the target domain has the best prediction accuracy and excels in prediction and is versatile in its application. The experimental results also reveal that this study provides a promising tool for predicting Lithium-ion batteries' SOH and strives to build a generalized model of the Lithium-ion batteries' SOH indicators.

Keywords: Battery health management, State of health prognosis, Transformer, Maximum mean discrepancy, Domain generation.

1 Introduction

In recent years, with the increasing demand for energy supply and the guidance of government policies, the rapid development of modern society calls for advanced energy storage technology(Bai et al., 2022; H. Chen et al., 2022). The popularity of lithium-ion batteries (LIBs) stems from their capacity to store a significant amount of energy, nearly zero memory effect, high open circuit voltage and long-lasting performance (Hou et al., 2023). Lithium-ion batteries have proven to be highly effective in electric vehicles (Kennedy & Philbin, 2019), medical electronics (Hu, Ye, Jain, & Schmidt, 2018), mobile devices (Mennik, Dinç, & Burat, 2023), etc. With the use of batteries, the aging problem of lithium batteries limits its energy storage and power output capacity. Additionally, it impacts how long the battery lasts, which is a crucial scientific issue in lithium-ion battery research. Therefore, analyzing the aging mechanism of LIBs and monitoring the change of performance parameters in the aging process, and accurately predicting the remaining useful life (RUL) have guiding significance for the design and production of LIBs as an important part of the health management of LIBs.

In the process of the life cycle decline of LIBs, the voltage, current, impedance and other parameters will change with the degradation process of the LIBs (H. Chen et al., 2022). The mechanism of battery health decline can be reflected through monitoring and analyzing these parameters. By integrating these parameters collected during charge and discharge, the state of charge (SOC) and state of health (SOH) indexes for each battery can be obtained. The SOC of batteries represents the proportion of the remaining charge of a lithium-ion battery to its full charge (Zheng Liu, Zhao, Qiu, Jing, & Yang, 2023). An accurate estimate of how much energy is left in the battery can help us know how long it will keep working. This information is important for using and taking care of the device. The SOH of batteries represents the current battery's capacity to store electrical energy relative to a new battery, representing the state of the battery from the beginning to the end of its life in percentage terms. As an indicator of the long-term utilize of the battery, the SOH is used to quantitatively depict the current performance status of the battery (Zhou, Zhang, Liu, & Yang, 2022) In light of previous work, This study will focus on battery SOH index research, establishing the SOH prognosis model, and providing a basis for the life cycle analysis of LIBs.

At present, there exist two different approaches for forecasting the state of health. One is a model-based method. Zhang et al. proposed a novel method combining voltage capacity model based incremental capacity analysis with support vector regression for battery SOH estimation (Y. Zhang, Liu, Wang, & Zhang, 2022). Mohamed Ahwiadi and Wilson Wang proposed an enhanced PF technology, which adopted an evolving fuzzy predictor technique fused into enhanced PF structure, and by using simulation experiments validate the effectiveness of the model (Ahwiadi & Wang, 2022). Ouyang et al. proposed a health state estimation method that combines the improved firefly algorithm to solve the particle dilution problem with the particle filter, and successfully achieved high accuracy in the battery SOH prediction problem (Ouyang, Ye, Xu, Wang, & Xu, 2023). Wen et.al proposed a battery SOH prediction model based on incremental capacity analysis and BP neural networks. This method through the analysis of the correlation of IC curve characteristics and SOH, using the least squares method to establish the relationship between the

temperature and the feature of IC curve mapping, SOH is obtained under different temperature prediction model (Wen, Chen, Li, & Li, 2022).

In spite of the fact that the model-based approach has the advantage of being able to penetrate into the nature of the object system and accomplish battery SOH prediction. However, it is frequently troublesome to set up exact mathematical models for complex dynamic systems in practical applications. Data-driven methods have been applied in numerous areas. For example, fault diagnosis (H. Wang, Li, & Ren, 2023), energy management (Shi, Liu, & Xiao, 2022) and remaining useful life prediction. Li et al. proposed an end-to-end hybrid neural network (NN), a connection of one-dimensional convolutional neural networks and active state-tracking long short-term memory neural networks. The model is utilized to capture various leveled features between several variables that influence the battery degradation process, as well as the time dependencies embedded in these features (P. Li et al., 2022). Chemali et al. propose a convolutional neural network-based (CNN) framework to directly estimate SOH using voltage, current, and temperature measured from the battery while charging, while also enhancing the training data with noise and error to improve accuracy (Chemali, Kollmeyer, Preindl, Fahmy, & Emadi, 2022). By including attention mechanism to the bidirectional long short term memory network, this strategy can precisely judge the multi-feature attention distribution, extract valuable information, and decrease the uncertainty of multi-step prediction (Z. Zhang, Zhang, Yang, & Zhang, 2022). Data-driven methods don't need to know about the object system beforehand (S. Liu, Duan, Wang, Zhang, & Hao, 2023). Based on the collected data, this method mines the covered up data through various information examination and handling strategies for prediction operation, in this way avoiding the shortcomings of model-based and knowledge-based prediction technology, and becoming a practical LIBs prediction strategy.

In recent years, transfer learning has been widely concerned by the academic community (Ma et al., 2022; Pinto, Wang, Roy, Hong, & Capozzoli, 2022; S. Yao, Kang, Zhou, Rawa, & Abusorrah, 2023). Its connotation is to transfer labeled data or knowledge structure from related fields, complete or improve the learning effect of target fields or tasks. Due to the many differences in the model and working condition of lithium-ion batteries, different lithium-ion batteries also have different changes in the health state during the decline process. Therefore, a model trained on one type of battery does not necessarily show high predictive accuracy on other battery datasets. Therefore, transfer learning method is proposed to solve the problem of lithium-ion battery health state prediction. Ma et al. proposed a convolutional neural network combined with improved domain adaptive method is proposed to construct SOH estimation model. The method uses CNN to automatically extract features from the original charging voltage trajectory, and uses the maximum mean discrepancy domain adaptive method to reduce the distribution difference between the training and test battery data (Ma et al., 2022). Yao et al. proposed a deep transfer learning method, which utilizes the capacity increment characteristics of partial charge and discharge segments to estimate battery capacity by training the deep transfer convolutional neural network (DTCNN) of source and target data simultaneously (J. Yao & Han, 2023). In summary, these transfer learning methods boil down to the study of domain adaptation. At its core is the ability of models learned from one type of battery data (the source domain) to

generalize to a different but related type of battery data (the target domain), aiming to solve the problem of distribution discrepancy between the source domain and the target domain.

Domain generalization is a further generalization of domain adaptive. The difference between domain adaptation and domain generalization is whether the target domain data can be obtained during training. The goal is to learn more robust models, and in the SOH prediction problem of lithium-ion batteries, the expectation of building a prediction model that fits many different types of batteries at the same time becomes a key problem. Based on this objective, this paper will study 40 battery data from 5 publicly available datasets, and strive to build a generalization model that conforms to 5 datasets concurrently.

The contributions of this paper are listed as follows:

• A domain generalization method for SOH prediction of multi-type battery datasets is proposed, which improves the prediction accuracy of multiple different types of lithium-ion battery datasets by adjusting the weight parameters in the model.

• The denoising autoencoder mechanism is adopted to reconstruct the inputs, which makes the model robust to the noise in the input.

• The multi-head self-attention mechanism prediction method can capture the dependencies between different positions in the sequence, and the residual neural network mechanism retains the input information of each sub-layer in the model, and effectively improves the long-term memory ability of the model (Zhu, Li, & Zhang).

The remainder of this paper starts with the theoretical background in the literature in Section 2. The proposed methods are presented in Section 3, and the research objects involved are presented in Section 4. The experimental process and results are presented in Section 5. We finish this paper with conclusions in Section 6.

2 Theoretical Background

2.1 Denoising Autoencoder

Typically, the sensor collects data along with some unavoidable noise. By training the network with noisy training data, the neural network can learn the input features without noise and the main features of noise. This can enhance the network's performance on test data. DAE employs an automatic encoder to add noise into the training data in training the entire network.

