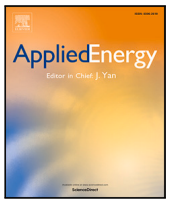
Applied Energy 377 (2025) 124629

Contents lists available at ScienceDirect

Applied Energy

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



A synthetic data generation method and evolutionary transformer model for degradation trajectory prediction in lithium-ion batteries

Haiyan Jin [a], Rui Ru [a], Lei Cai [a,*], Jinhao Meng [b], Bin Wang [a], Jichang Peng [c], Shengxiang Yang [d]

[a] School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, 710048, China
[b] School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, 710049, China
[c] Smart Grid Research Institute, Nanjing Institute of Technology, Nanjing 211167, China
[d] Institute of Artificial Intelligence, School of Computer Science and Informations, De Montfort University, Leicester, LE1 9BH, UK

ARTICLE INFO

Keywords:
Lithium-ion battery
Degradation trajectory prediction
Transformer model
Evolutionary framework

ABSTRACT

Identifying the long-term degradation of lithium-ion batteries in their early usage phase is crucial for the battery management system (BMS) to properly maintain the battery for practical use. Nevertheless, this procedure is challenging due to variations in the production and operating conditions of the battery. In recent years, it has been empirically proven that the data-driven method is a promising solution for handling the prediction of degradation. However, the lack of appropriate data remains the main obstacle that impacts the ultimate performance of the prediction. Furthermore, the prediction is also influenced by the setup of the predictor, which covers the structure of neural networks and their hyperparameters. The challenge of automating this process remains unresolved. In this study, we propose a novel degradation trajectory prediction framework. First, synthetic data is generated via a conditional generative adversarial network (CGAN), providing the characterization of the battery's degradation at an early stage and utilizing the argument data to alleviate the issue of insufficient data. Second, an evaluation method to evaluate the quality of the synthetic data is also provided. In addition, a selection method is proposed based on the diversity mechanism to further filter out the redundancy of synthetic data. These two sub-processes aim to promote the quality of the synthetic data. Finally, the synthetic data hybrid with real values is used for the training of a transformer model, whose architecture and hyper-parameters are automatically configured via an evolutionary framework. The experimental results show that the proposed method can achieve accurate predictions compared to its rivals, and its best configuration can be automatically configured without hand-crafted efforts.

1. Introduction

Lithium-ion batteries are increasingly popular in several fields due to their numerous benefits, making them suitable for a wide range of applications. Nevertheless, the limited duration of the battery hinders the development of its applications [1]. Excessive and frequent use will definitely decrease the battery lifespan. Failure to appropriately replace the battery might result in significant safety concerns [2,3]. Consequently, the maintenance and operation of the Lithium-ion battery is crucial for its applications, particularly in the field of battery energy storage for the electric cars and smart grids [4,5]. By accurately predicting the degradation of capacities during the early cycling phase, we are able to strategically arrange the schedule of battery replacements to prevent safety concerns and optimize the economic efficiency of the secondary use of the batteries [6].

The lithium-ion battery is a complex system with various factors affecting its lifespan, and these are often coupled with each other. This makes the ageing trajectories of degradation show a significant non-linear pattern. The Battery Management System (BMS) faces challenges in identifying material factors like lithium plating, increased resistance, and electrode saturation that contribute to the battery capacity decline [7]. As a result, accurately estimating the State of Health (SOH) and predicting their future ageing trajectories is tricky due to their dynamic and complex nature [8]. The use of the smart battery safety monitoring technique has resulted in an increase in the quantity and accuracy of the measured variables [9]. This advancement enables the prediction of battery degradation. Meanwhile, by utilizing the digital twin technique [10], the predicted ageing trajectories can

* Corresponding author.
E-mail address: cailei@xaut.edu.cn (L. Cai).

https://doi.org/10.1016/j.apenergy.2024.124629
Received 20 April 2024; Received in revised form 29 July 2024; Accepted 1 October 2024
Available online 22 October 2024

be transmitted to the battery management system. This not only enhances battery maintenance and safety but also provides a baseline for analyzing potential second-life applications from both technique and economic standpoint [11]. Recently, there have been numerous method try to make a prediction for the future of a battery, can be termed as health prognostic technique, such as the prediction of knee point [12], end of life (EOL) [13], remaining useful life (RUL) [14] or even the whole degradation trajectories [15]. These methods, according to their basic mechanism, can be broadly classified into three categories: model-based, data-driven, and hybrid methods.

The model-based methods, in ascending order of their complexity, are empirical model [16], equivalent circuit model (ECM) [17,18] and electrochemical model [19,20]. The empirical model is the most straightforward method which seeks to provide a mathematical formula that maps the number of cycles to the appropriate capacities. Some typical mathematical models, such as polynomial and exponential model are commonly used [21,22], and historical data is utilized to determine the appropriate parameters for these models. It is worth mention that relying solely on empirical models to predict capacities for long-term degradation is challenging since the battery degradation has a strong nonlinear behavior. Therefore, in practice, empirical models are frequently employed in conjunction with filter-based methods to dynamically update critical model parameters [23,24]. The primary goal of ECM is to simulate the entire working process of a battery through the use of a simplified circuit model comprised of a series of circuit elements. The typical works employed a fractional-order ECM to estimate model parameters via the input current and output terminal voltage of the battery. Furthermore, similar to the empirical model, the ECM commonly employs filter-based method to improve its prognostic capabilities over the long term [25,26]. The electrochemical model, as a type of mechanistic model, examines the internal electrochemical processes within the battery to explain its degradation patterns [27]. For example, as one of the representative models, the pseudo-2-dimensional (P2D) model employs certain equations and parameters that faithfully represent the battery's physical properties. These properties, such as solid-electrolyte interphase (SEI) resistance, electrolyte porosity can be considered as the key indicators for interring the future degradation trajectory of the battery [28].

The model-based methods do have some advantages. The empirical model is straightforward and easy to implement. The ECM model gives us the basic simulation of the working environment, some state of battery can be easily inferred by its state of circuit. The electrochemical model can output properties that have physical meaning, which can not only used for the estimation of the SOH and prognostic prediction but can also be used as the key feedback to the battery manufacturer to further improve the battery's quality. However, their disadvantages are also obvious. The model parameters of the empirical model depend on historical data, and its generalization tends to be weak when applied to the fresh batteries [26]. Although the computational efficiency of the ECM is still acceptable, achieving a trade-off between model complexity and prediction accuracy is challenging when the ECM is used in conjunction with a filter-based method. The computational complexity of electrochemistry models is high, which imposes significant limits when applied to BMS [27]. Furthermore, constructing a precise electrochemical model necessitates a wide range of knowledge and an in-depth understanding of the complex internal mechanics of the battery.

