

Article

Domain Generalization Using Maximum Mean Discrepancy Loss for Remaining Useful Life Prediction of Lithium-Ion Batteries

Wenbin Li * , Yue Yang  and Stefan Pischinger 

Chair of Thermodynamics of Mobile Energy Conversion Systems, RWTH Aachen University, Forckenbeckstraße 4, 52074 Aachen, Germany; yang_yue@tme.rwth-aachen.de (Y.Y.); pischinger_s@tme.rwth-aachen.de (S.P.)

* Correspondence: li_w@tme.rwth-aachen.de

Abstract: The capacity of Lithium-ion batteries degrades over the time, making accurate prediction of their Remaining Useful Life (RUL) crucial for maintenance and product lifespan design. However, diverse aging mechanisms, changing working conditions and cell-to-cell variation lead to the inhomogeneous cell lifespan and complicated life prediction. In this work, a data-driven algorithm based on stacked Long Short Term Memory (LSTM) encoder-decoders is proposed for RUL prediction. The encoder and upstream decoder form an autoencoder framework for feature extraction. The encoder and the downstream decoder form the encoder-decoder framework for RUL prediction. To enhance generalization during training, the Maximum Mean Discrepancy (MMD) loss is included in the autoencoder framework. The similarity of aging patterns is analyzed during splitting source and target datasets through k-means and Density-Based Spatial Clustering of Applications with Noise (DBSCAN). The Euclidean metric with accumulated Equivalent Cycle Number (ECN) sequence during aging shows better performance for similarity-based data splitting than the Dynamic Time Wrapping (DTW) distance metric based on capacity fading trajectory. The experimental results indicate that the proposed algorithm can provide accurate RUL prediction using 5% fading data and shows good generalization with Coefficient of Determination (R^2) score of 0.98.



Academic Editors: Diptish Saha, Juan C. Vasquez and Najmeh Bazmohammadi

Received: 8 April 2025

Revised: 3 May 2025

Accepted: 12 May 2025

Published: 14 May 2025

Citation: Li, W.; Yang, Y.; Pischinger, S. Domain Generalization Using Maximum Mean Discrepancy Loss for Remaining Useful Life Prediction of Lithium-Ion Batteries. *Batteries* **2025**, *11*, 194. <https://doi.org/10.3390/batteries11050194>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: lithium-ion battery (LIB); RUL; time series forecasting; domain adaption

1. Introduction

Lithium-Ion Battery (LIB)s are used in a wide range of applications due to their high energy density, excellent power response performance and long lifespan, including electric vehicles and stationary applications [1]. Like many other electrochemical storage systems, the available capacity and power capability of batteries decrease during their lifespan. However, the aging test takes lots of time to determine the reference for the lifespan design of the entire system. Therefore, an accurate lifetime prediction is required for in-time maintenance and to shorten the development cycle [2]. The target of RUL prediction is to give the remaining number of cycles in the early aging stage with short history data.

The degradation of LIBs is complicated with multiple effects coupled together, which is also influenced by electrode materials and preparation processes [3,4]. Many previous works have simulated the battery lifetime and disclosed various aging mechanisms, including loss of active materials [5–8], loss of lithium inventory [6–8] and growth of Solid Electrolyte Interface (SEI) film [7–10]. As shown in Table 1, the common ways to predict RUL can be divided into physical models [11,12], empirical models [13–16] and data-driven methods [17–21].

Physical models derive capacity degradation by considering various aging mechanisms caused by side reactions in the source terms of partial differential equations [11]. Prada et al. [12] proposed an electrochemical thermal aging model to derive the capacity loss and resistance growth. Various aging mechanisms are considered, including the loss of active materials at the anode and the cathode, loss of lithium inventory and SEI growth. In the aging model, the operating conditions like C-rate and Depth of Discharge (DoD) are transformed into over-potential and embedded in the format of Arrhenius equation together with temperature to calculate the degradation rate in (1):

$$i_s = A \cdot \exp\left(-\frac{B}{R \cdot T} \cdot (\Phi - C \cdot i_{app})\right) \quad (1)$$

where i_s is the side reaction current for degradation, i_{app} is the cell current, Φ is the half-cell potential, R is the constant for ideal gas, T stands for the temperature and A , B and C are the coefficients to be parameterized.

Physical models consider specific degradation individually and can precisely predict degradation under multiple aging stresses. However, aging models become complex due to increasing source terms as additional aging mechanisms are included. Meanwhile, the parameterization work for the coefficients in source terms is also challenging through full cell tests due to the difficulty of observing real-time internal battery dynamics [22].

Empirical models divide capacity degradation into calendric aging Q_{loss}^{cal} and cyclic aging Q_{loss}^{cyc} , which are calculated using semi-empirical equations in (2) and (3) [13] fitted by experimental data [15]:

$$Q_{loss}^{cal} = B_{cal}(SOC) \cdot \exp\left(-\frac{E_{cal}}{R \cdot T}\right) \cdot t^{z_{cal}} \quad (2)$$

$$Q_{loss}^{cyc} = B_{cyc}(I) \cdot \exp\left(-\frac{E_{cyc} + \alpha \cdot |I|}{R \cdot T}\right) \cdot Ah^{z_{cyc}} \quad (3)$$

where B_{cal} and B_{cyc} are coefficients related to storage State of Charge (SOC) and cycling C-rate, E_{cal} and E_{cyc} are activation energies for calendric and cyclic aging, R is the constant for ideal gas, T stands for the thermodynamic temperature, α is the accelerating coefficient depending on C-rate, z_{cal} and z_{cyc} are exponential terms for degradation due to SEI growth and the diffusion limited process, t is the storage time for calendric aging and Ah is the charge throughput during cycling. The empirical model proposed in [14] considers the effects of various operation factors, including storage SOC, DoD, temperature and C-rate. Instead of the absolute capacity loss, the degradation rate caused by each aging factor is calculated using an equation similar to the Arrhenius function in (2) and (3). The total capacity loss rate is considered as the accumulating product of each effect. The parameters in (2) and (3) are fitted by experimental data as functions of equivalent operation conditions through time averaging by rainflow counting. In [16], the degradation coefficients are considered in a similar way. However, the fitting is implemented through polynomial functions.

Empirical models predict degradation by curve fitting using polynomial and Arrhenius equations and consider various aging factors in the coefficients without detailed knowledge of aging mechanisms [14,16]. However, the size of coefficients also increases and the fitting accuracy decreases as complicated degradation patterns caused by varying operation conditions are considered.

Data-driven models based on time series forecasting capture various aging patterns and predict RUL by leveraging historical aging data through data mining. The algorithms are optimized through backward propagation and gradient descent instead of complex parameterization in physical models and empirical models. Time series data-driven models are typically categorized into single-shot forecasting and iterative prediction [23] as shown in Table 1. The iterative framework based on LSTM in [24] utilizes previous prediction

results as input for subsequent forecasts until the outputs approach End of Life (EOL). However, the deviation tends to increase in long-term predictions due to the accumulating errors during iterations. The single-shot structure in [17] processes historical aging data and generates the entire prediction sequence in one step. The accuracy of the single-shot framework increases when the starting prediction point is located at later aging stage but also encounters decreased accuracy for long sequence prediction due to the limited availability of EOL degradation information in the early aging data.

The diversity of aging trajectories due to varying aging stresses presents another challenge for data-driven methods. Numerous studies [18–21] demonstrate good prediction accuracy using the NASA [25] and CALCE datasets [26], while the battery lifespan shows limited variation ranges. The total cycle number in the CALCE dataset ranges from 400 cycles to 1000 cycles, and the total cycle number in the NASA dataset ranges from 100 to 200 cycles. However, a newly published open-source dataset [27] reveals a more significant diversity in aging cycles, ranging from 50 cycles to over 8000 cycles. Increasing attention is focused on the adaptability of RUL algorithms to inhomogeneous battery lifespans.

To deal with the divergence of lifespan, Transfer Learning (TL) technologies such as Fine-Tuning (FT) [28–30] and feature-based learning [31] are receiving increasing attention in recent works. Feature-based TL is focused on improving the adaptation of feature extraction between source domain and target domain. In [31], the researcher applies a Simsiam model to RUL prediction by sharing weight of the feature extraction components between cells with similar aging behavior. The proposed model shows good accuracy using a dataset with inhomogeneous lifespans ranging from 800 to 1895 cycles. The similarity in degradation is determined according to the absolute difference of the total cycle number, while the aging process in between is ignored. The LSTM-based model with FT in [28–30] freezes the upstream layers in the pre-trained model and replaces the downstream part with newly built layers trained by partial target domain data with similar degradation patterns to the source domain. The Euclidean distance is used to determine the similarity of the degradation curves. However, the Euclidean metric requires time series to have the same length and sampling rate, which is not practical in the application with inhomogeneous lifespans. In addition, the model modified by FT does not fit another target dataset with different aging patterns when the entire dataset shows huge lifespan divergency and is divided into multiple degradation categories.