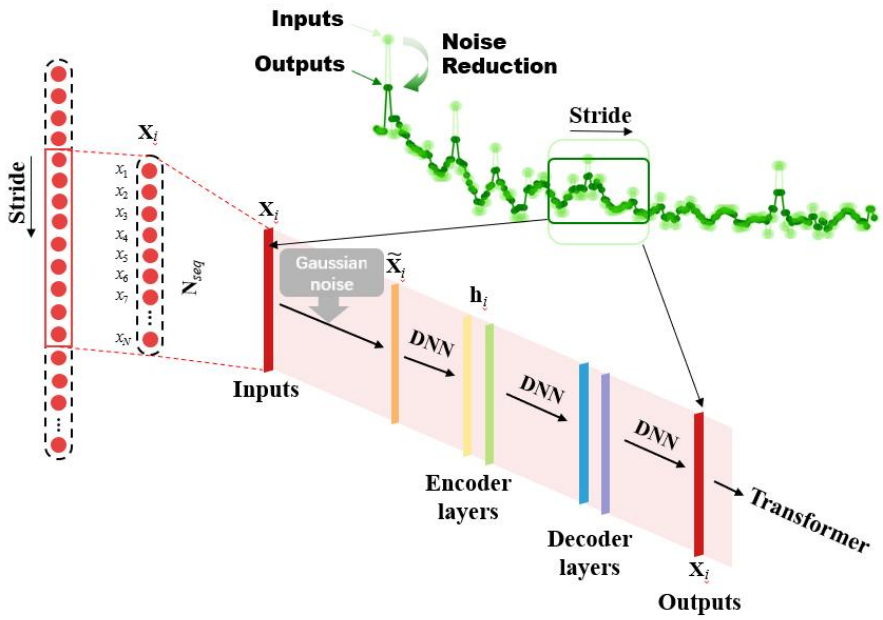

Fig. 1. The process of Denoising autoencoder

Firstly, the time series containing the health status of the lithium-ion battery will be segmented into the network architecture and represented as $x_i = [x_1, x_2, ..., x_N]$, where N represents the length of the cut segment. Subsequently, Gaussian noise of covariance $\sigma^2 I$ is added to vector x_i to obtain a time series sample, which expressed as $\tilde{x}_i = x_i + \varepsilon, \varepsilon \sim N(0, \sigma^2 I)$. This corresponds to conditional density $q_\sigma(\tilde{x}_i \mid x_i) = \frac{1}{(2\pi)^{d/2}\sigma^d}\exp(-\frac{1}{2\sigma^2}\| \tilde{x}_i - x_i \|^2)$

Subsequently, the time series with noise $\tilde{x}_i$ is transmitted into the neural network to obtain a hidden representation h_i, which expressed as,

$$h_i = \text{encode}(\tilde{x}_i) = \varphi(w^T \tilde{x}_i + b) \quad (1)$$

where w represents the weight term, $*^T$ represents transpose the vector *. b represents the bias term, and φ represents the nonlinear activation function.

Then the hidden representation h_i is decoded into reconstruction x_i', which expressed as,

$$x_i' = \text{decode}(h_i) = \varphi(w'^T h_i + b') \quad (2)$$

where w' represents the weight term, b' represents the bias term.

The LeakyReLU function, as the activation function φ in the training process, is defined as:

$$\text{LeakyReLU}(x) = \begin{cases} x & , x > 0 \\ ax, & x < 0 \end{cases} \quad (3)$$

The LeakyReLU function changes the negative zero gradient by setting the negative input to the least linear component, thus avoiding dead neurons. And the weight parameter α in Eq.(3) is set to 0.01 in this study.

Finally, optimizing the network weight parameters in the training process minimizes the expected square reconstruction error between $\tilde{x}_i$ and x_i. Therefore, the objective function of the DAE network can be expressed as,

2.2 Transformer

Originally proposed for natural language processing, Transformer has become the preferred model for solving NLP problems and replacing the traditional RNN model (Vaswani et al., 2017). In recent years, the model has excelled in resolving research issues across a wide range of areas, such as computer vision (Ze Liu et al.), speech signal processing (Dong, Xu, & Xu), etc. By incorporating the Attention mechanism, this model enhances communication between positions in a sequence. Thus this research involved using the Transformer model to estimate the SOH of lithium-ion batteries.

2.2.1 Positional Encoding

When preparing time series, conventional neural network models will handle each value within the time series one by one in a specific arrange, such as the sliding procedure of the convolutional kernel in the convolutional neural network. This permits the sequential information to be naturally preserved during processing and does not require extra processing.

However, Transformer needs to process the time series of inputs at the same time, which makes the sequential information lost when entering the network. The Transformer use Positional Encoding to indicate the position of each value in the sequence. This method obtains the position and order of each value in the sequence by adding positional encoding to the input (P.-C. Chen et al., 2021).

A simple encoding method that uses sine and cosine functions to encode input vectors, expressed as,

$$PE(t,i)=\begin{cases}\sin(t/10000^{2k/d}), & if\ \ i=2k\\ \cos(t/10000^{2k/d}), & if\ \ i=2k+1\end{cases} \tag{4}$$

where t denotes the time step of the sequence, and d represents the encoding dimension (J. Zhang, Li, Tian, Luo, & Yin, 2023).

$$L=\frac{1}{n}\sum_{i=1}^{n}\|\tilde{x}_i-x_i\|^2 \tag{5}$$

2.2.2 Multi-head Self-Attention

The Self Attention mechanism can directly associate any two values in the input vector, so the distance between the distance-dependent features is greatly shortened, which is conducive to effectively extracting distant features. Given the representation of the (l-1)-th layer, the i-th (i ∈ [0, h]) attention is defined as,

$$head_i = Attention(T^{l-1}W^{Q}_{i}, T^{l-1}W^{K}_{i}, T^{l-1}W^{V}_{i}) \tag{6}$$

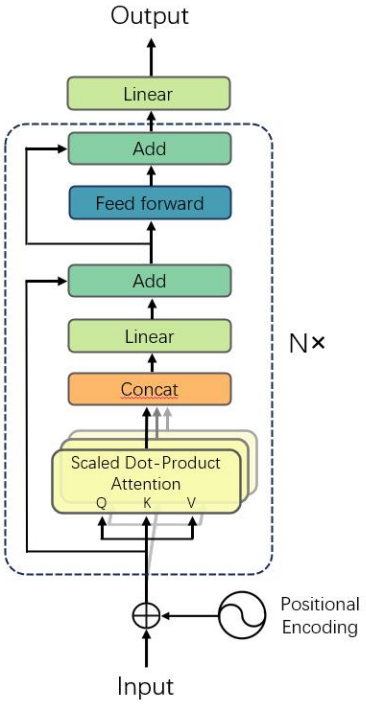

Fig. 2. The architecture of Transformer.

where W^{Q}_{i}, W^{K}_{i}, and W^{V}_{i} are weight matrix, and Q, K, V are defined as query, key and value. T^{l-1} and h parallel attention functions. The calculation process of the self-attention mechanism is defined as,

$$Attention(Q,\ K,\ V)=softmax(\frac{QK^T}{\sqrt{d_h}})V \tag{7}$$

where $d_h = d/h$. Scaling by $\sqrt{d_h}$ alleviates the situation where the gradient becomes small

after passing softmax due to the large values after query and key dot multiplication (Su, Xiong, & Yang, 2023).

The multi-head attention mechanism further details the attention layer, multiplies it with multiple groups of W^Q, W^K and W^V, generates different attention subspaces by using different projection matrices of query, key and value, and finally splices the output of all subspaces together to form the final output, which is expressed as,

$$MultiHead(T^{l-1}) = Concat(head_1, head_2, ..., head_h)W^O \quad (8)$$

where W^O is a weight matrix (D. Chen, Hong, & Zhou, 2022).