In contrast, data-driven methods necessitate considerably less understanding of the internal mechanisms of batteries. The majority of data-driven methods utilize historical data to construct a predictor that maps a set of explicit or implicit features to the battery lifetime. During this process, it is unnecessary to consider the mechanism of the battery's degradation. We follow the categorization method proposed in Ref. [29] to divide the data-driven prognostic into three distinct categories based on their input–output structure: single point prediction, iterative or rolling prediction, and sequence-to-sequence prediction. The first one uses features captured from charge or discharge signals to predict a single property related to the prognostic. As a representative work, Ref. [30] constructs statistical features of discharge capacities with only 100 cycles to accurately predict the remaining useful cycle. Besides that, we believe that the identification of the knee point, which reflects the critical points where the battery's degradation turns nonlinear, can also be categorized within this category [7,31], as the output of this prediction is also a single scale value. Rather than predicting a single point, the iterative prediction seeks to predict multiple steps or even the complete degradation trajectory in an iterative manner. For example, Xiaosong, etc., utilize a Long Short-term Memory (LSTM) neural network to iteratively predict the capacity one step ahead. The error caused by the non-linear degradation property was eliminated by combining a model-based method with a particle filter [23]. In Ref. [32], the LSTM combined with the Gaussian Progress Regression (GPR) model to recursively predict the future capacity and RUL. The iterative approach is an enhanced form of single-point prediction for degradation curve prediction. However, this method is computationally expensive since it requires several iterations to obtain the final degradation curves. Moreover, the iterative prediction may be at risk for inaccuracy as a result of the accumulation of errors and noise. The last category of data-driven methods is solving the prediction of degradation trajectory as a sequence-to-sequence (Seq2Seq) problem [33]. Instead of breaking down the predicting process into multiple iterations, Seq2Seq learning aims to optimize the entire system directly. Some neural networks, especially deep learning methods, have a strong ability for end-to-end learning, making the prediction of degradation trajectories more reliable. For example, Xu, etc. proposed a novel method to hybridize physics-based and data-driven methods for degradation trajectory prediction; the ultimate predictor is an autoencoder consisting of two LSTMs to directly map the early capacity curve to the remaining degradation trajectory [29]. Qian, etc., proposed an attention-based multi-source Seq2Seq method to consider both historical information and future SOH degradation trends to predict the long-term SOH [33]. In summary, the three kinds of data-driven methods all utilize historical data to construct a black-box predictor for the prognostic of the cells. Since the data-driven method can provide a greater generalization for the fresh cells, it has increasingly gained popularity in the research community.

Although significant progress has been made in the data-driven methods, difficulties are still encountered in capacity degradation trajectory prediction for lithium-ion batteries. The main challenge is insufficient data. Data-driven methods heavily rely on the data they use for training, especially the data that is labeled. However, there is only a small amount of data available since conducting such an ageing experiment on a battery is too slow and expensive; it is impractical and costly to do so. Thus, access to quality data can be considered as the bottleneck in the development of data-driven methods for the prediction of degradation trajectory. Recently, some efforts were made to try to alleviate this challenge. The first attempt to handle the problem of data insufficiency is to focus on the predictor. For example, some semi-supervised methods are employed to fully take advantage of the unlabeled data [34,35]. In general, these methods address the challenge of a lack of labeled data. However, certain unlabeled data can be directly used as important supplements in the training process. Nevertheless, the semi-supervised learning method is highly dependent on the assumption that the unlabeled data follows the same distribution as the labeled data and, therefore, faces difficulties when incorporating domain knowledge [36]. The second method to addressing the issue of insufficient data is to continue focusing on the predictor and define the prior distribution using transfer learning technology and models trained on related tasks [37–39]. While it is proven that transferring knowledge from the source domain can partly alleviate the data shortage issue, bridging the gap between the source and target domains remains to be an enormous task. Instead of focusing the model on handling the challenge of data shortage, expanding the training dataset using empirical or generative models would be a promising solution. For

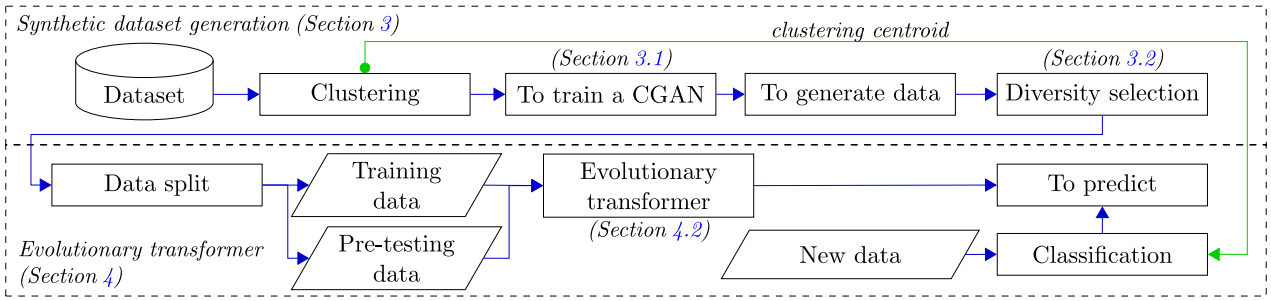


Fig. 1. The main procedure of the proposed degradation trajectory prediction method.

example, In Ref. [40], the synthetic data were initialized via polynomial function and used for training the CNN model. In addition, Qiu, etc., utilize a conditional generative adversarial network to enrich training data for state-of-charge (SOC) estimation [41]. Given that using data augmentation can directly address the challenge of data shortages, it shows huge potential to resolve the data insufficiency problem.

Data augmentation appears to be a direct solution to the issue of insufficient data, but it is not a silver bullet. There are still some obstacles to overcome when attempting to predict degradation using synthetic data. The first problem is how to generate synthetic data that is more realistic in nature. At present, evaluation methods for the generated data are absent. Most generative methods exist that evaluate their works via visual quality, which is directly assessed by human judgment. In practice, however, it is impracticable to manually evaluate every single synthetic data point. The second issue is to ensure the diversity of synthetic data. Given the inherent randomness in most data generation techniques, it is crucial to boost the diversity of the synthetic data in order to improve the predictor's ability to generalize. The third issue mainly comes from the configuration of the predictor, particularly in the case of deep learning models. In practice, a more flexible solution is required to configure the essential architecture and hyper-parameters in order to ensure the performance of the predictor. However, this process is typically carried out in a trial-and-error manner, which is impractical and costly to do so.