Table 1. RUL algorithm overview.

Algorithm	Advantages	Challenges
Physical model [12]	Detailed aging mechanisms under different operating conditions are considered	Difficult parameterization, increasing parameter size and decreasing accuracy under complicated aging patterns [22,32]
Empirical model [13,14,16]	Consider various aging factors through curve fitting ignoring detailed degradation mechanisms	
Iterative data-driven model [24]	Use aging history data for degradation prediction under various working conditions	Accumulating prediction error during iterations [23]
Single-shot data-driven model [17]		Increasing deviation for long sequence prediction with early aging data [23]

The lifespan variation under varying aging factors and the data splitting of source and target datasets are not discussed in detail in the other works. Varying prediction steps caused by different lifespans present a significant challenge for the adaptability of the data-driven model, which ultimately leads to a decreasing accuracy of RUL prediction. In addition, the varying lifespan also makes it a great challenge to determine the similarity of degradation patterns while splitting the source and domain datasets for TL since the commonly used Euclidean distance-based metric requires time series with the same length and sampling rate. To this end, the incremental aging cycle sequence is used as the input of the RUL model in this work instead of the capacity fading trajectory. The aging cycle sequence is derived from the accumulating ECNs from Beginning of Life (BOL) to EOL based on a uniform State of Health (SOH) interval. Since the EOL threshold in the application is usually fixed, the aging cycle sequence can also be treated as a fixed-length time series with SOH recognized as the time axis. The MMD loss is used as the regularization term for TL instead of FT to improve the generality of feature representation extracted from the input ECN sequence so that the model can be adapted to multiple categories with different aging patterns determined by the Euclidean distance of the ECN sequence.

The dataset is firstly split into source domain and target domain through clustering based on the Euclidean distance of aging cycle sequence, where the centroid-based clustering (k-means) and the density-based clustering (DBSCAN) are compared. During the training process, the aging cycle sequences from source and target domain are fed into LSTM models for feature learning and RUL prediction. The proposed model in Figure 1 consists of one encoder and two decoders based on LSTM. The encoder and prediction decoder form the sequence-to-sequence framework for single-shot RUL prediction and the encoder and reconstruction decoder form the autoencoder framework for feature learning. In the autoencoder framework, the MMD metric is used to minimize the discrepancy of the feature presentation between target domain data and source domain data. The extracted features are fed to the prediction decoder for RUL prediction in a sequence-to-sequence framework. After training, the sequence-to-sequence framework can be extracted for RUL application. The main contributions are summarized as follows:

- (1) An incremental aging cycle sequence with fixed length is proposed as the input of the time series RUL prediction model. In this way, the RUL problem is transformed from a variable-length time series forecasting problem to a constant-length time series forecasting problem.
- (2) The original dataset is split into source domain and target domain through clustering based on the Euclidean distance metric. The split data within the same category show excellent homogeneity.
- (3) An autoencoder embedded with MMD loss is implemented for domain generalization to improve the adaptability of RUL prediction for batteries with divergent lifespans aged under various degradation impacts considering C-rate, DoD, temperature and storage SOC.

The remainder of this article is structured as follows. The architecture of the RUL prediction model is introduced in Section 2. In Section 3 the training and validation results are analyzed. Section 4 presents the main conclusions.

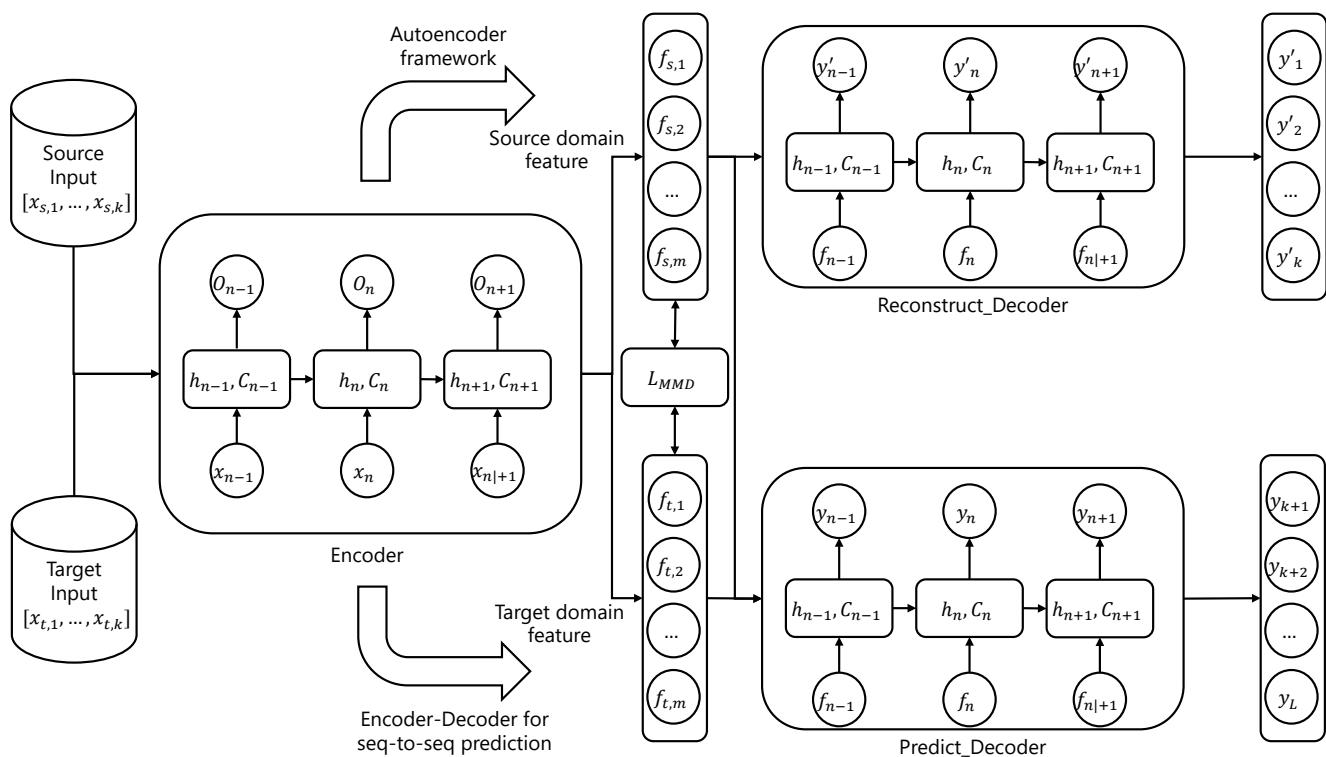


Figure 1. RUL prediction based on LSTM encoder–decoder framework.

2. Methods

2.1. Data Pre-Processing

An aging dataset from KIT [27] is used, which provides cyclic degradation and calendric degradation under various aging factors, including temperature, C-rate, DoD and SOC. The aging test is conducted upon Nickel Manganese Cobalt (NMC) batteries with a nominal capacity of 3 Ah. As illustrated in Figure 2a, the lifespan shows huge divergency ranging from 50 to 8700 cycles. The degradation accelerates when the aging test involves charging at low temperatures and cycling at high temperature.

As shown in Figure 2b, the capacity regeneration can be observed during the aging test. Therefore, the aging process can be divided into local regeneration and global degradation [33–35]. However, the battery loses the regenerated capacity quickly in several cycles and the mechanism behind this is still not clear. Therefore, more attention is paid to global degradation in RUL, including the capacity trajectory and direct lifespan prediction till specified EOL thresholds. To this end, the Discrete Wavelet Transform (DWT) algorithm in Figure 3 is used to divide local regeneration and global degradation. The input data are decomposed into detail coefficients, the components in high-frequency bandwidths, and approximation coefficients, the components in low-frequency bandwidths.