2.2.3 Feed-Forward

After the Attention operation, the fully-connected forward network will continue the computation on top of the encoder and decoder, thus performing the same operation on the vector at each position, which is expressed as,

$$FFN(x) = Leaky\text{ReLU}(xW_1 + b_1)W_2 + b_2 \quad (9)$$

and T^l could be obtained from the previous MultiHead (T^{l-1}), which expressed as,

$$T^l = FFN(Multihead(T^{l-1})) \quad (10)$$

Finally, in order to predict future battery SOH sequence, i.e., $\hat{y}_i = (x_{N+1}, x_{N+2}, ...)$, a fully connected layer is constructed to map the representation learned by the last transformer cell to get the output result. And The fully connected layer chose LeakyReLU as the activation function, which avoids the problem of neuron death by introducing a small slope so that there is also a non-zero gradient when the input is less than 0, which expressed as,

$$\hat{y}_i = Leaky\text{ReLU}(W_p T^l + b_p) \quad (11)$$

where W_p and b_p denotes weight and bias matrix.

2.3 Transfer learning

In general, it is difficult to obtain training samples due to the long experimental period required for the lifetime charge-discharge cycle experiment of LIBs. In recent years, the academic community is seeking to leverage the existing knowledge to accelerate the acquisition of new knowledge. Numerous studies have been conducted. (Kim, Choi, Kim, & Choi, 2021; Yihuan Li, Li, Liu, Wang, & Zhang, 2021; Yang Li & Tao, 2020).

Reviewing the traditional deep learning prediction model, its strategy is to select a certain number of samples for training, and verify the remaining samples after the training is completed, to establish a prediction model that meets the experimental conditions of the sample. However, the training and verification samples may differ considerably due to factors such as the testing process, sample types, and operating environment.

Previous work has also successfully demonstrated the feasibility of two different types of batteries for transfer learning (Zhu, Li, Zhang, & Zhang, 2023). However, training batteries to fit more unknown labels at the same time as the target domain remains a challenge. The key lies in how to dig out the common characteristics of multiple types of battery health state decline, and use the characteristics as the basis for prediction models. Therefore, this study will select one type of battery as the source domain sample (with labels), two or more types of batteries as the target domain sample (no labels), and the remaining types of batteries as validation samples. The feasibility of extracting the common characteristics of various battery SOH degradation types from the model was studied.

2.4 Maximum Mean Discrepancy

This study will measure the differences that exist between multiple battery datasets based on the Maximum Mean Discrepancy (MMD).

It is defined as the square distance between the kernel embeddings of marginal distributions in the reproducing kernel Hilbert space (RKHS), which can be expressed as,

$$MMD_k(P, Q) = \| E_p[\phi(x^s)] - E_Q[\phi(x^t)] \|^2_{H_k} \quad (12)$$

where H_k denotes the RKHS endowed with a characteristic kernel k, moreover, it conforms to the property of $MMD_k(P, Q) = 0$ if and only if $P = Q$ (Gretton, Borgwardt, Rasch, Schölkopf, & Smola, 2012).

At the same time, Different methods of computing MMD loss will also produce varying effects. That is because different kernels can embed probability distributions in different RKHS, in which different RKHS can emphasize different orders of sufficient statistics. However, the MMD kernel function cannot be changed, so it is difficult to pick the best one. Due to this, the study decided to adopt multi-kernel MMD loss to evaluate the discrepancy between different domains. Therefore, a mixture of N_k RBF kernels could be expressed as,

$$k(x^s, x^t) = \sum_{i=1}^{N_k} k_{\sigma_i}(x^s, x^t) \quad (13)$$

where k_{σ_i} represents a Gaussian kernel with bandwidth parameter σ_i.

3 Proposed Method

3.1 Prognosis Index

The decline in LIBs is occurring gradually and is a complex procedure. Usually, when a battery is recharged and subsequently depleted, it is referred to as a cycle. The battery deteriorates with increased usage over time. A measure of battery health is introduced as a predictive indicator, namely SOH, which is defined as,

$$S_t = \frac{C_t}{C_0} \times 100\% \quad (14)$$

where C_t represents the capacity at a specific charge cycle stept, and C_0 represents the battery's rated capacity (Fan, Xiao, Li, Yang, & Tang, 2020). Additionally, comparing the current charge capacity of the battery with its initial capacity enables us to grasp the changes in the battery's capacity following each charge and discharge cycle.

For the capacity C_t of a battery during a complete charge/discharge cycle, the current during the discharge phase of the battery can be captured and integrated over the time t, which difined as,

$$C_t = \int_{t_a}^{t_b} i(t)dt \quad (15)$$

where t_a and t_b denote the initial time and end time of the battery discharge process, respectively. Therefore, A data acquisition system is designed using the above formula to track the charging and discharging of lithium-ion batteries. The current signal of each cycle process is extracted successively, and then the changing trend of SOH in the life cycle of the lithium-ion battery is obtained.

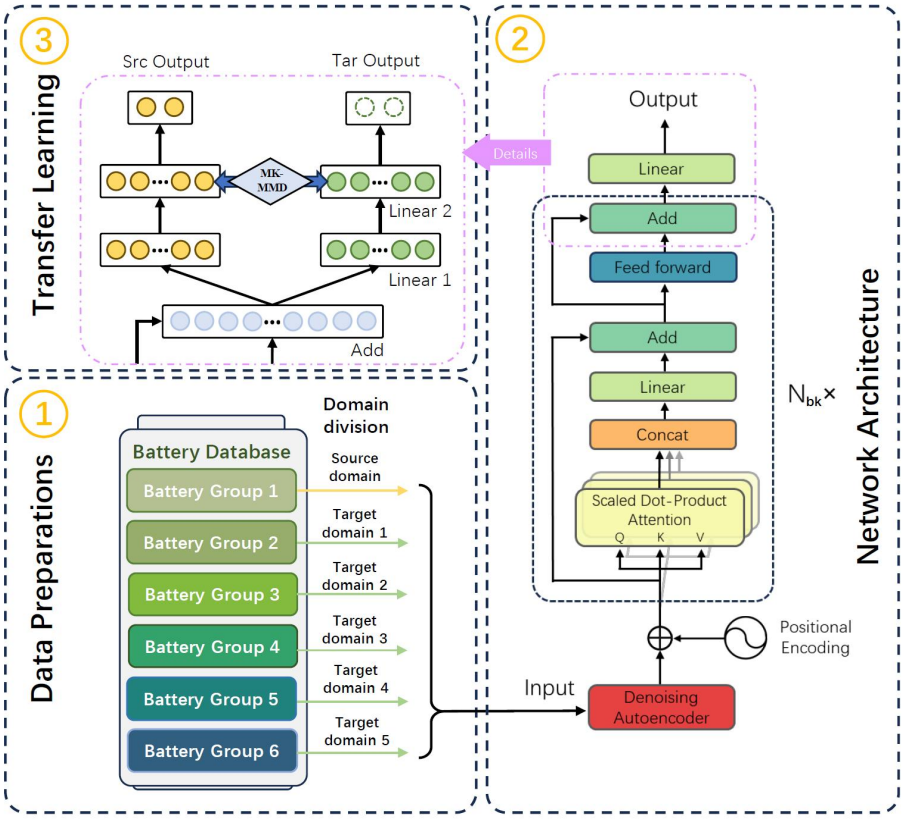

Fig. 3. The architecture of the proposed method

3.2 Model Architecture

This section will explain the proposed model architecture. The whole model takes the prediction model composed of Autoencoder and Transformer as the main architecture, multi-kernel MMD as the index to measure the differences between different domains, and Figure 3 shows the architecture of the proposed method.

First, the individual battery datasets, which follow different distributions, determine the source and target domains through a domain partitioning process. Select a type of battery as the source domain that contains labels throughout the training process. The rest of the battery data is used as the target domain, which does not contain labels during the entire training process. Each type of battery dataset contains experimental data of multiple battery charge and discharge cycle experiments during the whole life cycle, from which time series of SOH changes used across the entire experiment process are extracted and divided into vectors with sequence length N_{seq} of 16, which are used as input values of the neural network. Then the vector from each dataset is fed into the same network to begin the forward transmission process.