Confronting the challenges mentioned above, this paper introduces a novel evolutionary transformer model. This model, which leverages a synthetic data generation technique, is designed to predict the long-term degradation of lithium-ion batteries. The unique aspects of this model and its potential impact on the field are the primary focus of this paper.

1. To produce synthetic data, a Conditional Generative Adversarial Network (CGAN) is employed. Our goal is not to simply replicate an existing generative model. Instead, CGAN is specifically customized by collaborating with a clustering procedure. The clustering centroid can serve as the condition input for the CGAN.
2. Based on the basic principle of generative data, which is to get as close as possible to the realistic data, we design an evaluation method for assessing the quality of generated data. In addition, a selective procedure is proposed based on the diversity mechanism to filter out the redundancy of synthetic data.
3. A novel transformer-CNN model is proposed to promote the prediction performance. In order to enhance the smoothness of the prediction results, a transformer model is expanded by several 1D convolution layers. Furthermore, we propose an evolutionary method to automate the optimization of architecture and hyper-parameters. The configuration of the proposed Transformer-CNN can be automatically determined without any manual intervention.

The paper is outlined as follows. First, the main procedure of the proposed method, as presented in Section 2, provides a snapshot of the method. In Section 3, the technique of generating synthetic data is elaborated, and corresponding evaluation and diversity selection methods are also presented. Section 4 gives the details of the proposed evolutionary transformer-CNN model. The experiment setup, which refers to the datasets, performance metrics, and peer method, is given in Section 5. The experimental results and corresponding discussion are given in Section 6.

2. The main procedure of the proposed method

As depicted in Fig. 1, the whole procedure of the proposed degradation trajectory early prediction method consists of two parts. The first main part is to generate synthetic degradation data. There are four fundamental stages, including clustering, training a generator, generating data, and diversity selection. Once the reasonable synthetic dataset is prepared, it can be used to construct a transformer model to predict the early degradation trajectories. The training procedure for the evolutionary transformer is depicted in the lower part of Fig. 1.

Next, the generation of synthetic datasets will be described in detail, followed by the training of an evolutionary transformer and the corresponding prediction method.

3. Method for generating battery capacity degradation curves based on CGAN

Despite the fact that data-driven methods have acquired popularity due to their high accuracy, they typically require a large amount of data for model training, and the fundamental requirement of these data is that they must be diverse enough to ensure the generalization of the training model. In practical applications, the collection of battery degradation data is influenced by various factors such as battery operating conditions, loads, and data collection methods. Acquiring various battery degradation data for the purpose of research is extremely costly, requiring a substantial investment of time and money. Thus, understanding how to acquire diverse datasets is extremely valuable for research purposes. This paper uses data generation techniques to augment existing datasets, increasing the sample size and diversity of battery degradation data in order to obtain more extensive and diverse battery degradation data under various application scenarios. The detailed generalization approach will be described as follows.

3.1. Method for generating battery capacity degradation curves

In this work, we utilize CGAN to enhance the existing Lithium-ion battery degradation trajectory data. CGAN is a variant of the Generative Adversarial Network (GAN), consisting of a generator and a discriminator. As depicted in Fig. 2, the generator converts input random noise and conditional vector into some false samples that are similar

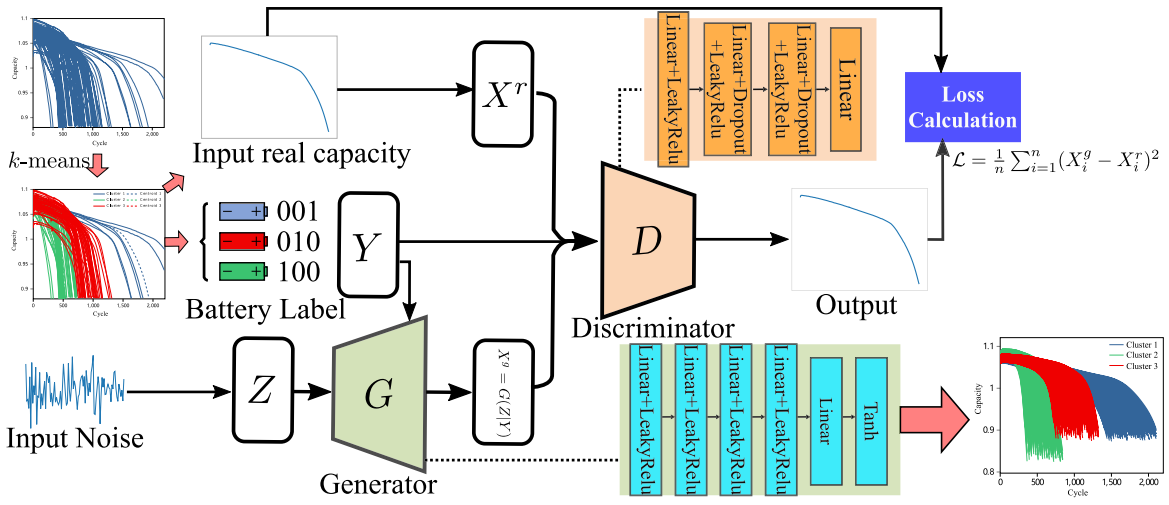


Fig. 2. The conditional-GAN for data synthesis.

to the real data. Meanwhile, the discriminator is used to distinguish between real data and fake data generated by the generator. The generator gradually improves the quality of fake samples by learning the characteristics of real data. In CGAN, the generator and discriminator are jointly trained, and their objective functions are mutually adversarial. The generator's objective is to generate false data that is similar to real data, while the discriminator's objective is to correctly identify the fake data generated by the generator and real data. In the application of CGAN to Lithium-ion battery degradation trajectory, the generator takes noise vector and conditional vector as input to generate fake degradation trajectory data that meets the conditions, and the training process makes the output of the generator closer to the real data. The discriminator is used to compare the differences between the generated fake data and the real data, helping the generator improve the authenticity of the generated fake data. Both the generator and the discriminator learn and optimize together to eventually generate a set of fake data samples with higher authenticity, improving the existing Lithium-ion battery degradation trajectory data.