Figure 2b shows the reconstructed capacity fading curves based on the detail coefficients of each decomposition level and the approximation coefficients. As illustrated on the left side, the wavelet transform decomposes the input data into several frequency bandwidths from high frequency (level 1) to low frequency (level 4). The Daubechies 4 wavelet function is used for DWT, and the detail coefficients are processed by an adaptive soft threshold method in (4) and (5) before the reconstruction of the capacity fading trajectory.

$$T(x) = \begin{cases} x - \lambda, & \text{if } x > \lambda \\ x + \lambda, & \text{if } x < -\lambda \\ 0 & \text{if } |x| \leq \lambda \end{cases} \quad (4)$$

$$\lambda = \frac{1}{0.6745} \cdot x_{mid} \cdot \sqrt{2 \cdot \log(N)} \quad (5)$$

where x is the detail coefficient and $T(x)$ is the filter function. The threshold λ is adapted according to the median value x_{mid} of the coefficients and the length of detail coefficients N based on quartile fraction.

The SOH curve on the right side is reconstructed from the approximation coefficients and filtered detail coefficients. The measurement noise and the local regeneration are removed by the DWT and the reconstructed data are used to train the long-term RUL prediction model.

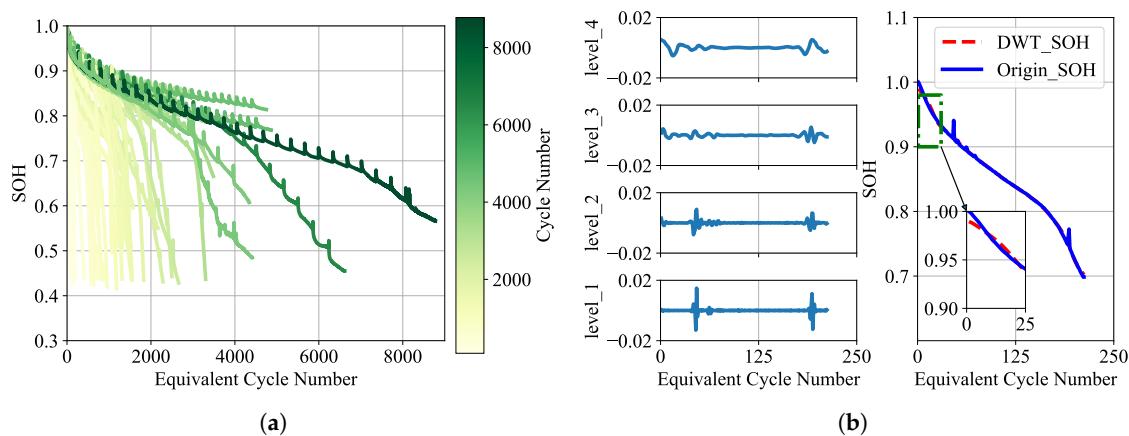


Figure 2. Battery capacity fading trajectories: (a) Capacity fading trajectories of all batteries. (b) Capacity fading processed by Wavelet Packet Decomposition.

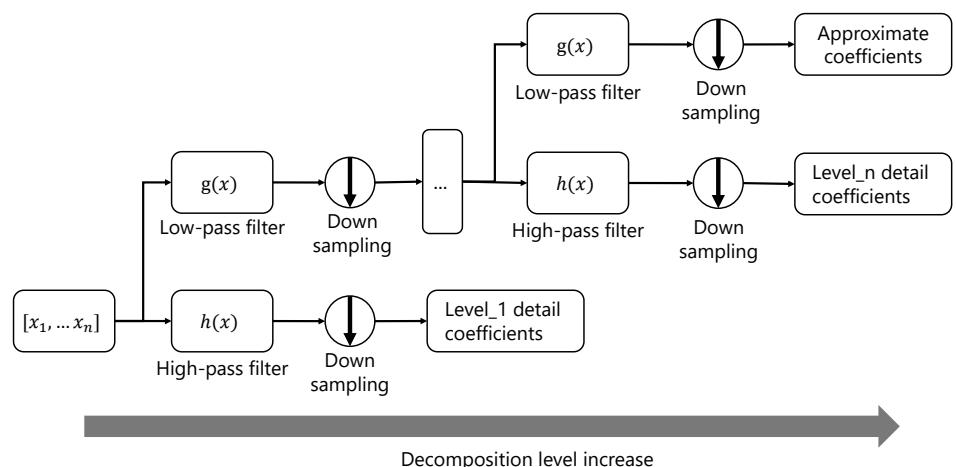


Figure 3. Signal decomposition based on wavelet transformation.

2.2. Extracting Aging Cycle Sequence

The capacity fading trajectory based on an equal ECN interval in the early aging stage is a commonly used input for the data-driven RUL model [18–21]. The SOH is the most easily obtainable data in comparison to the other processed signals, since the other proposed features such as capacity–voltage matrices in [31] and Incremental Capacity Analysis (ICA) peaks in [28] require long-term history data under a Constant Current (CC) profile.

Nevertheless, it is still a significant challenge to directly use capacity fading data with diverging lifespans in RUL algorithms considering the accumulation error for long-term prediction and difficulty of splitting source and target domain datasets. Therefore, the aging cycle sequence is proposed in this work as the input for single-shot long-term RUL prediction.

As illustrated in Figure 4, the aging cycle sequence is derived from accumulating ECN during degradation based on an equal interval of SOH degradation. The SOH fading trajectories processed by DWT are firstly resampled based on an equal interval. A 1% step is used in this work, which is also the common resolution in Battery Management Systems in reality. The ECN is resampled based on the processed SOH accordingly, and the accumulating aging cycle sequence with constant length is obtained.

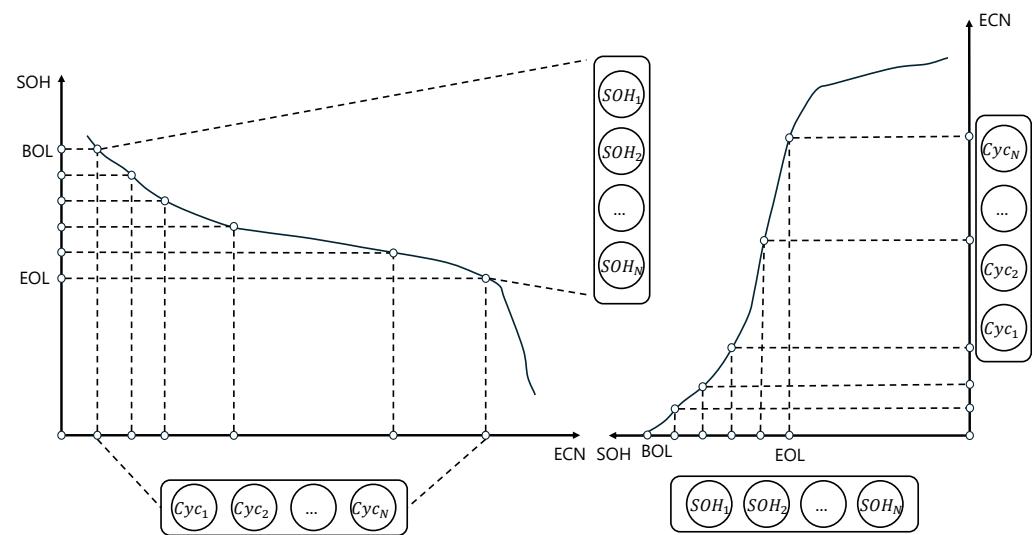


Figure 4. Aging cycle sequence extracted from SOH fading curve.

In addition, an aging cycle sequence also supports comparing similarities between time series based on Euclidean distance. Euclidean distance is the most commonly used similarity metric, which accumulates a squared deviation at each time step. However, Euclidean distance cannot evaluate the distance between curves with different lengths and sampling rates. To this end, DTW distance is proposed in [36] to evaluate the similarity of degradation, which rematches the curves with different sampling rates and lengths through dynamic programming and calculates the deviation between matched elements. However, the DWT distance does not follow the triangle rules when it is used to compare the distance. By using an aging cycle sequence, the length of sequence is fixed since it only depends on SOH threshold at EOL. The equal SOH interval can also be recognized as the fixed sampling rate of the aging cycle sequence. The benefits of the aging cycle sequence for similarity comparison will be discussed together with domain adaption in Section 3.1.

2.3. Domain Adaption in Autoencoder Framework

As illustrated in Figure 1, the encoder and reconstruction decoder are based on a stacked LSTM layer in the autoencoder framework to train the feature extraction. The encoder takes aging cycling as input and outputs extracted features to the decoder to reconstruct the input aging cycle sequence. Therefore, the encoder is trained to operate as a feature extractor for the following sequence-to-sequence RUL prediction.