The main architecture of the network is composed of two modules: noise reduction self-coding and Transformer. First, the input time series is transmitted to the Denoising Autoencoder module to reduce potential problems caused by data noise. The noised data is encoded into a hidden layer by adding Gaussian noise to the sequence and transmitting it into an encoder constructed by a two-layer DNN. The decoder constructed by the same two-layer DNN decodes the hidden

layer representation into the original representation, thus achieving the noise reduction process. The dimensions of each hidden layer remain the same, with a length of 8.

After that, the reconstructed sequence is encoded by position to give the global information of the sequence to save the position of the values of each position in the sequence. Then, the sequences enter the Multi-Head Attention layer to achieve information fusion. An Add layer is also included above the Multi-Head Attention to indicate residual connections to prevent network degradation. The Feed-Forward layer then maps the data to a high-dimensional space and then to a low-dimensional space through a linear transformation, which is used to extract deeper features.

Finally, the end of the network consisting of two linear layers is used for transfer learning training. Where, the MMD value of the source domain and target domain data will be calculated in the last linear layer, representing the distance between the two distributions during the training process, and the distance will be set in the objective function during the training process. Since the number of target domains is more than one, it is necessary to calculate the corresponding MMD values of source domains and multiple target domains respectively, and add the MMD values to the objective function for optimization, so as to narrow the distribution discrepancy between source domains and target domains after neural network training.

The output value of the neural network represents the SOH value of the battery in a future period of time. This study focuses on the establishment of a generalization model in the field of battery dataset, so the output value is studied in the simplest case, i.e., the SOH of one cycle in the future.

To improve operational effectiveness, Mini-batch strategy is used in the preparing prepare of neural network. This strategy divides the different types of battery data into several batches, and each batch of data is passed into the neural network to upgrade the parameters, and the set of data in a batch jointly decides the direction of the current training gradient. The computational load of network training is essentially decreased, and the issue of conflicting data quantity of diverse datasets cannot be presented into the neural network at the same time is solved.

3.3 Optimization Objective

Traditional training aims to minimize the difference between the predicted value and the actual value. Therefore, when constructing the objective function, it is necessary to compare the output value of the neural network, i.e., the predicted value with the real value. During training, the network weight parameters are optimized and predicted value and the real value is gradually reduced. Root Mean Square Error (RMSE) is a commonly uused way to measure how well a forecast works, which is defined as,

$$L_p = RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{16}$$

where N denotes the number of samples involved in RMSE loss, and each epoch of RMSE loss during training is expressed as L_p (W. Wang & Lu).

In addition, the differences between source domains and different domains should be made up during training. As previously stated, the loss of multi-kernel MMD can be used as a measure. In order to normalize the MMD loss value between different domains, the source domain dataset can be expressed with the number "1", and the remaining datasets used as the target domain are represented with the numbers "2, 3, 4...". Thus, the MMD loss between the target domain dataset p and the

source domain can be expressed as $L_{m;1,t}$, which is calculated by the following formula:

$$L_{m;1,t} = \sum_{k \in K} MMD_k(P,Q) \ , t = 2,3,... \quad (17)$$

where K denotes the kernel set. $MMD_k(P, Q)$ denotes the operation of MMD values between hidden layers P and Q using kernel k. In general, the last hidden layer of the network is often utilized as the P and Q for network training.

A complete loss term is obtained by integrating the indexes of multiple optimization objectives. Therefore, the final loss L can be defined as:

$$L = L_P + \sum_{t=1}^{T} l_{1,t} L_{m;1,t} \quad (18)$$

Where $l_{1,t}$ represent the MMD loss weights of the source and target domains, identified as t. T represents the total number of domains divided into target domains. The configuration of different weights will affect the generalization ability of the model. In the experimental setup, the influence of different weight combinations on SOH prediction of battery datasets with multiple target domains will be investigated.

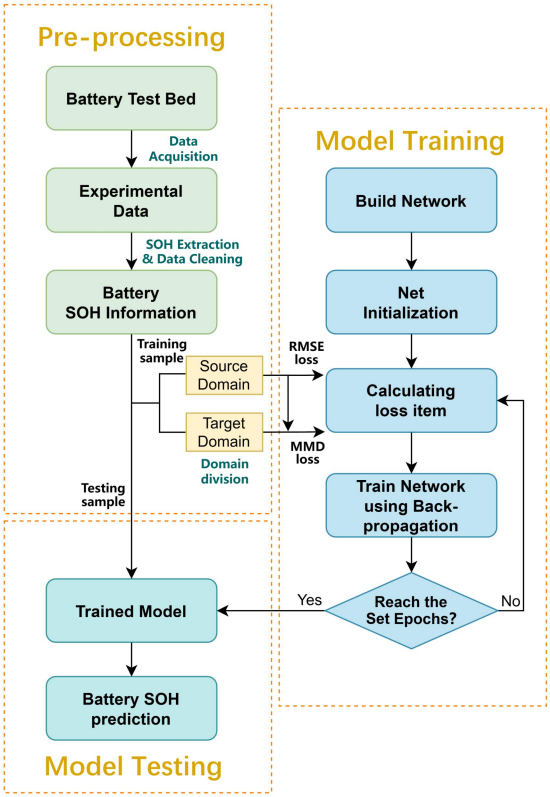

Fig. 4. Flowchart of the proposed prognostic method.

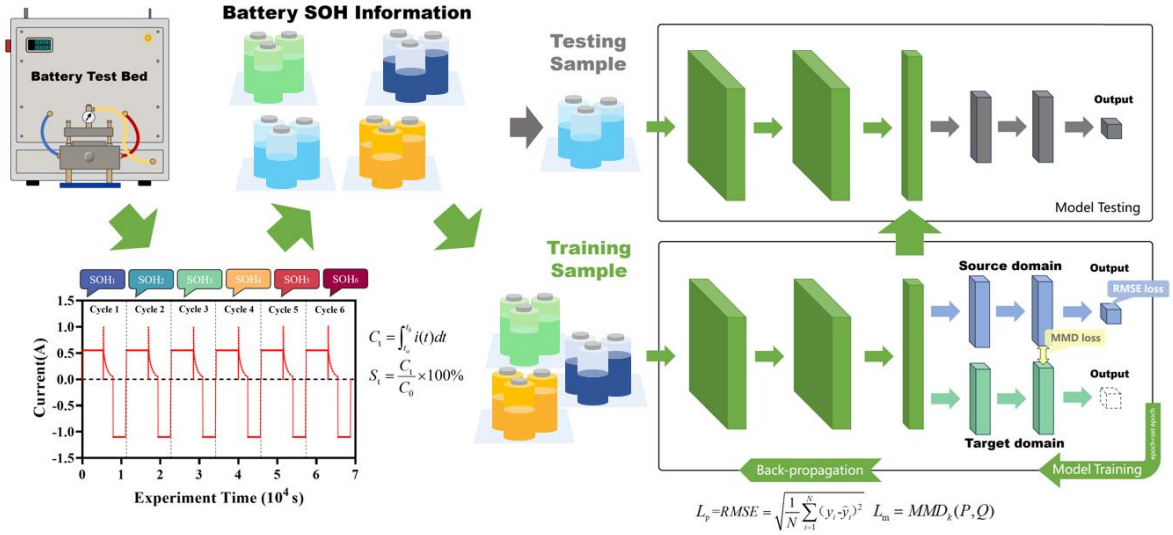

Fig. 5. The diagram of SOH prognosis process for various types of lithium-ion batteries

3.4 Prognosis Procedure

The entire process of prognosis can be clarified in this way. Firstly, the experimental data of various types of batteries are obtained according to the battery test bed. The SOH obtained through the experimental data, and the outliers were initially removed through the data cleaning.