The initial step in data synthesis is the preparation of training data. Initially, we use the k-means algorithm to cluster the present degradation curves into three categories: short, medium, and long lifespan [42]. For each category, one-hot code is used as a label to represent the actual capacity degradation curve, three categories is represented as 001, 010, and 100, respectively. After that, the degradation curve closet to the clustering centroid is considered as the representative curve for the current category, and its corresponding one hot label are selected as input for the data synthesis.

In the training phase, multiple copies of the representative curve serve as the training dataset for a CGAN model. The structure of generator and discriminator are elaborated in Fig. 2. In particular, the generator accepts a noise vector and conditional labels as inputs and generates synthetic data that should resemble actual data from the desired distribution. The proposed generator is composed of four BatchNorm 1D layers and five linear layers. The discriminator, on the other hand, accepts both actual data samples with their labels and fake data samples generated by the generator with their labels. Its purpose is to differentiate between real and fake data. The current discriminator consists of four linear layers and two Dropout layers.

Once the CGAN model's training is complete, it can be utilized for data synthesis. Firstly, the corresponding labels for the synthesized curves need to be selected and converted into the respective conditional vectors, ensuring that the format of the conditional vectors remains consistent with the CGAN model training process. Next, the prepared conditional vectors and random noise are inputted into the trained generator network. The synthesized battery capacity degradation trajectory curves can be obtained through the output of the generator network. Finally, the reasonableness and reliability of the synthesized curves can be evaluated using the algorithm designed in this paper to ensure that the synthesized curves meet the requirements.

3.2. The selection of the synthetic data based on the maximum diversity

From the above subsection, we can see that the CGAN has the capability to produce synthetic data that is closely related to the original values. The data obtained, however, exhibit a degree of stochasticity, raising a challenge for the proposed prediction. Thus, for the synthetic data, we still need a selection operation to further filter out the low-quality synthetic data, which is redundant to each other. The primary concept underlying this selection operation is to iteratively select synthetic data that exhibits a relatively high level of variation in comparison to the other data within the candidate pool.

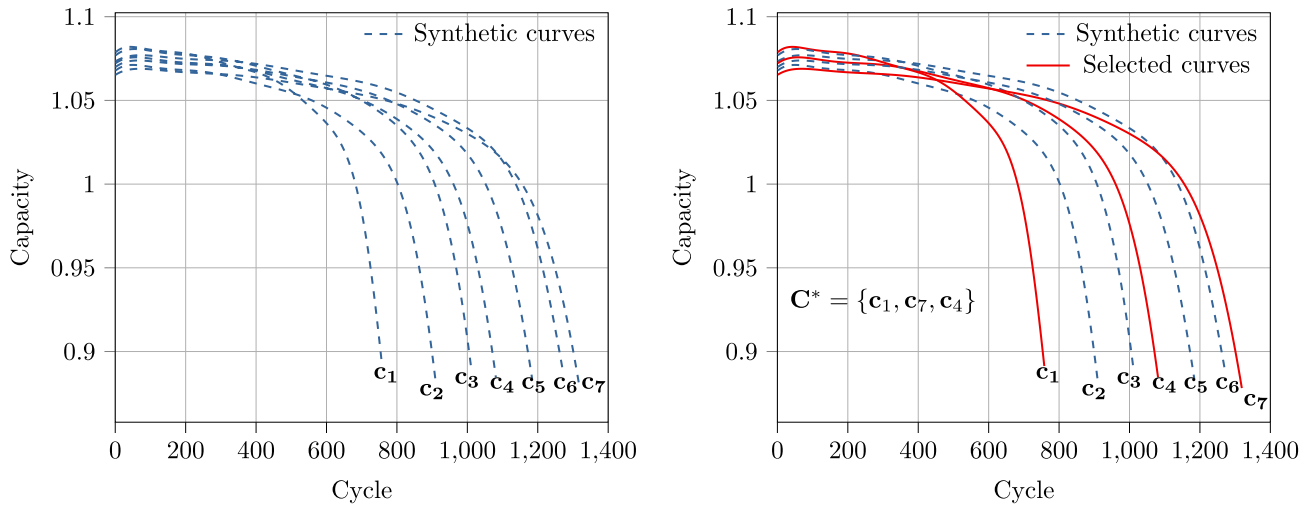
To select synthetic data with high diversity, we first need to define the similarity between two degradation curves, which could qualify the degree to which the two curves follow a similar trajectory as they progress along their respective axes. Here, we utilize dynamic time warping (DTW) to calculate the similarity between two curves [43]. The DTW was initially developed to quantify the similarity between two temporal sequences, even in the case of differences in their temporal progress. It is capable of computing the distance between two time series of varying lengths. Different from Euclidean Distance, which can only measure similarity between points of identical length, DTW can handle situations where two sequences do not align at the same time steps. This situation also frequently occurs in the generated curves, making comparing two curves more challenging. Consequently, given the present conditions, DTW is more suited for comparing two curves in here.

In addition, the distance of a single curve to a curve set is defined as the smallest one among the distances between this curve and all the curves in the set, which can be represented as follows,

$$d(\mathbf{c}_i, \mathbf{C}) = d(\mathbf{c}_i, \mathbf{c}_p) \tag{1}$$

where $\mathbf{c}_p = \underset{\mathbf{c}_j \in \mathbf{C}}{\operatorname{argmin}}\, d(\mathbf{c}_i, \mathbf{c}_j)$.

Once the method of measuring similarity is determined, we can design a selection procedure to find the synthetic data that have the maximum diversity. The proposed method can be demonstrated through Algorithm 1. To be more specific, we utilize the trained optimal generator to generate a batch of capacity degradation curves and define that as the $\mathbf{C}$. First, to randomly select a curve $\mathbf{c}_\mathbf{r}$ from the generated set $\mathbf{C}$. We need to find a curve $\mathbf{c} \in \mathbf{C}$ that has the maximum distance to



(a) The initial synthetic curves (b) The selected curves according to the porposed method

Fig. 3. Illustration of maximal diversity selection method.