The MMD loss L_{MMD} in (6) is introduced as the regularization term at the feature representation during training to minimize the deviation of features x_s and x_t extracted from source domain and target domain.

$$\begin{aligned} L_{\text{MMD},p} = & \frac{1}{m \cdot (m-1)} \sum_i^m \sum_{j \neq i}^m k_p(x_{s,i}, x_{s,j}) \\ & - \frac{2}{m \cdot n} \sum_i^m \sum_j^n k_p(x_{s,i}, x_{t,j}) \\ & + \frac{1}{n \cdot (n-1)} \sum_i^n \sum_{j \neq i}^n k_p(x_{t,i}, x_{t,j}) \end{aligned} \quad (6)$$

$$k_p(x_i, x_j) = \exp\left(-\frac{|x_i - x_j|^2}{2 \cdot \sigma_p^2}\right) \quad (7)$$

where m and n are the batch size of the source domain data and target domain data, and k stands for the kernel function used to calculate the divergency. The Gaussian function in (7) is the commonly used kernel function in MMD loss, and σ^2 is the bandwidth in Gaussian kernel.

The kernel function converges to zero when the deviation term in (7) exceeds $3 \cdot \sigma_p$, indicating a degradation in kernel value and a saturation of MMD loss while the features from x_s and x_t show great deviation if a fixed bandwidth is used. In addition, the degradation of the kernel value also indicates that the Gaussian kernel function is sensitive to deviation within $2 \cdot \sigma_p$ range, since the kernel value is already lower than 0.05 at $3 \cdot \sigma_p$ point. To this end, a multi-kernel MMD loss $L_{\text{mk-MMD}}$ in (8) is used in this work, where the bandwidth of each kernel is updated during the training process. As shown in (9), the bandwidth of each Gaussian kernel scales exponentially with a base of two, where the average variance in the L2 distance between x_s and x_t is located in the center. The calculation complexity increases with increasing kernel amount N in (8) and (9). In this work, five kernel functions are used to calculate $L_{\text{mk-MMD}}$.

$$L_{\text{mk-MMD}} = \sum_{p=1}^N L_{\text{MMD},p} \quad (8)$$

$$\sigma_p^2 = \frac{\sum_i^{m+n} \sum_{j \neq i}^{m+n} (x_i - x_j)^2}{(m+n) \cdot (m+n-1)} \cdot 2^{p-\frac{N}{2}} \quad (9)$$

To realize domain adaption for feature extraction, the aging data need to be split into source domain and target domains according to the similarity of aging patterns. Therefore, DBSCAN and k-means algorithms are compared in this work to divide the aging dataset into different categories based on the Euclidean distance between aging cycle sequences in Section 2.2.

As shown in Figure 5, DBSCAN [37] is a density-based clustering algorithm that is specified by the neighbor region size and minimized clustering size. The points within the neighbor region are assigned as the same category and the categories with small clustering size are recognized as noise points. However, the amount of categories in centroid-based k-means clustering needs to be specified and the noise points in the DBSCAN algorithm will be assigned to the k-means cluster, which increases the inhomogeneity within the category. The detailed classification results will be discussed in Section 3.1. The category with the most data will be selected as the source domain and all the other samples will be used as the target domain in TL.

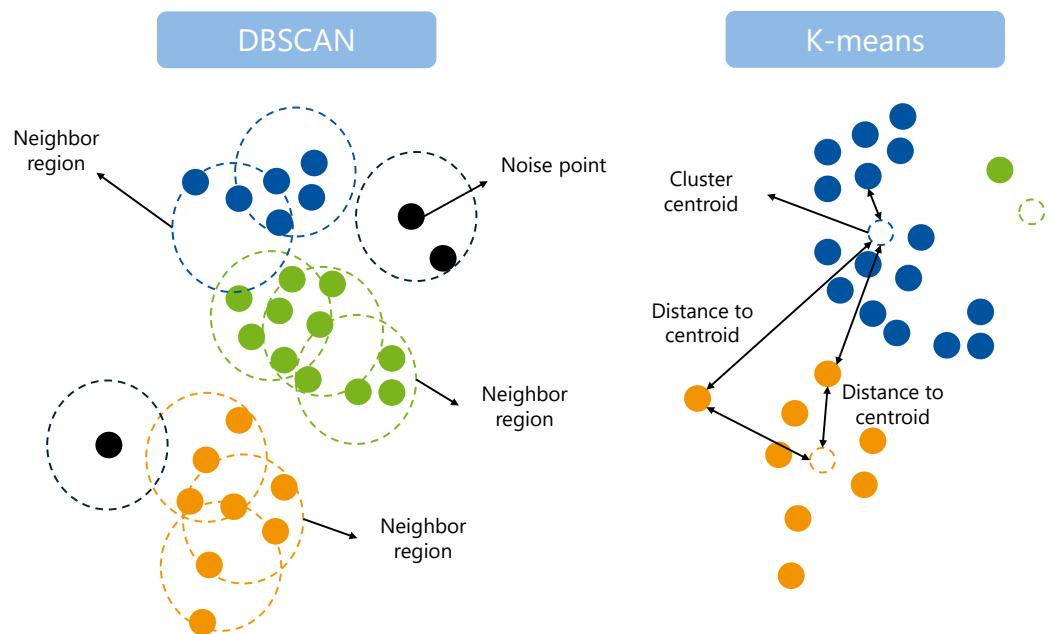


Figure 5. Comparison between k-means and DBSCAN.

2.4. RUL Prediction in Encoder–Decoder Framework

As illustrated in Figure 1, the encoder and predictor decoder based on stacked LSTM layers form the single-shot framework for RUL prediction. The encoder takes an aging cycle sequence in the early aging stage with an SOH degradation of 5% and extracts features for RUL prediction in the decoder. The feature presentation from the encoder is trained in autoencoder framework as mentioned in Section 2.3. The predictor decoder takes the features and gives an aging cycle sequence prediction till EOL.

The Mean Squared Error (MSE) loss in (10) is used to train the decoder, where N is the batch size of the input data, y_{pred} is the output of decoder and y_{true} is the ECN in the aging cycle sequence.

$$\text{MSE} = \frac{1}{N} \cdot \sum_i (y_{pred,i} - y_{true,i})^2 \quad (10)$$

3. Results and Discussion

The dataset from KIT [27] using an NMC battery with a nominal capacity of 3 A h is cited. The KIT dataset combines cyclic and calendric aging, considering various aging stresses including C-rate, DoD, temperatures and storage SOC. The detailed testing profile can be found in [27].

3.1. Classification of Different Aging Patterns

To implement domain adaption, 89 cell samples from the KIT dataset that have been aged to 80% SOH are split into source and target domains through clustering. Two commonly used distance metrics, namely the DTW and the Euclidean distance, are compared using k-means and DBSCAN clustering. The DTW distance is derived from the SOH fading sequence on the left side of Figure 4 and the Euclidean distance is derived from the aging cycle sequence on the right side of Figure 4. The following equations are used to calculate Euclidean and DTW distance:

$$d(x_i, y_i) = \sqrt{(x_i - y_i)^2} \quad (11)$$

$$D_{\text{euclidean}}(X, Y) = \sqrt{\sum_i (d(x_i, y_i))^2} \quad (12)$$

$$D_{\text{DTW}}(x_i, y_j) = d(x_i, y_j) + \min \begin{cases} D_{\text{DTW}}(x_{i-1}, y_{j-1}) \\ D_{\text{DTW}}(x_i, y_{j-1}) \\ D_{\text{DTW}}(x_{i-1}, y_j) \end{cases} \quad (13)$$

where $d(x_i, y_i)$ stands for the distance of each dimension in vectors $X = (x_1, x_2, \dots)$ and $Y = (y_1, y_2, \dots)$, $D_{\text{euclidean}}(X, Y)$ stands for the Euclidean distance between vectors X and Y and $D_{\text{DTW}}(x_i, y_j)$ stands for the accumulated DTW distance from pair (x_1, y_1) to (x_i, y_j) .

The distance matrix heat maps in Figure 6 illustrate the DTW distance of SOH fading sequences and Euclidean distance of aging cycle sequences pairwise, where the distance value is normalized between 0 and 1 separately. For the convenience of visualization, Multi-Dimension Scaling (MDS) is applied to the original distance matrix. The SOH fading curve and aging cycle sequences are downsized to a two-dimensional space as shown in Figure 7. MDS is a dimensionality reduction technique that keeps the similarity of pairwise distance between the original and transformed data. The reconstructed points from different categories are marked with different colors according to the k-means clustering result in Figure 8.

According to the heat map in Figure 6, the distance matrix from the Euclidean metric shows more significant contrast between aging samples than the matrix from DTW metric, especially for the cases with sample numbers around 62 and 85 in the Euclidean distance matrix. The MDS reconstructed clustering result in Figure 7 also indicates a more reasonable classification in the case of Euclidean distance-based clustering in the right sub-figure, where the reconstructed points from the same category show a more condense distribution and an obvious divergency between different clustering categories is observed.