For the training process, the model architecture was built and the network parameters were initialized according to the proposed method. And the battery SOH data from the training sample is input into the network for forward propagation calculation and obtain each loss item value. Then the network weight parameters are updated according to the acquired loss items in the process of network backpropagation. The established network model is obtained when the training is finished for the certain epoch.

Finally, for the testing process, the test sample is passed into the trained model to obtain the predicted value of the battery SOH. The flowchart of the entire training process is shown in Figure 4 and Figure 5.

4 Data Description

4.1 CALCE battery dataset

The experimental subjects in this study were derived from the CALCE dataset at the University of Maryland (Xing, Ma, Tsui, & Pecht, 2013). This dataset provides experimental test data on LIBs, including continuous full cycle and partial cycle, storage, dynamic drive curves, open-circuit voltage measurements, and impedance measurements. In this study, two kinds of columnar batteries were selected as research objects, i.e., CS2 and CX2 batteries. Figure 6 (a) and (b) shows the SOH changes of the two groups of batteries from the beginning of the experiment to the end of life.

4.2 NASA battery dataset

The experimental dataset comes from NASA's Ames Center for Prediction Excellence data repository. The experiment was carried out on the battery prediction test bench. Multiple charge and discharge cycles were carried out at room temperature for a large number of commercially available 18650 LIBs, and accelerated aging tests were carried out. This study selected four kinds of batteries that are widely used at present, namely batteries B005, B006, B007 and B0018, for comparison. Detailed battery health data are shown in Figure 6 (c).(W. Zhang, Li, & Li, 2020)

4.3 UL-PUR battery dataset

The experimental dataset comes from the open dataset of Underwriters Laboratories Inc. - Purdue University (UL-PUR) (Juarez-Robles, Jeevarajan, & Mukherjee, 2020). In this study, six batteries in the dataset were selected as experimental data and represented by UL001, UL002, UL003, UL004, UL005 and UL006 respectively. Their health status curves are shown in Figure 6 (d).

4.4 HNEI battery dataset

The experiment dataset comes from the Hawaii Natural Energy Institute (HNEI), which studied a batch of Model 18650 LIBs manufactured by IG Chemical Co., LTD. (Seoul, South Korea) for use in laptop computers (Devie, Baure, & Dubarry, 2018). Among them, 15 LIBs were used to study the cyclic aging characteristics of batteries. In this study, these 15 batteries were used with HN001, HN002, ..., HN015 indicates. Figure 6 (e) shows the battery SOH change curve of the dataset.

4.5 SNL battery dataset

The dataset from Sandia National Laboratories covers three types of LIBs. One type of battery used in this experiment is the 18650 battery, manufactured by mpanasonic with electrode material NCA, containing 7 battery experimental data. According to the difference of charge and discharge cycle experiments, each battery is marked with "xC", x represents the discharge rate of the battery. For instance, "0.5C_001" indicates that the discharge is carried out at a constant current rate of 0.5C, and the first battery under this experimental condition is labeled "001". Figure 6 (f) shows the battery SOH change curve of the dataset.

Table 1 shows the basic information of all lithium-ion battery datasets. Figure 7 illustrates the differences between the batteries of different battery groups, where two of the batteries of each group were arbitrarily selected for plotting.

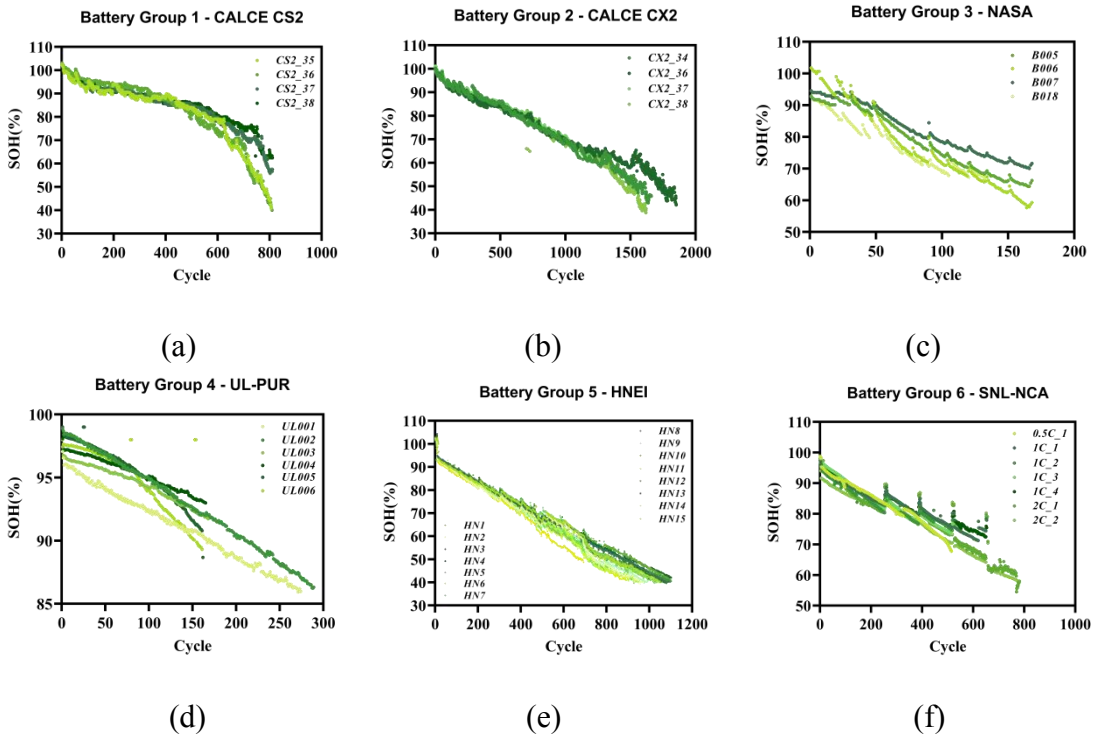

Fig. 6. Plot of SOH trends for the six LIBs datasets in the experiment

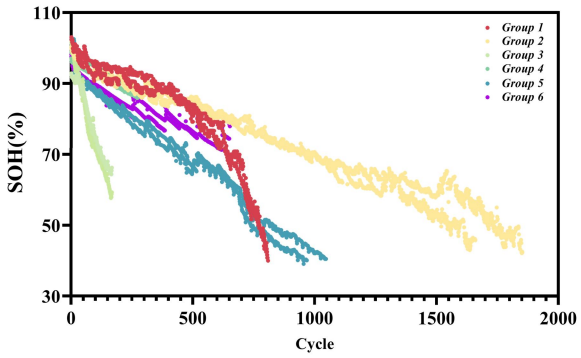

Fig. 7. Comparison of differences in SOH data of 6 battery group

5 Experimental Study

5.1 Details

The objective function of the proposed network has been defined in Eq.18. By setting the weight parameter $l_{1,t}$ item before the $L_{m;1,t}$ of the source domain and each target domain, you can set the specific target domain and the contribution degree of the target domain to transfer learning. The main purpose of this experiment is to find a generalization model that is suitable for multiple domains at the same time. Therefore, it is a key task to make the model obtain relatively small accuracy in various fields by setting different weight parameter combinations.

In this experiment, it is determined that battery group 1 is the source domain data, namely CS2 of CALCE dataset. The remaining battery groups are used to set the target domain. Therefore, the weight parameters $l_{1,2}$, $l_{1,3}$, $l_{1,4}$ and $l_{1,5}$ correspond to the CX2 dataset from CALCE, the NASA dataset, the UL-PUR dataset, the HNEI dataset and the SNL dataset respectively. Table 2 shows these weight combinations chosen for this experiment, covering the types of target domains ranging from 1 to multiple.