Algorithm 1: Pseudo code of maximal diversity curve selection

Input: Related curve set $\mathbf{C}$ and number of curves to be selected n
Output: The selected curve $\mathbf{C}^*$

1 $\mathbf{C}^* \leftarrow \emptyset$, $k \leftarrow 1$;
2 Randomly select a curve $\mathbf{c}_r$ in $\mathbf{C}$, find a curve $\mathbf{c}$ in $\mathbf{C}$ with the largest distance to the $\mathbf{c}_r$, $\mathbf{C}^* \leftarrow \mathbf{C}^* \cup \{\mathbf{c}\}$, $\mathbf{C} \leftarrow \mathbf{C} \backslash \{\mathbf{c}\}$;
3 **while** $k \neq n$ **do**
4 Find a curve $\bar{\mathbf{c}} \in \mathbf{C}$ with the largest distance to $\mathbf{C}^*$;
5 $\mathbf{C}^* \leftarrow \mathbf{C}^* \cup \{\bar{\mathbf{c}}\}$, $\mathbf{C} \leftarrow \mathbf{C} \backslash \{\bar{\mathbf{c}}\}$;
6 $k \leftarrow k + 1$
7 **return** $\mathbf{C}^*$

$\mathbf{c_r}$, and to move the $\mathbf{c}$ from $\mathbf{C}$ to $\mathbf{C}^*$ (line 2). After that, the algorithm enters the second stage (lines 3–6). During each iteration at this stage, we try to find the generated curve $\bar{\mathbf{c}} \in \mathbf{C}$ which has the largest distance to $\mathbf{C}^*$. Finally, the loop is terminated when n curves are selected.

In order to better understand the proposed maximal diversity selection method, Fig. 3 gives an example of how we select a limited number of curves from the synthetic dataset. Assume that we have seven synthetic degradation curves, denoted as $\mathbf{C} = \{\mathbf{c_1} \dots \mathbf{c}_7\}$. Our objective is to identify a subset of three degradation curves that represent the overall degradation trend for the current cluster. As outlined in Algorithm 1, the initial step involves the random selection of a curve $\mathbf{c}_r$. Assume that the curve $\mathbf{c}_6$ has been selected as the one. The $\mathbf{c}_1$ has the highest DTW distance to $\mathbf{c}_6$. Consequently, we designate $\mathbf{C}^*$ as the set containing $\mathbf{c_1}$, and proceed to remove the $\mathbf{c}_1$ from $\mathbf{C}^*$. Next, the method proceeds to execute the loop operations, where $\mathbf{c}_7$ is identified as having the largest distance to $\mathbf{C}^*$ and is subsequently included in $\mathbf{C}^*$. Ultimately, $\mathbf{c}_4$ has the greatest distance to $\mathbf{C}^* = \{\mathbf{c}_1, \mathbf{c}_7\}$, and is therefore regarded as the final selection for the construction of $\mathbf{C}^*$.

The aforementioned description reveals that the objective of the proposed method is to maximize the diversity of the data that is generated. The potential advantages of this process can be summarized as two aspects. Firs is efficiency: instead of using all synthetic data as the training set, a small amount of synthetic data with maximal diversity can reduce scale with minimal informative loss. The second benefit is promoting prediction generalization. The presence of a wide range of different data enables the surprised learning model to be exposed to a greater verity of information, which in turn enhances the model's ability to generalize.

3.3. The evaluation of the synthetic data

After generating synthetic data, it is imperative to assess the quality of the synthetic data. One fundamental principle of a generation model is its ability to synthesize data that closely matches the realistic data. Furthermore, it is important that the training data does not include the testing data in order to prevent the model from simply memorizing the original data. Taking these principals into consideration, we design a straightforward yet efficient evaluation method for assessing the quality of the synthetic data. The basic idea behind this method is to use the real data to evaluate the synthetic dataset, since the data that used for evaluation never involved for the training of CGAN. By getting the synthetic data closer to the real data, the CGAN is able to produce novel data that matches the unexperienced data, as opposed to solely relying on the memorization of the original training data.

Algorithm 2: Pseudo code of evaluation of the synthetic data

Input: Training Data $\mathbf{Y}$, Testing Data $\mathbf{Y}'$, $(\mathbf{Y} \cap \mathbf{Y}' = \varnothing)$
Output: Similarity Score S and corresponding index I

1 {discriminator, generator} ← `train_CGAN`($\mathbf{Y}$, L);
2 **foreach** $l \in |\mathbf{Y}'|$ **do**
3 $v \leftarrow$`generator.generate_synthetic`(l) ;
4 $\{s, i\} \leftarrow$ `find_maximum_similarity`(v, $\mathbf{Y}'$);
5 $S \leftarrow \{s\}$, $I \leftarrow \{i\}$;
6 **return** S, I

The procedure of the evaluation is described in Algorithm 2. First the training and testing data are considered as the input for the evaluation method. Prior to that, it is essential to ensure that the testing data does not appear in the training set, i.e. $\mathbf{Y} \cap \mathbf{Y}' = \varnothing$. After training the CGAN, the generator is utilized to generate a synthetic data, denoted as v (line 1). Subsequently, we deliberately search for the data in the $\mathbf{Y}'$ that has the highest similarity to v, and corresponding index of the realistic dataset is recorded (lines 2–5). DTW is still utilized in here to handle the situation that pairs of data has different cycles. Finally, the data in the testing set that are closet to the synthetic one and

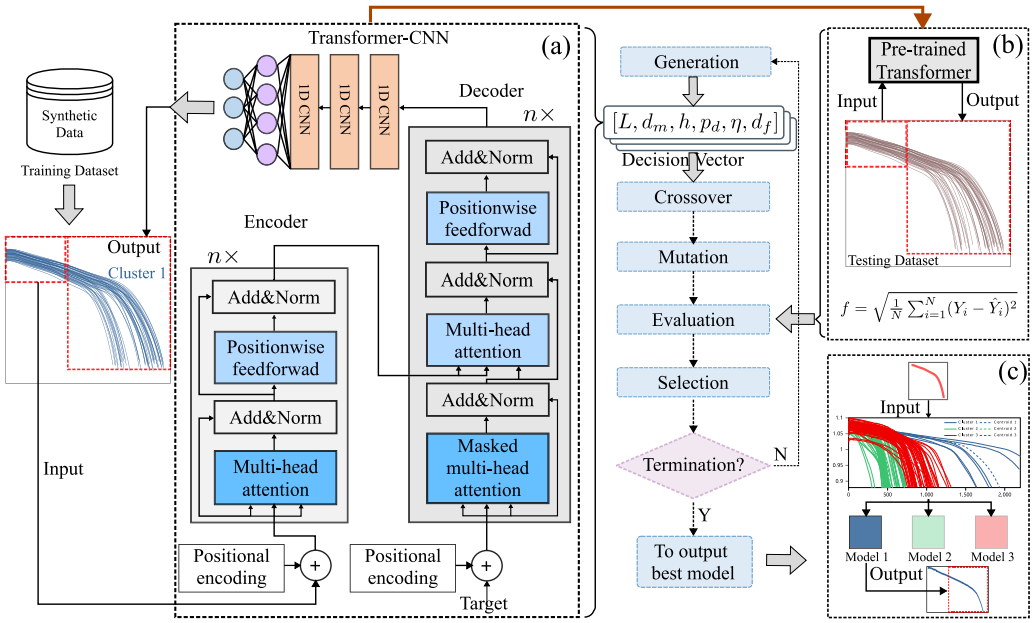


Fig. 4. Overall structure of the evolutionary transformer-CNN model.

their corresponding index are considered as the output for the latter qualitatively evaluation (lines 6). We can utilize this output to calculate the similarity score between the synthetic and realistic data.