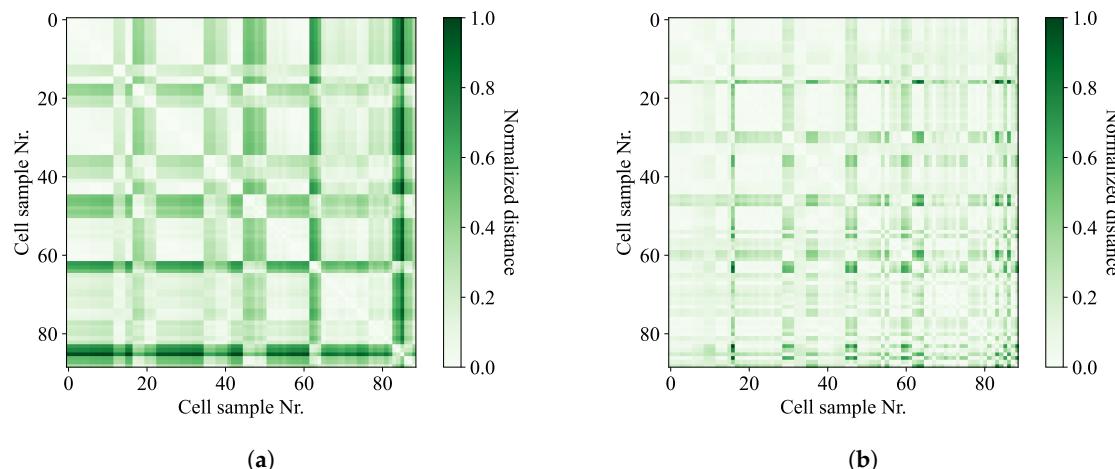


Figure 6. Distance-based clustering: (a) Distance matrix based on Euclidean metrics. (b) Distance matrix based on DTW metrics.

The clustering result of aging curves based on DTW and the Euclidean distance metrics using k-means and DBSCAN algorithms are illustrated in Figure 8. Similar to Figure 7, the k-means with Euclidean metric in Figure 8b also indicates a better clustering result. The divergency between different aging cycle sequences in the same aging cluster in Figure 8b are smaller than the case in Figure 8a based on DTW metric.

In addition to k-means, DBSCAN clustering results also indicate a better performance with the Euclidean metric in Figure 8d, where a more homogeneous aging pattern can be observed within the same aging category than the case in Figure 8c. The number of clusters in the k-means algorithm needs to be customized while the DBSCAN considers distance density and provides adaptive category number as shown in Figure 5, and the remaining

curves are considered as noise inputs. Considering the adaptive determination of cluster number, results from DBSCAN clustering are used in the following section and the noise data are used in the target domain validation.

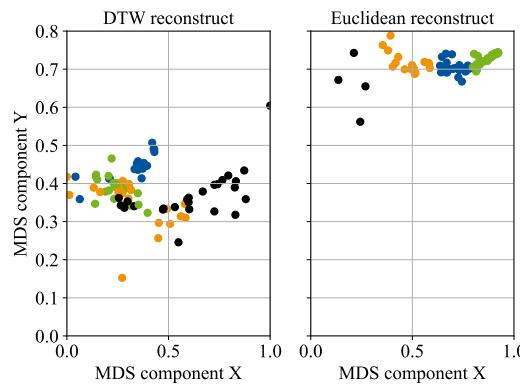


Figure 7. k-means clustering results in 2-D reconstructed space.

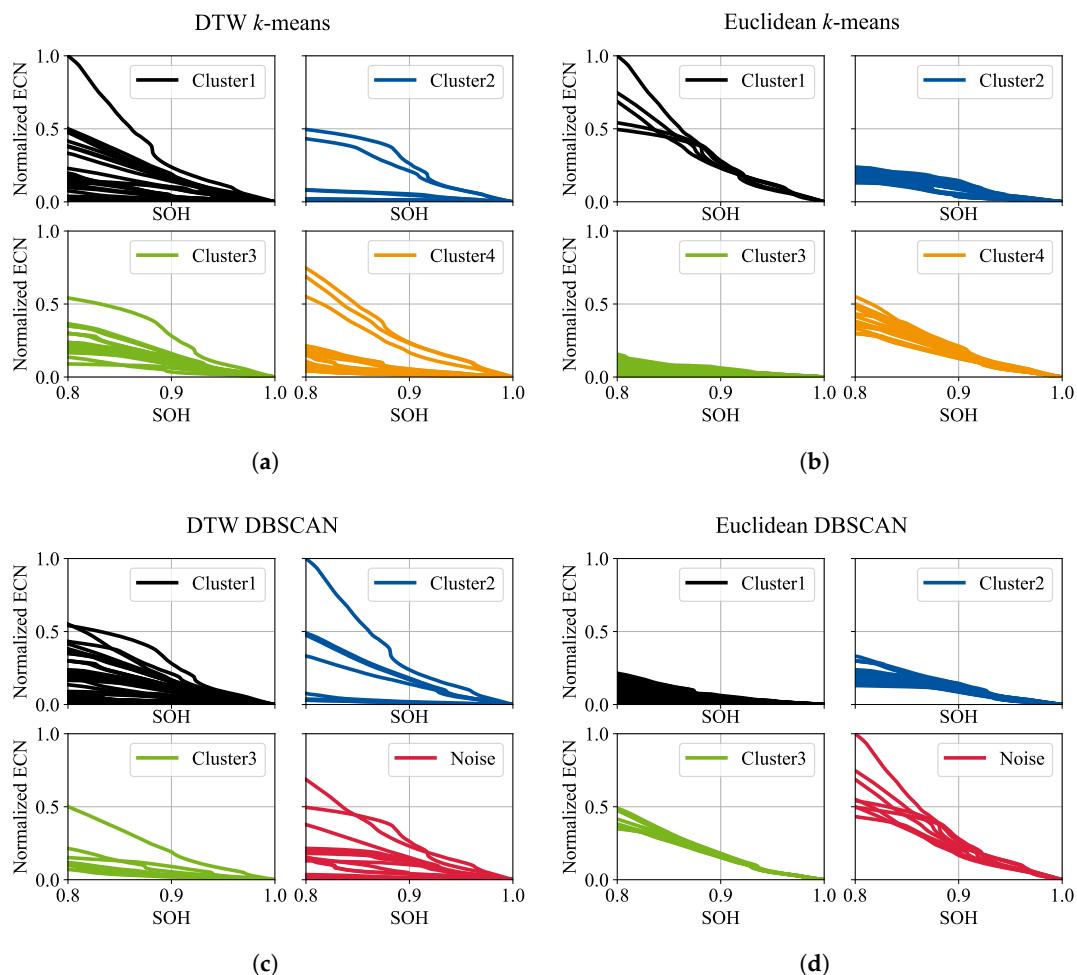


Figure 8. Clustering result using k-means and DBSCAN: (a) k-means clustering based on DTW distance. (b) k-means clustering based on Euclidean distance. (c) DBSCAN clustering based on DTW distance. (d) DBSCAN clustering based on Euclidean distance.

3.2. RUL Prediction Using Domain Adaption

Figure 1 shows three models with the same architecture but different training and prediction strategies implemented with aging cycle sequence for RUL prediction, namely the single-shot structure models with and without MMD adaption, and the iterative predic-

tion model with MMD adaption. The target domain and the source domain are divided into training, validation and testing datasets independently and randomly by a ratio of 0.5:0.2:0.3.

Figure 9a,b show the output of the LSTM encoder from single-shot models without and with MMD adaption. The encoder outputs are firstly scaled to the same range by a Min-Max scaler separately, and the mean values with variation range of each feature dimension are illustrated. The encoder features from the model with MMD show more centralized distribution with smaller variance on all feature dimensions, indicating a smaller divergency in the encoder feature presentation for the model with MMD loss.