According to the data in Table II, the target domain is set for one of the battery datasets from the transfer tasks T_1 to T_4. From the transfer task T_5 to T_{10}, 2 types of battery datasets were set as target domains. Since this study only considered the feasibility of model domain generalization under setting multiple target domains, the weight parameters were set to consistent values. Similarly, the transfer task T_{11} to T_{13} sets 3 kinds of batteries as the target domain, and the transfer task T_{14} is a transfer learning task with all 4 kinds of batteries as the target domain.

At the same time, battery group 1 and battery group 2 were also utilized as the source domain and other battery groups were utilized as the target domain for training in this study. For the RMSE losses of these two types of batteries during training, they are expressed as L_{p1} and L_{p2}, respectively.

Similarly, the MMD loss of the two different types of batteries is expressed in $L_{m;\,a,b}$, where a indicates the battery string number in the source domain, and b indicates the battery string number in the target domain. a indicates the battery string number in the source domain, and b indicates the battery string number in the target domain. Therefore, the loss term can be expressed as,

$$L = L_{p1} + L_{p2} + l_{1,3} \times L_{m;1,3} + l_{1,4} \times L_{m;1,4} + l_{1,5} \times L_{m;1,5} + l_{2,3} \times L_{m;2,3} + l_{2,4} \times L_{m;2,4} + l_{2,5} \times L_{m;2,5} \quad (19)$$

Table 1. The overview of battery dataset.

Type	CALCE - CS2	CALCE - CX2	NASA-18650
Rated capacity:	1100 mAh	1300 mAh	1100 mAh
Train sample:	CS2_35, CS2_36, CS2_37	CX2_34, CX2_36, CX2_37	B005, B006, B007
Test sample(s):	CS2_38	CX2_38	B018
Experimental conditions	Temperature: 25℃ •Charge at a constant current rate of 0.5C until the voltage reaches 4.2V and then 4.2V was sustained until the charging current dropped to below 0.05A. •The discharge cut off voltage was 2.7V. Cycled at constant current of 1C.	Temperature: 25℃ •Charge at a constant current rate of 0.5C until the voltage reaches 4.2V and then 4.2V was sustained until the charging current dropped to below 0.05A. •The discharge cut off voltage was 2.7V. Cycled at constant current of 0.5C.	Temperature: 25℃ •Charge at 1.5A constant current until the voltage reaches 4.2V, then continue charging in constant voltage until the charge current drops to 0.02A. •Discharge at a constant current of 2A until the voltages of batteries 5, 6, 7 and 18 drop to 2.7V, 2.5V, 2.2V and 2.5V respectively.
Type	UL-PUR - 18650	HNEI - 18650	SNL- 18650
Rated capacity:	3400 mAh	2800 mAh	3200 mAh
Train sample:	UL001, UL002, …, UL005	HN001, HN002, …, HN013	0.5C_001, 1C_001, 1C_002, 1C_003, 1C_004, 2C_001
Test sample (s):	UL006	HN014, HN015	2C_002
Experimental conditions	Temperature: 23℃ •Charge at a constant current rate of 0.5C until the voltage reaches 4.2V and then 4.2V was sustained until the charging current dropped to below 0.05A. •The discharge cut off voltage was 2.7V. Cycled at constant current of 0.5C.	Temperature: 25℃ •Charge at 0.5C constant current. •Discharge at 1.5C constant current	Temperature: 25℃ •Charge at 0.5C constant current. • Discharge: Refer to "xC" on the front of the battery label, where x stands for discharge at a constant current rate value.

Table 2. Weight parameter Settings under different transfer tasks.

Transfer tasks	$l_{1,2}$	$l_{1,3}$	$l_{1,4}$	$l_{1,5}$
T_0	0	0	0	0
T_1	**40**	0	0	0
T_2	0	**40**	0	0
T_3	0	0	**40**	0
T_4	0	0	0	**40**
T_5	**20**	**20**	0	0
T_6	**20**	0	**20**	0
T_7	**20**	0	0	**20**
T_8	0	**20**	**20**	0
T_9	0	**20**	0	**20**
T_{10}	0	0	**20**	**20**
T_{11}	**10/3**	**10/3**	**10/3**	0
T_{12}	**10/3**	**10/3**	0	**10/3**
T_{13}	0	**10/3**	**10/3**	**10/3**
T_{14}	**8**	**8**	**8**	**8**

Table 3. Weight parameter Settings under two training source domains.

Transfer tasks	$l_{1,3}$	$l_{1,4}$	$l_{1,5}$	$l_{2,3}$	$l_{2,4}$	$l_{2,5}$
$T_{0'}$	0	0	0	0	0	0
T_1'	**40**	0	0	0	0	0
T_2'	0	**40**	0	0	0	0
T_3'	0	0	**40**	0	0	0
$T_{4'}$	0	0	0	**40**	0	0
T_5'	0	0	0	0	**40**	0
T_6'	0	0	0	0	0	**40**
T_7'	**20**	0	0	**20**	0	0
T_8'	**20**	0	0	0	**20**	0
T_9'	**20**	0	0	0	0	**20**
T_{10}'	0	**20**	0	**20**	0	0
T_{11}'	0	**20**	0	0	**20**	0
T_{12}'	0	**20**	0	0	0	**20**
T_{13}'	0	0	**20**	**20**	0	0
T_{14}'	0	0	**20**	0	**20**	0
T_{15}'	0	0	**20**	0	0	**20**

Table 3 shows the setting of different weight parameters under this loss item. The experiment covers transfer learning under experimental conditions with two batteries as source domain and the other batteries as target domain. And these transfer learning tasks are represented by T_1', T_2' to T_{15}' respectively.

Meanwhile, transfer task T_0 is set to 0 for all the weight parameters, i.e., only the source domain dataset is trained separately. Aimed at assessing the model analysis model of the optimization effect, experimental conditions will be a baseline comparison with other transferring situations.

Table 4. Training process parameter setting table.

Parameter	Value
Training epochs	2000
Learning rate lr	0.005
Mini-batch size N_{batch}	64
Transformer blocks N_{bk}	1
Neurons in Feed-forward layer	64
Heads number in attention mechanisms	8
Gaussian kernels number	8

In the network training process, the backpropagation optimization method is used to update the model coefficients, and the Adam optimizer is used to update the parameters in the network (Kingma & Ba, 2014). Each training sample was randomly divided into multiple small batches, containing 64 samples. Each small batch sample is optimized by using the total loss term defined in Eq.18 as the optimization target. The learning rate value is set to 0.005 and the number of training epochs is 2000.

For the proposed network architecture setup, the number of multi-head attention mechanisms used in the Transformer model is set at 8. The number of Transformer blocks N_{bk} shown in Figure 2 is set to 1. The hidden layer size in the Feed-forward layer is set to 64. For the MMD loss module, the number of

Gaussian kernels is set to 8. Table 4 sorted out the parameters set in this experiment.

5.2 Results

Table 5 shows the RMSE errors of different transfer learning tasks on the test sets of different battery datasets. Compared with task T_0, the tasks from T_5 to T_{10} transfer learning combination with two types of battery datasets as the target domain has the lowest test error value, and the accuracy of T_5 and T_8 is better than that of other transfer tasks. The second accuracy is to set a type of battery dataset transfer tasks from T_1 to T_4, and the test error is reduced to a certain extent compared with T_0. However, transfer tasks from T_{11} to T_{14} showed a decrease in prediction accuracy. These tasks chose 3 or 4 types of batteries as the target domains, and the prediction results showed high prediction error values in some target domains. In addition, in addition to showing the above statistics, in order to show the overall distribution of errors, Figure 8 shows the distribution of the absolute error $|y-\hat{y}|$ between the predicted and actual values. The figure shows the error distribution of different transfer learning tasks on the test set. The chart removes 1% of the predicted outliers. As can be seen from the figure, when two types of battery data are selected as the target domain, the method proposed in this study has the best performance on different test sets, which is consistent with the results in Table 5 and Table 6.

Table 6 shows the prediction results of selecting specific target domain data as research objects under the conditions of two source domains. The experiment shows that

Table 5. The RMSE values (%) of the test sample.