4. Early battery degradation trajectory prediction with transformer model

Once the synthetic curves depicting battery degradation are generated, they can serve as an instance of reference for predicting the trajectory of early battery degradation. In this work, the Transformer-CNN model is utilized for the prediction of capacity degradation. Furthermore, the optimization of the main hyperparameters and structure of the encoder and decoder in the transformer model, which ultimately decides the prediction's final performance, is accomplished by the use of an evolutionary strategy.

4.1. Transformer model

The transformer model is a neural network design that has gained significant prominence and has been widely employed in several domains and applications [44]. In contrast to conventional recurrent neural networks, the transformer model employs self-attention processes for the purpose of processing sequential input. Typically, the model has two main components: the encoder and the decoder. The encoder is responsible for extracting features from the input sequence, while the decoder generates the target sequence based on the encoder's output and self-attention mechanism. The self-attention mechanism enables the model to weigh the importance of different tokens in sequential data dynamically.

One of the key strengths of the transformer model is its ability to handle sequential data with remarkable effectiveness. Unlike its predecessors, such as RNN and LSTM [45], the transformer does not suffer from the vanishing gradient problem, making it capable of capturing long-range dependencies. Multi-head attention, another critical component, allows the model to attend to different aspects of the input sequence concurrently, enhancing its ability to learn complex patterns and relationships.

For the training phase, the original degradation and its synthetic degradation curves are fed into the transformer model. The input for the transformer model is 100 cycles of the degradation curves, and the output is the rest of the cycles. It is worth mentioning that despite the input of the proposed model being the capacity time series until the present time point (100 cycles for the current circumstance), it still relies on battery sensors and a BMS to obtain the measurable variables. Current capacities can be determined by evaluating the measured currents and voltages, as stated in Ref. [46]. Consequently, battery sensors play a crucial role as a data source in the proposed model. By using the digital twin technique [10], the measured variables can be sent to the proposed method for supervised training. The updated model can then be transmitted to the BMS to accurately predict degradation trajectories.

4.2. The training process of evolutionary transformer-CNN model

After the synthetic data process, we have three different groups of synthetic data that have maximum diversity. Next, we need to learn these data according to the proposed evolutionary transformer-CNN model. In order to promote the prediction's performance, we separately trained three transformer models for each cluster. Fig. 4 illustrates the overall structure of the transformer model for one specific cluster.

First, for the transformer model itself, as depicted in Fig. 4(a), the architecture of the transformer model is slightly different from the traditional one. Instead of only following a linear transformation, we added several 1D convolution layers behind the decoder. The reason for this modification is that it can significantly improve the smoothness of the output curve.

For the training process, the dataset is first divided into training and pre-testing datasets. The training dataset is used to train a specific transformer model. The input data for the model is the first 100 cycles of the degradation trajectories, and the output of the model is the rest of the cycles. Each potential transformer model's ultimate predictive performance is determined by both architecture and hyper-parameters. For the architecture of the transformer model, the key variation is the number of encoders and decoders, which can be represented as L. The choice of the number of layers (encoders and decoders) in a transformer model can impact both the model's capacity and its ability to learn from data effectively. For the hyper-parameters of the model, there are five parameters that can determine the final performance of the training.

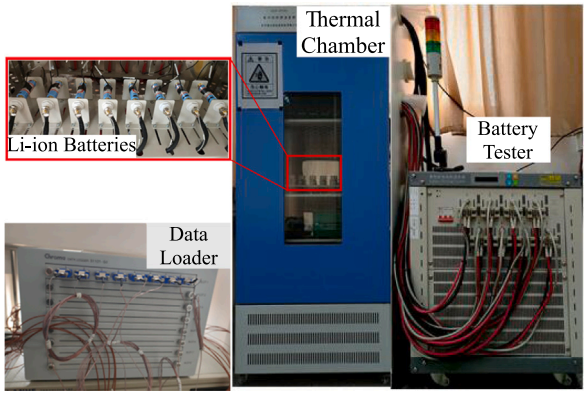


Fig. 5. The battery test bed for our private dataset.

Table 1
The boundary of each decision value.

Decision value	L	d_m	h	p_d	η	d_f
Boundary	[1, 10]	[1, 128]	[1, 8]	[0, 1]	[0.0001, 0.001]	[1, 128]

The first one is the positional encoding dimension d_m, which determines the dimensionality of the positional encodings added to input embeddings. The second parameter is the number of attention heads h, which specifies how many attention mechanisms are used in multi-head attention. Turning the number of heads can adjust the model's ability to capture different types of relationships in the data. The third parameter is the dropout rate p_d to control the probability of dropping out neurons during training, helping models to prevent overfitting. Next, learning rate η is a key parameter that determines the rate at which a model's weights are updated during training. The last one is the forward dimension d_f, which determines the dimensionality of the forward neural network within each layer. Adjusting this dimension can affect the models' capacity and computational efforts.

In practice, determining the architecture and hyper-parameters of the transformer model is cumbersome. We intend to delegate this process to an evolutionary optimization process. As depicted in the middle flow of Fig. 4, each decision vector $[L, d_m, h, p_d, \eta, d_f]$ in the evolutionary process can represent a specific transformer model. For the decision vector, the boundary for each decision value is presented in Table 1.