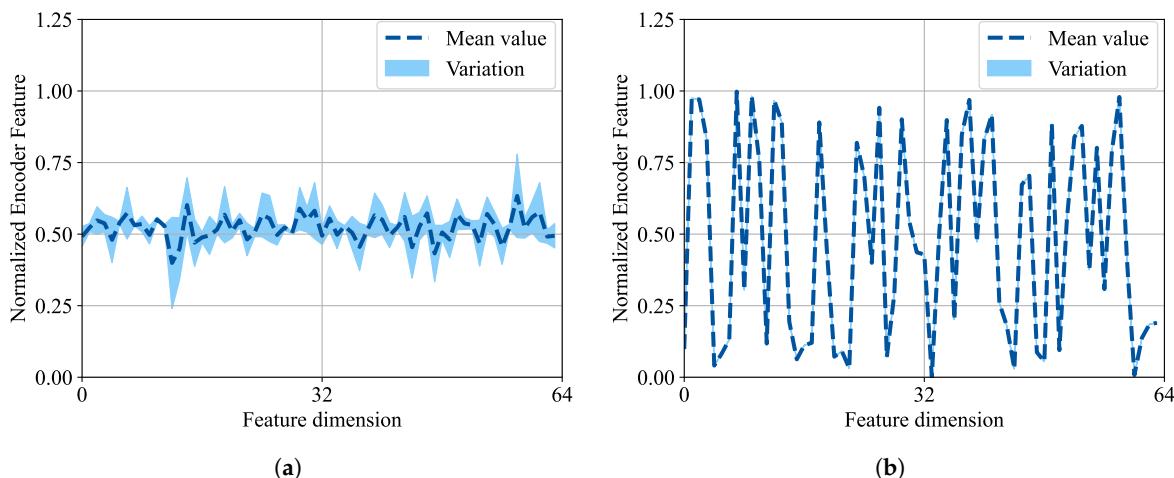


Figure 9. Encoder feature divergency comparison between models with and without MMD adaption: (a) encoder feature presentation without MMD; (b) encoder feature presentation with MMD.

The benefit of MMD domain adaption can also be shown by the RUL prediction results of all the cells in the testing sub-dataset as illustrated in Figure 10d. The RUL in the testing dataset ranges from 160 to 4000 ECN, which is related to SOH fading from 95% to 80% in Figure 2a. Due to the huge divergency in RUL, the R^2 in Equation (14) is used for evaluation, where y_{pred} is the predicted value, y_{true} is the measured value and \bar{y} is the mean value. The single-shot model with MMD loss obtains a better generalization with an R^2 score of 0.98 in comparison to the one without MMD loss with an R^2 score of 0.965.

$$R^2 = 1 - \frac{\sum_i (y_{true,i} - y_{pred,i})^2}{\sum_i (y_{true,i} - \bar{y}_{true})^2} \quad (14)$$

The results of the iterative RUL model with MMD adaption is shown in Figure 10d. The iterative model is also based on the encoder-decoder architecture in Figure 1. The encoder takes early aging data from 100% to 95% SOH for feature extraction, which is the same case as the single-shot model. However, the decoder takes the hidden state from the encoder and the predicted ECN from previous iterations as input for single-step prediction within each iteration till the EOL is reached. Due to the accumulating error, the iterative model shows increasing deviation of ECN prediction as it comes to the EOL. The R^2 score decreases from 0.98 to 0.91 and the maximum RUL deviation increases from 150 to over 600 ECN in Figure 10d when an iterative framework is used in the predictor decoder. According to the error distribution in Figure 11, the single-shot model with MMD loss also shows less Relative Error (RE) of ECN prediction. In comparison to the one without MMD loss with RE between $\pm 20\%$ and the iterative model with RE beyond -10% , most of the RE for the single-shot model with MMD concentrated within $\pm 10.8\%$ at 80% SOH.

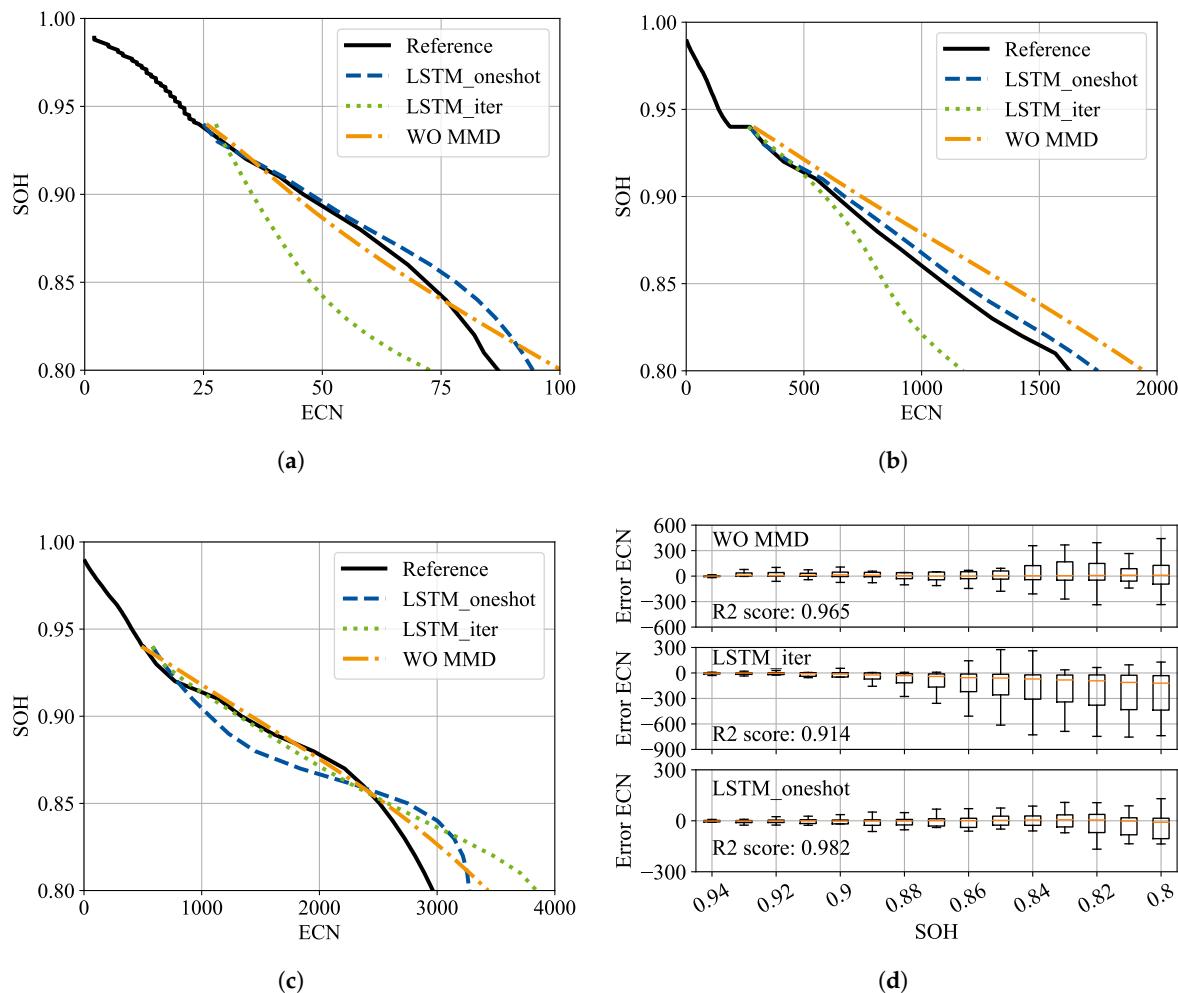


Figure 10. SOH fading prediction: (a) short lifespan example; (b) middle lifespan example; (c) long lifespan example; (d) prediction deviation for all testing cells.

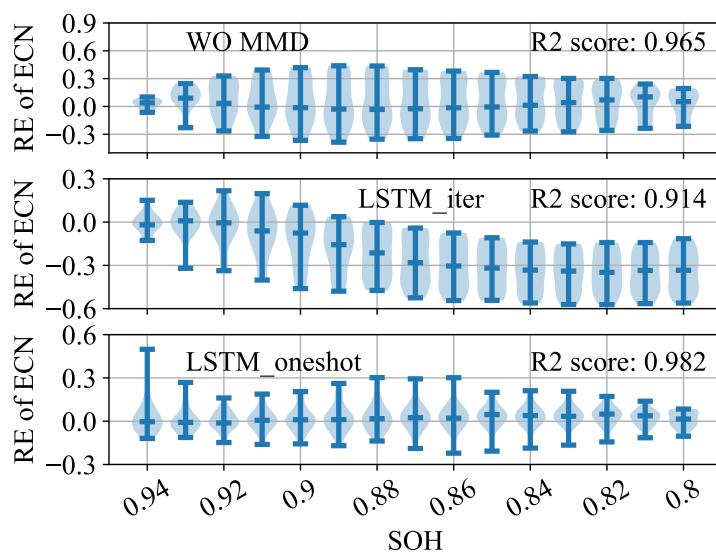
Table 2 summarizes the prediction results for batteries with short, middle and long lifetimes evaluated by RE in (15), where y_{pred} is the predicted value and y_{true} is the measured ECN in aging cycle sequence. The prediction of the aging cycle sequence is derived from the iterative model with MMD adaption, and the single-shot model with and without MMD adaption. The aging cycle sequences are transposed to SOH fading curves and shown in Figure 10. The single-shot model with MMD adaption shows better generalization to batteries from different domains with relative deviation between 8.4% and 10.4% than the one without MMD adaption with relative deviation ranging from 15.3% to 19.3%. For the model using iterative prediction, an obvious deviation up to 29.7% can be observed due to the accumulating error.

$$RE = \frac{|y_{pred,i} - y_{true,i}|}{y_{true,i}} \cdot 100\% \quad (15)$$

The R^2 and RE of RUL algorithms in other works using data-driven algorithms are summarized in Table 3, including LSTM, Transformer and Graph neural network [28–30,38,39]. The proposed single-shot framework takes fewer historical data points and shows good generalization and accuracy on a larger datasets with huge lifespan diversity, with an R^2 of 0.982 and RE within 10.8%.