Transfer tasks	T_0	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}	T_{13}	T_{14}
Battery Group 2	1.404	1.231	1.221	1.327	1.709	**0.959**	1.008	1.004	0.959	2.197	2.403	1.977	2.221	1.372	1.091
Battery Group 3	1.551	2.003	1.698	1.982	3.178	**1.530**	1.631	1.500	1.464	1.921	3.599	5.270	1.960	3.520	1.547
Battery Group 4	1.104	0.735	0.642	0.693	0.961	**0.480**	0.481	0.464	0.472	2.405	2.062	1.803	1.064	0.610	0.634
Battery Group 5	1.084	1.032	3.843	1.670	1.768	**0.603**	0.669	1.094	0.648	1.757	1.861	2.093	8.633	1.263	1.056
Battery Group 6	0.855	0.732	0.668	**0.198**	0.805	**0.152**	0.403	0.239	0.183	1.652	0.789	1.527	0.761	0.685	0.421

Table 6. The RMSE value (%) of the test sample under two training source domains.

Transfer tasks	T_0'	T_1'	T_2'	T_3'	T_4'	T_5'	T_6'	T_7'	T_8'	T_9'	T_{10}'	T_{11}'	T_{12}'	T_{13}'	T_{14}'	T_{15}'
Battery Group 3	1.620	1.562	1.481	1.803	1.429	2.086	3.391	**1.220**	**1.161**	**1.216**	1.627	1.207	1.665	1.312	1.203	2.004
Battery Group 4	0.464	0.403	**0.319**	0.400	**0.342**	1.396	0.653	0.366	**0.267**	0.345	0.595	0.371	1.505	0.770	0.345	0.550
Battery Group 5	0.580	**0.553**	0.709	0.686	0.546	1.107	1.444	1.776	**0.421**	0.594	0.648	1.175	1.026	0.633	**0.442**	0.817
Battery Group 6	0.348	0.292	0.427	0.391	0.226	0.828	1.061	0.300	**0.132**	0.226	0.547	**0.213**	0.750	0.454	**0.129**	0.329

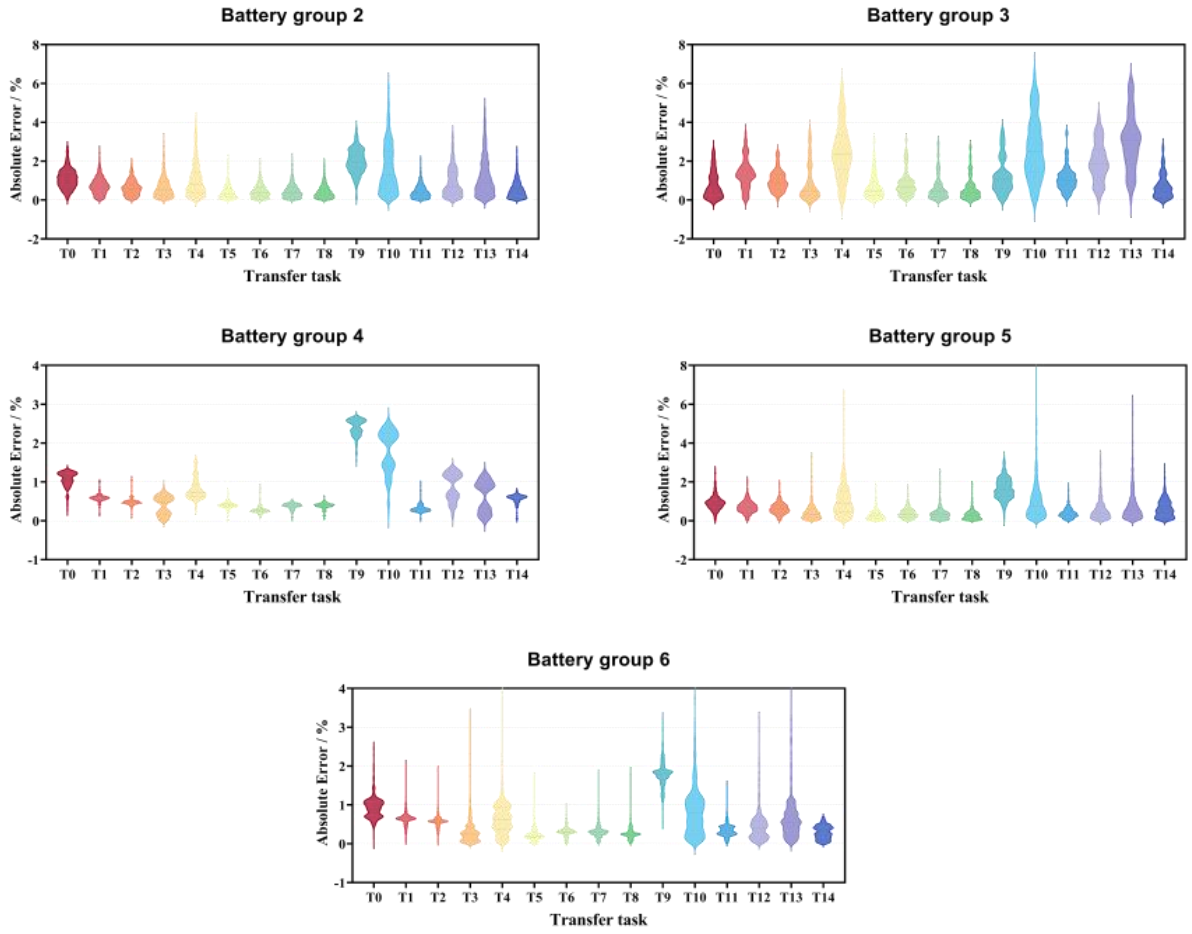

Fig. 8. Absolute error distribution of the target domain test sample.

the accuracy of the selected test is not necessarily improved compared with that of no transfer learning task T_0'.

In fact, This result is caused by multi-objective optimization. When faced with RMSE loss of the prediction accuracy of the representative source domain and MMD loss of the difference between multiple source domains and target domains, due to the easy internal conflict between targets, the optimization of one target is at the cost of the deterioration of other targets, so it isn't easy to find a single optimal solution, instead, a compromise is made between them. So as to make the overall goal as optimal as possible.

At the same time, the experimental results show that when some types of battery datasets are selected as the target domain, the accuracy of other battery datasets that are not involved in transfer learning may also be improved to some extent. This shows that the proposed model has learned the typical characteristics of various types of batteries during the learning process, and provides a possibility for the subsequent establishment of a predictive model that can be compatible with more battery datasets.

6 Conclusions

In this study, a method for predicting SOH for multiple types of LIBs is proposed. Based on

the Transformer network architecture, the proposed method introduces the maximum mean discrepancy loss term between multiple types of battery datasets for training. The experimental results show that the model achieves the highest accuracy on two types of battery datasets, which is a promising prediction method.

Results also show that by setting a specific weight combination of loss terms, the prediction accuracy of the target domain can be improved. Therefore, future work will focus on the variety of weight loss terms and design a specific algorithm to enable the model to automatically adjust the weight parameters before the loss terms, so as to establish a high-precision model suitable for more battery predictions, and further explore the realization of reliable online prediction methods for LIBs. At the same time, the future work will also adopt some methods to screen and save the prediction model with high prediction accuracy. The accuracy of the model is further optimized through the process of online verification.

References

1. Ahwiadi, M., & Wang, W., "An enhanced particle filter technology for battery system state estimation and RUL prediction," Measurement **191,** 110817 (2022).

2. Bai, S., Bi, X., Han, C., Zhou, Q., Shang, W.-L., Yang, M., He, H., "Evaluating R&D efficiency of China's listed lithium battery enterprises," Frontiers of Engineering Management **9(3),** 473-485 (2022).

3. Chemali, E., Kollmeyer, P. J., Preindl, M., Fahmy, Y., & Emadi, A., "A convolutional neural network approach for estimation of li-ion battery state of health from charge profiles," *Energies* **15(3),** *1185.* (2022).