After the generation steps, there is a population that contains different kinds of transformer models, and each one needs to be trained by the training dataset. After some basic evolutionary operations such as crossover and mutation, the population has doubled in size. In the evaluation step, each pre-trained transformer model is evaluated via the pre-testing dataset, as depicted in Fig. 4(b). Within this step, we first need to define an objective function that guides the evolutionary process. Considering the purpose of optimization is to find the optimal model that has the best performance, the root mean square error (RMSE) between the predicted value Y and the ground truth $\hat{Y}$ is considered as the ultimate objective function, which can be represented as follows.

$$f = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2} \tag{2}$$

The entire evolutionary process repeats itself until the termination condition is met. The output of the evolutionary process is an optimal solution capable of representing the greatest performance for degradation trajectory prediction. Once three models representing the best estimate for the low, medium, and long battery lifetimes have been developed, they can be incorporated into a final model to predict future degradation trajectories. To be more specific, as depicted in Fig. 4(c), once a degradation with early cycles is obtained, the centroid of the clustering procedure described in Section 3.1 is used as a reference. DTW is still used in here to compare the similarity between the early degradation trajectories and the centroid to determine which category of lifetime the current degradation trajectory belongs to. Thus, the clustering centroid is considered a representation of the specific category. When a target category is determined, its pre-trained transformer model will be utilized to predict the rest of the degradation trajectory.

5. Experiment setup

5.1. Battery datasets

In this work, three datasets are used to validate the proposed method, and corresponding details are summarized as follows.

- Maryland dataset [47]: The first dataset utilized in this study is from the Center for Advanced Life Cycle Engineering (CALCE) at the University of Maryland. Three lithium cobalt oxide ($LiCoO_2$) batteries with a capacity of 1.1 Ah have been selected for this study. These batteries are specifically identified as CS_35, CS_36, and CS_37 and undergo 0.5C constant current constant voltage (CC-CV) charge and 1C CC discharge, whose voltage ranges from 2.7 V to 4.2 V.
- MIT-Stanford dataset [30]: The dataset has been made accessible by academics affiliated with MIT and Stanford University. The collection comprises 124 commercially available lithium iron phosphate/graphite batteries, and the nominal capacity is 1.1 Ah. The deterioration patterns of the dataset in this ageing experiment has a broad range, spanning from 400 to 1200 cycles, due to the use of various fast-charging protocols. Thus, the use of the MIT-Stanford dataset for validating the effectiveness of the capacity degradation trajectory prediction presents a greater challenging because various fast-charging policies employed.
- On-road Vehicles dataset [48]: To validate the effectiveness of the proposed trajectory prediction method in real-world conditions, we utilize the charging data of battery packs for commercial electric vehicles (EV). The total time of EV operating exceeds two years. The data is collected using charging devices, which receive battery charging data via control area network communication during the charging process.
- Our private dataset: Four 1.5 Ah 18650 NMC-based Li-ion batteries are charged with the CC-CV model and discharged with constant current on our battery test bench, which is shown in Fig. 5. The voltage of the Li-ion battery is limited between 2.5 V to 4.2 V, and the discharging current is 15 A for all the cells. As shown in Fig. 5, we also include a thermostat to keep the temperature at 25 °C.

It is worth noting that the number of batteries accessible for Maryland and our private dataset is limited, so the clustering function is disabled for these two datasets. To be more specific, all the batteries can be classified into a single category for these two datasets. The clustering procedure is no longer required, and just a handful of these batteries are used for training.

5.2. Performance metrics

In this paper, two performance metrics are considered, which can be represented as follows.

- Mean absolute error (MAE): It is essentially to find the absolute difference between each predicted value and its corresponding ground truth.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|Q_{pred,i} - \hat{Q}_{ref,i}| \tag{3}$$

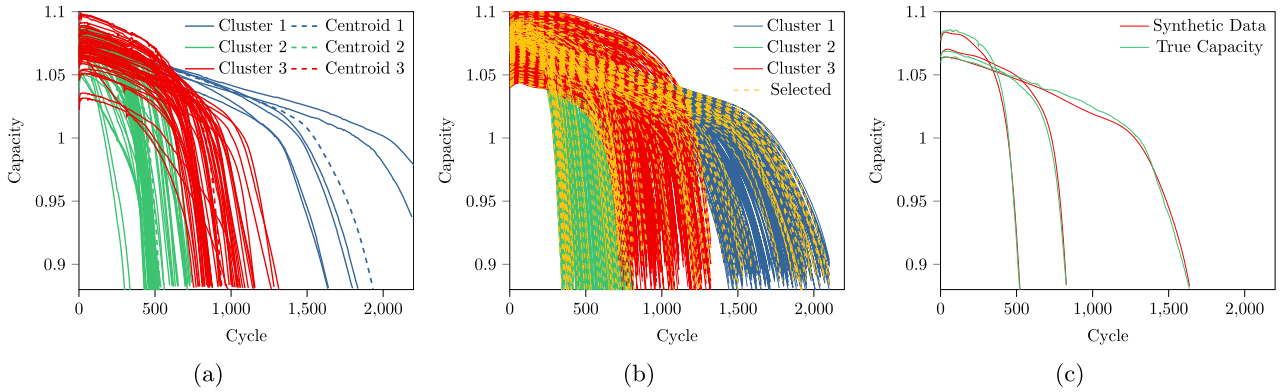


Fig. 6. The results of synthetic dataset generation. (a) The original MIT-Stanford dataset and corresponding clustering results. (b) The results of synthetic dataset and the selected values by the proposed diversity selection method. (c) The synthetic degradation trajectories that best matches the true value.

- Root mean square error (RMSE): It quantifies the average magnitude of errors between predicted values and actual observed values. Compared with MAE, it is more sensitive to outliers or extreme values in the data.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Q_{pred,i} - \hat{Q}_{ref,i})^2} \tag{4}$$

where N is the total number of test sets, Q_{pred} is the predicted capacity, and Q_{ref} is the true capacity.

5.3. Peer methods

The selected peer methods versus our proposed `EVO-TransformerCNN` are outlined as follows. These three methods have covered all three categories of the techniques mentioned in Ref. [29].