Table 2. RUL prediction results for target dataset

Model	Actual RUL	Predicted RUL	RE (%)
Single-shot with MMD	87	94	8.4
	1629	1748	7.3
	2964	3272	10.4
Iterative with MMD	87	100	16.5
	1629	1176	-27.7
	2964	3844	29.7
Single-shot without MMD	87	100	15.3
	1629	1944	19.3
	2964	3433	15.8

**Figure 11.** SOH fading prediction error distribution for all testing cells.

The LSTM models with FT for TL are used in [28–30], which freeze partial networks of models pre-trained on the source dataset and further train the remaining networks on the target dataset. Performance in the target dataset becomes optimized with the increase in FT data in the target domain. In [28], the authors take the aging data from BOL to 90% SOH and fine-tune the pre-trained LSTM model on battery samples with an RUL of 443 and 318 ECN and obtain accurate RUL prediction with an RE of 7.38%. The average deviation ranges from 140 to 24 ECN when the FT data increase from 10% to 90% for a battery dataset with total lifespan between 400 and 700 ECN, indicating the RE decrease from 20% to 3.5% with increasing data for FT. In [29], the authors use four battery samples from the NASA dataset and take the capacity sequence of the first 15 ECN as input for RUL prediction, which is related to over 20% of the total life span. The TL is still based on FT and the RE of prediction on the target battery ranges from 9.0% to 12.7%. In [30], the authors apply FT to LSTM with attention mechanism and realize RUL prediction with RE between 4.66% and 14.89% using aging history with more than 10% SOH loss. In [38], a Transformer is used for RUL prediction to decrease the calculation time through parallel calculation. The results show a small difference when LSTM and Transformer are used on the NASA dataset due to the short total lifespan within around 200 ECN, while an obvious optimization can be observed on the CALCE dataset with up to 800 ECN. The Transformer shows that much benefit since the total length of the ECN sequence used in this work is 20 and the calculation time is within 20 ms. The Graph network in [39] takes input with aging data with 5% to 10% SOH loss and provides an R^2 between 0.92 to 0.982. In comparison to the reference

works, the proposed model in this work obtains similar accuracy with mean deviation of 150 ECN in Figure 10d and 11% RE in Figure 11 using 5% degradation data and a larger aging dataset with lifespan varying from 50 to 4000 ECNs.

Table 3. Performance comparison with other methods.

Model Type	Dataset	Prediction Start Point	R ² /Mean RE (%)
LSTM-FT [28]	/	≤90% SOH	RE: 7.38%
LSTM-FT [29]	NASA [25]	≥20% total lifespan	RE: 9% to 12.7%
LSTM-FT with attention [30]	/	<89% SOH	RE: 4.66% to 14.89%
Transformer [38]	NASA [25] and CALCE [26]	16 to 64 ECN (Aprox. 90% SOH)	RE: 7.6% to 22.5%
Graph Neural Network [39]	NASA [25]	90% to 95% SOH	R ² : 0.92 to 0.982

4. Conclusions

An encoder–decoder-based model is proposed in this work for RUL prediction. MMD loss is included in the feature presentation of autoencoder framework during training to optimize the generalization of RUL model for batteries with different lifespans. The main conclusions can be summarized as follows:

- (1) The incremental aging cycle sequence is used as the input and output. Therefore, the Euclidean distance metric can be used to compare the similarity of degradation curves. According to the clustering results using DBSCAN and k-means algorithms, the Euclidean metric-based clustering using an incremental aging cycle sequence shows more homogenies of aging patterns within the same cluster than the DTW metric-based case using capacity fading curves.
- (2) The MMD is used as the extra loss term of encoder feature presentation for TL. The model with MMD loss shows better generalization to batteries with different lifespans with an R² score of 0.982 than the one without MMD loss. Additionally, more densely distributed feature presentation can be observed in the MDS reconstruction figure for the model using MMD during training, indicating more homogeneous feature extraction and better generalization.
- (3) The proposed single-shot framework model 5% aging data and shows RUL prediction with an RE between 7% and 11% for batteries with different life cycles between 80 and 3000 ECN. In comparison to the single-shot model without MMD and the iterative prediction model, the absolute deviation reduces by 250 and 650 ECNs and the RE reduces by 19%, from 29.7% to 10.4%.

The future work may be further focused on the training strategies of feature representation learning to increase the generalization of RUL model to batteries with diverging lifespans. The adversarial training can be introduced in the encoder training to increase the homogeneity of the features extracted from source and target domains.

Author Contributions: Funding acquisition, S.P.; Methodology, W.L. and Y.Y.; Resources, S.P.; Software, W.L.; Validation, Y.Y.; Writing—review and editing, W.L. and Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Open-source dataset from KIT [27] is used.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

R ²	Coefficient of Determination
BOL	Beginning of Life
CC	Constant Current
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DOD	Depth of Discharge
DTW	Dynamic Time Wrapping
DWT	Discrete Wavelet Transform
ECN	Equivalent Cycle Number
EOL	End of Life
FT	Fine-Tuning
ICA	Incremental Capacity Analysis
LIB	Lithium-Ion Battery
LSTM	Long Short-Term Memory
MDS	Multi-Dimension Scaling
MMD	Maximum Mean Discrepancy
MSE	Mean Squared Error
NMC	Nickel Manganese Cobalt
RE	Relative Error
RUL	Remaining Useful Life
SEI	Solid Electrolyte Interface
SOC	State of Charge
SOH	State of Health
TL	Transfer Learning

References

1. Hannan, M.A.; Hoque, M.M.; Hussain, A.; Yusof, Y.; Ker, P.J. State-of-the-Art and Energy Management System of Lithium-Ion Batteries in Electric Vehicle Applications: Issues and Recommendations. *IEEE Access* **2018**, *6*, 19362–19378. [[CrossRef](#)]
2. Afshari, S.S.; Cui, S.; Xu, X.; Liang, X. Remaining Useful Life Early Prediction of Batteries Based on the Differential Voltage and Differential Capacity Curves. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 6500709. [[CrossRef](#)]
3. Joshi, B.; Samuel, E.; Kim, Y.I.; Lee, H.S.; Swihart, M.T.; Yoon, S.S. Exploring the potential of MIL-derived nanocomposites to enhance performance of lithium-ion batteries. *Chem. Eng. J.* **2023**, *461*, 141961. [[CrossRef](#)]
4. Joshi, B.; Samuel, E.; il Kim, Y.; Yarin, A.L.; Swihart, M.T.; Yoon, S.S. Progress and potential of electrospinning-derived substrate-free and binder-free lithium-ion battery electrodes. *Chem. Eng. J.* **2022**, *430*, 132876. [[CrossRef](#)]
5. Christensen, J.; Newman, J. Cyclable Lithium and Capacity Loss in Li-Ion Cells. *J. Electrochem. Soc.* **2005**, *152*, A818–A829. [[CrossRef](#)]
6. Birk, C.R.; Roberts, M.R.; McTurk, E.; Bruce, P.G.; Howey, D.A. Degradation diagnostics for lithium ion cells. *J. Power Sources* **2017**, *341*, 373–386. [[CrossRef](#)]
7. Vetter, J.; Novák, P.; Wagner, M.R.; Veit, C.; Möller, K.C.; Besenhard, J.O.; Winter, M.; Wohlfahrt-Mehrens, M.; Vogler, C.; Hammouche, A. Ageing mechanisms in lithium-ion batteries. *J. Power Sources* **2005**, *147*, 269–281. [[CrossRef](#)]
8. Wang, Z.; Zhao, Q.; Wang, S.; Song, Y.; Shi, B.; He, J. Aging and post-aging thermal safety of lithium-ion batteries under complex operating conditions: A comprehensive review. *J. Power Sources* **2024**, *623*, 235453. [[CrossRef](#)]
9. Pinson, M.B.; Bazant, M.Z. Theory of SEI Formation in Rechargeable Batteries: Capacity Fade, Accelerated Aging and Lifetime Prediction. *J. Electrochem. Soc.* **2012**, *160*, A243–A250. [[CrossRef](#)]
10. Yang, H.; Li, X.; Fu, K.; Shang, W.; Sun, K.; Yang, Z.; Hu, G.; Tan, P. Behavioral description of lithium-ion batteries by multiphysics modeling. *DeCarbon* **2024**, *6*, 100076. [[CrossRef](#)]
11. Edge, J.S.; O’Kane, S.; Prosser, R.; Kirkaldy, N.D.; Patel, A.N.; Hales, A.; Ghosh, A.; Ai, W.; Chen, J.; Yang, J.; et al. Lithium ion battery degradation: what you need to know. *Phys. Chem. Chem. Phys.* **2021**, *23*, 8200–8221. [[CrossRef](#)]
12. Prada, E.; Domenico, D.D.; Creff, Y.; Bernard, J.; Sauvant-Moynot, V.; Huet, F. A Simplified Electrochemical and Thermal Aging Model of LiFePO₄-Graphite Li-ion Batteries: Power and Capacity Fade Simulations. *J. Electrochem. Soc.* **2013**, *160*, A616–A628. [[CrossRef](#)]