4. Chen, D., Hong, W., & Zhou, X., "Transformer network for remaining useful life prediction of lithium-ion batteries," *Ieee Access* **10,** 19621-19628 (2022).

5. Chen, H., Zhang, T., Gao, Q., Han, Z., Jin, Y., Li, L., Xu, X., "Assessment and management of health status in full life cycle of echelon utilization for retired power lithium batteries," *Journal of Cleaner Production* 134583 (2022).

6. Chen, P.-C., Tsai, H., Bhojanapalli, S., Chung, H. W., Chang, Y.-W., & Ferng, C.-S., "A simple and effective positional encoding for transformers," *arXiv preprint* 2104.08698 (2021).

7. Devie, A., Baure, G., & Dubarry, M., "Intrinsic variability in the degradation of a batch of commercial 18650 lithium-ion cells," *Energies* **11(5),** 1031 (2018).

8. Dong, L., Xu, S., & Xu, B., "*a no-recurrence sequence-to-sequence model for speech recognition," Speech-transformer* (2018).

9. Fan, Y., Xiao, F., Li, C., Yang, G., & Tang, X., "A novel deep learning framework for state of health estimation of lithium-ion battery," *Journal of Energy Storage **32,*** 101741 (2020).

10. Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., & Smola, A, "A kernel two-sample test," *The Journal of Machine Learning Research* **13(1),** 723-773 (2012).

11. Hou, S., Yi, M., Jiang, F., Lu, L., Ren, J., Fuhrmann, M., . . . Lai, X., "Ultrasonic testing-based method for segmental calibration and quantitative estimation of the electrolyte content in lithium-ion batteries," *Measurement* 113101 (2023).

12. Hu, C., Ye, H., Jain, G., & Schmidt, C., "Remaining useful life assessment of lithium-ion batteries in implantable medical devices," *Journal of Power Sources* **375,** 118-130 (2018).

13. Juarez-Robles, D., Jeevarajan, J. A., & Mukherjee, P. P., "Degradation-safety analytics in lithium-ion cells: Part I. Aging under charge/discharge cycling," *Journal of The Electrochemical Society* **167(16),** 160510 (2020).

14. Kennedy, D., & Philbin, S. P., "Techno-economic analysis of the adoption of electric vehicles," *Frontiers of Engineering Management* **6(4),** 538-550 (2019).

15. Kim, S., Choi, Y. Y., Kim, K. J., & Choi, J.-I., "Forecasting state-of-health of lithium-ion batteries using variational long short-term memory with transfer learning," *Journal of Energy Storage **41***, 102893 (2021).

16. Kingma, D. P., & Ba, J. Adam:, "A method for stochastic optimization," *arXiv preprint arXiv 1412.6980* (2014).

17. Li, P., Zhang, Z., Grosu, R., Deng, Z., Hou, J., Rong, Y., & Wu, R., "An end-to-end neural network framework for state-of-health estimation and remaining useful life prediction of electric vehicle lithium batteries," *Renewable and Sustainable Energy Reviews **156***, 111843 (2022).

18. Li, Y., Li, K., Liu, X., Wang, Y., & Zhang, L., "Lithium-ion battery capacity estimation—A pruned convolutional neural network approach assisted with transfer learning," *Applied Energy **285*****,** 116410 (2021).

19. Li, Y., & Tao, J., "*CNN and transfer learning based online SOH estimation for lithium-ion battery,"* Chinese Control And Decision Conference (CCDC) (2020).

20. Liu, S., Duan, X., Wang, S., Zhang, J., & Hao, Y., "Optimization of dual field plate AlGaN/GaN HEMTs using artificial neural networks and particle swarm optimization algorithm," *IEEE Transactions on Device and Materials Reliability* (2023).

21. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., . . . Guo, B., *"Hierarchical vision transformer using shifted windows," Swin transformer* (2021).

22. Liu, Z., Zhao, Z., Qiu, Y., Jing, B., & Yang, C., "State of charge estimation for Li-ion batteries based on iterative Kalman filter with adaptive maximum correntropy criterion," *Journal of Power Sources **580*****,** 233282 (2023).

23. Ma, G., Xu, S., Yang, T., Du, Z., Zhu, L., Ding, H., & Yuan, Y., "A transfer learning-based method for personalized state of health estimation of lithium-ion batteries," *IEEE Transactions on Neural Networks and Learning Systems* (2022).

24. Mennik, F., Dinç, N. İ., & Burat, F., "Selective recovery of metals from spent mobile phone lithium-ion batteries through froth flotation followed by magnetic separation procedure," *Results in Engineering **17*****,** 100868 (2023).

25. Ouyang, T., Ye, J., Xu, P., Wang, C., & Xu, E., "Estimation of state-of-charge and state-of-health for lithium-ion battery based on improved firefly optimized particle filter," *Journal of Energy Storage **68*****,** 107733 (2023).

26. Pinto, G., Wang, Z., Roy, A., Hong, T., & Capozzoli, A., "Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives," *Advances in Applied Energy **5*****,** 100084 (2022).

27. Shi, Q., Liu, C., & Xiao, C., "Machine learning in building energy management: A critical review and future directions," *Frontiers of Engineering Management* **9(2),** 239-256 (2022).

28. Su, L., Xiong, L., & Yang, J., "Multi-Attn BLS: Multi-head attention mechanism with broad learning system for chaotic time series prediction," *Applied Soft Computing* ***132*****,** 109831 (2023).

29. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I., "Attention is all you need," *Advances in neural information processing systems* 30 (2017).

30. Wang, H., Li, Y.-F., & Ren, J., "Machine learning for fault diagnosis of high-speed train traction systems: A review," *Frontiers of Engineering Management* 1-17 (2023).

31. Wang, W., & Lu, Y., "*Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model,"* (2018).

32. Wen, J., Chen, X., Li, X., & Li, Y., "SOH prediction of lithium battery based on IC curve feature and BP neural network," *Energy* ***261*****,** 125234 (2022).

33. Xing, Y., Ma, E. W. M., Tsui, K.-L., & Pecht, M., "An ensemble model for predicting the remaining useful performance of lithium-ion batteries," *Microelectronics Reliability* **53(6),** 811-820 (2013).

34. Yao, J., & Han, T., "Data-driven lithium-ion batteries capacity estimation based on deep transfer learning using partial segment of charging/discharging data," *Energy* ***271*****,** 127033 (2023).

35. Yao, S., Kang, Q., Zhou, M., Rawa, M. J., & Abusorrah, A., "A survey of transfer learning for machinery diagnostics and prognostics," *Artificial Intelligence Review* **56(4),** 2871-2922 (2023).

36. Zhang, J., Li, X., Tian, J., Luo, H., & Yin, S., "An integrated multi-head dual sparse self-attention network for remaining useful life prediction," *Reliability Engineering & System Safety* ***233*****,** 109096 (2023).

37. Zhang, W., Li, X., & Li, X., "Deep learning-based prognostic approach for lithium-ion batteries with adaptive time-series prediction and on-line validation," *Measurement* ***164*****,** 108052 (2020).

38. Zhang, Y., Liu, Y., Wang, J., & Zhang, T., "State-of-health estimation for lithium-ion batteries by combining model-based incremental capacity analysis with support vector regression," *Energy* ***239*****,** 121986 (2022).

39. 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* ***204*****,** 112093 (2022).

40. Zhou, K., Zhang, Z., Liu, L., & Yang, S., "Energy storage resources management: Planning, operation, and business model," *Frontiers of Engineering Management* **9(3),** 373-391 (2022).

41. Zhu, Y., Li, X., & Zhang, W., "*Deep Residual Net-Based Prognosis Method for Lithium-ion Batteries with Information Fusion from Different Scales,"* (2022).

42. Zhu, Y., Li, X., Zhang, Y., & Zhang, W., "Cross-domain prognostic method of lithium-ion battery in new energy electric aircraft with domain adaptation," *IEEE Sensors Journal* (2023).