13. Petit, M.; Prada, E.; Sauvant-Moynot, V. Development of an empirical aging model for Li-ion batteries and application to assess the impact of Vehicle-to-Grid strategies on battery lifetime. *Appl. Energy* **2016**, *172*, 398–407. [CrossRef]
14. Fioriti, D.; Scarpelli, C.; Pellegrino, L.; Lutzemberger, G.; Micolano, E.; Salamone, S. Battery lifetime of electric vehicles by novel rainflow-counting algorithm with temperature and C-rate dynamics: Effects of fast charging, user habits, vehicle-to-grid and climate zones. *J. Energy Storage* **2023**, *59*, 106458. [CrossRef]
15. Ali, M.A.; Da Silva, C.M.; Amon, C.H. Multiscale Modelling Methodologies of Lithium-Ion Battery Aging: A Review of Most Recent Developments. *Battat* **2023**, *9*, 434. [CrossRef]
16. Xu, W.; Cao, H.; Lin, X.; Shu, F.; Du, J.; Wang, J.; Tang, J. Data-Driven Semi-Empirical Model Approximation Method for Capacity Degradation of Retired Lithium-Ion Battery Considering SOC Range. *Appl. Sci.* **2023**, *13*, 11943. [CrossRef]
17. Li, W.; Sengupta, N.; Dechent, P.; Howey, D.; Annaswamy, A.; Sauer, D.U. One-shot battery degradation trajectory prediction with deep learning. *J. Power Sources* **2021**, *506*, 230024. [CrossRef]
18. Guo, X.; Yang, Z.; Liu, Y.; Fang, Z.; Wei, Z. A Hybrid Approach Based on Gaussian Process Regression and LSTM for Remaining Useful Life Prediction of Lithium-ion Batteries. In Proceedings of the 2023 IEEE Transportation Electrification Conference & Expo (ITEC), Detroit, MI, USA, 21–23 June 2023; pp. 1–4. [CrossRef]
19. Richardson, R.R.; Osborne, M.A.; Howey, D.A. Gaussian process regression for forecasting battery state of health. *J. Power Sources* **2017**, *357*, 209–219. [CrossRef]
20. Lin, Y.H.; Tian, L.L.; Ding, Z.Q. Ensemble Remaining Useful Life Prediction for Lithium-Ion Batteries With the Fusion of Historical and Real-Time Degradation Data. *IEEE Trans. Veh. Technol.* **2023**, *72*, 5934–5947. [CrossRef]
21. Li, Z.; Li, A.; Bai, F.; Zuo, H.; Zhang, Y. Remaining useful life prediction of lithium battery based on ACNN-Mogrifier LSTM-MMD. *Meas. Sci. Technol.* **2023**, *35*, 016101. [CrossRef]
22. Su, C.; Chen, H.J. A review on prognostics approaches for remaining useful life of lithium-ion battery. *IOP Conf. Ser. Earth Environ. Sci.* **2017**, *93*, 012040. [CrossRef]
23. Hu, X.; Xu, L.; Lin, X.; Pecht, M. Battery Lifetime Prognostics. *Joule* **2020**, *4*, 310–346. [CrossRef]
24. Mou, J.; Yang, Q.; Tang, Y.; Liu, Y.; Li, J.; Yu, C. Prediction of the Remaining Useful Life of Lithium-Ion Batteries Based on the 1D CNN-BLSTM Neural Network. *Battat* **2024**, *10*, 152. [CrossRef]
25. Saha, B.; Goebel, K. *Battery Data Set*; NASA Prognostics Data Repository, NASA Ames Research Center: Moffett Field, CA, USA, 2007.
26. Xing, Y.; Ma, E.W.; Tsui, K.L.; Pecht, M. An ensemble model for predicting the remaining useful performance of lithium-ion batteries. *Microelectron. Reliab.* **2013**, *53*, 811–820. [CrossRef]
27. Luh, M.; Blank, T. Comprehensive battery aging dataset: capacity and impedance fade measurements of a lithium-ion NMC/C-SiO cell. *Sci. Data* **2024**, *11*, 1004. [CrossRef]
28. Che, Y.; Deng, Z.; Lin, X.; Hu, L.; Hu, X. Predictive Battery Health Management With Transfer Learning and Online Model Correction. *IEEE Trans. Veh. Technol.* **2021**, *70*, 1269–1277. [CrossRef]
29. Chen, X.; Liu, Z.; Sheng, H.; Wu, K.; Mi, J.; Li, Q. Transfer learning based remaining useful life prediction of lithium-ion battery considering capacity regeneration phenomenon. *J. Energy Storage* **2024**, *76*, 109798. [CrossRef]
30. Chou, J.H.; Wang, F.K.; Lo, S.C. A Novel Fine-Tuning Model Based on Transfer Learning for Future Capacity Prediction of Lithium-Ion Batteries. *Batteries* **2023**, *9*, 325. [CrossRef]
31. Du, J.; Zhang, C.; Li, S.; Zhang, L.; Zhang, W. Two-stage prediction method for capacity aging trajectories of lithium-ion batteries based on Siamese-convolutional neural network. *Energy* **2024**, *295*, 130947. [CrossRef]
32. Severson, K.A.; Attia, P.M.; Jin, N.; Perkins, N.; Jiang, B.; Yang, Z.; Chen, M.H.; Aykol, M.; Herring, P.K.; Fraggedakis, D.; et al. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy* **2019**, *4*, 383–391. [CrossRef]
33. Huang, Y.; Zhang, P.; Lu, J.; Xiong, R.; Cai, Z. A transferable long-term lithium-ion battery aging trajectory prediction model considering internal resistance and capacity regeneration phenomenon. *Appl. Energy* **2024**, *360*, 122825. [CrossRef]
34. Meng, H.; Geng, M.; Xing, J.; Zio, E. A hybrid method for prognostics of lithium-ion batteries capacity considering regeneration phenomena. *Energy* **2022**, *261*, 125278. [CrossRef]
35. Bodnár, D.; Mouli, G.R.C.; Ďurovský, F.; Bauer, P.; Qin, Z. Semi-Empirical Model of Nickel Manganese Cobalt (NMC) Lithium-Ion Batteries Including Capacity Regeneration Phenomenon. *IEEE Trans. Transport. Electricif.* **2024**, *11*, 797. [CrossRef]
36. Zhang, S.; Wu, S.; Cao, G.; Chen, S.; Wang, Z.; Wang, N. Aging trajectory and end-of-life prediction for lithium-ion battery via similar fragment extraction of capacity degradation curves. *J. Clean. Prod.* **2024**, *436*, 140686. [CrossRef]
37. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining(KDD-96), Portland, OR, USA, 2–4 August 1996; pp. 226–231.

38. Chen, D.; Hong, W.; Zhou, X. Transformer Network for Remaining Useful Life Prediction of Lithium-Ion Batteries. *IEEE Access* **2022**, *10*, 19621–19628. [[CrossRef](#)]
39. Wei, Y.; Wu, D. State of health and remaining useful life prediction of lithium-ion batteries with conditional graph convolutional network. *Expert Syst. Appl.* **2024**, *238*, 122041. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.