- LSTM based Sequence-to-sequence model (`Seq2Seq-LSTM`) [29]: The LSTM based Sequence-to-sequence consists of an encoder and a decoder. The encoders obtain the input data and encode it into a latent vector, and then the decoder is responsible for decoding the data to build the final output data. The present study utilizes the initial 100 cycles of the degradation trajectory as input and the remaining cycles of the trajectory as output. It is worth mentioning that the physical-based features in work [29] are not used here for a fair comparison.
- Gaussian Process Regression (GPR): We consider GPR as a competitor since it is frequently used for the predicting of the ageing of the batteries [49,50]. The initial 100 cycles of the capacities are deemed as the input for training a GPR in a recursive manner. Subsequently, the future capacities are predicted iteratively in a sliding-window manner: at each step of the iteration, a singular capacity is predicted based on its preceding l capacities, where l serves as the representation of the window size.
- Naive transformer model (`Nav-Transformer`): This is the transitional transformer model for sequence-to-sequence prediction without following CNN layers. In addition, neither the structures nor the hyperparameters are intentionally not optimized. The primary reason for this particular method being considered a competitive method is its ability to directly assess the sensitivity of prediction performance to configuration and the attended CNN layers.
- Transformer CNN model without optimization (`TransformerCNN -1&2`): These models are the exact replica of the proposed prediction model, with the exception that no architecture or hyper-parameter optimization operations were performed. In other words, the decision values related to the architecture and hyper-parameter are randomly assigned in order to verify the necessity of the optimization.

6. Experiment results and discussion

6.1. The results of the synthetic dataset generation

In this subsection, we investigate the effectiveness of the proposed synthetic dataset generation method as introduced in Sections 3.1 to 3.3, respectively. The generated synthetic dataset will be presented intuitively and be further evaluated by using the method introduced in Section 3.3. Fig. 6 shows the proposed generation method by conducting experiments on the original MIT-Stanford dataset.

As can be seen from Fig. 6, the original dataset can be clustered into three different types. The centroids of these clusters are also displayed in Fig. 6(a). The synthetic dataset generated by using the proposed CGAN is presented in Fig. 6(b). It is evident from these results that the proposed method is capable of producing a variety of synthetic data for each clustering. In the meantime, the dashed yellow lines that represent the selected degradation trajectories provide evidence that the diversity selection method that was proposed is capable of identifying representative instances from the original synthetic dataset. Fig. 6(c) illustrates a comparison between the synthetic data and the real capacity discovered through the implementation of the evaluation method suggested in Section 3.3. As the training of CGAN does not include these true degradation trajectories, it can be inferred that the proposed method is capable of generating synthetic data that matches the real data.

Table 2 provides a comprehensive evaluation of the synthetic trajectories. We employ Algorithm 2 to choose a synthetic set of data that has the most similarities when compared to the realistic data. The error between realistic and synthetic data is then calculated. The ultimate metrics for each dataset were established by measuring the mean errors. Based on the findings presented in Table 2, it is evident that there is a small difference between the real and synthetic data. The ability of the proposed generative method to provide reliable data is confirmed.

6.2. The results of early battery degradation trajectory prediction

In this subsection, the proposed method is verified on three different kinds of batteries. First, for our private dataset, a set of eight cells is selected for the training process, and the other four cells is considered as the testing set to evaluate the performance. It is assumed that 100 cycles will be given, and the remaining cycles need to be predicted. Before training an estimator, these four cells are directly considered as four types to generate four types of synthetic degradation trajectories. Ultimately, the first 100 cycles of the training data and synthetic data are both chosen as the input, and the remaining cycles are considered as the output for the training process. It is worth mentioning that due to the limited size of the dataset, the clustering procedure is disabled

Table 2
The RMSE and MAE for the synthetic data.

Maryland dataset		MIT dataset		The private dataset		Vehicle dataset	
RMSE (%)	MAE (%)	RMSE (%)	MAE (%)	RMSE (%)	MAE (%)	RMSE (%)	MAE (%)
1.68	1.33	0.69	0.47	3.95	2.31	1.32	1.04

Table 3
The RMSE and MAE of the trajectory prediction for our private dataset.

Method	Cell-4		Cell-5		Cell-7		Cell-8	
	RMSE (%)	MAE (%)	RMSE (%)	MAE (%)	RMSE (%)	MAE (%)	RMSE (%)	MAE (%)
Seq2Seq-LSTM	18.42	13.34	2.38	1.87	11.32	8.14	3.45	2.85
GPR	32.51	16.91	2.55	1.50	20.50	10.73	5.39	3.53
Nav-Transformer	6.82	5.32	1.33	1.54	4.46	3.54	**1.91**	**1.51**
TransformerCNN-1	11.97	10.19	2.33	1.96	6.85	5.65	4.89	3.74
TransformerCNN-2	15.11	12.47	1.67	1.47	4.82	4.06	2.06	1.62
EVO-TransformerCNN	**5.62**	**3.42**	**1.08**	**0.76**	**3.72**	**2.26**	1.92	1.56

* The best results are highlighted in bold face, with a gray background.

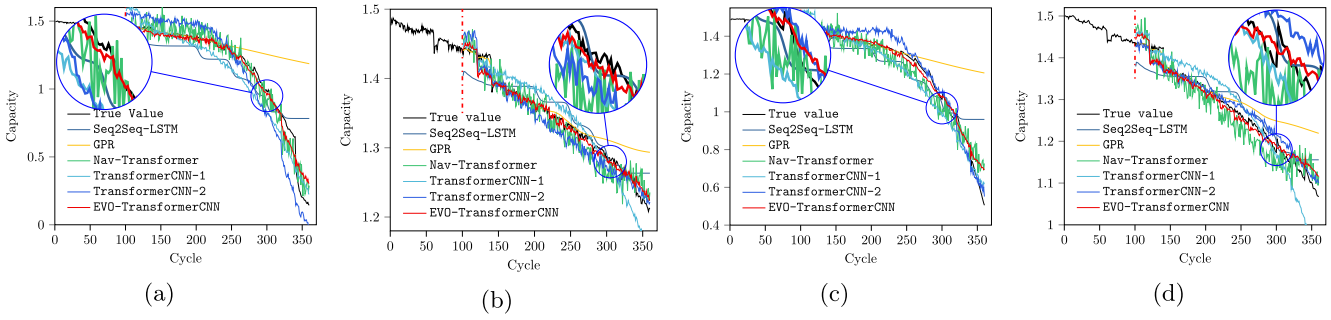


Fig. 7. The results of degradation trajectory prediction for our private dataset. (a)–(d) demonstrate the prediction results for Cell 4, 5, 7 and 8, respectively.

for both our private dataset and the Maryland dataset. It means that the assumption that there is just one category is recognized for the synthetic method. In addition, the proposed EVO-TransformerCNN and all its peers utilized the original and synthetic data as the training set for a fair comparison.

Our private dataset exhibits a notable decline in the initial degradation trajectories for certain batteries, which poses a greater problem for accurately predicting degradation. Fig. 7 plots the predicted trajectories for the four methods on testing cells, thereby providing an intuitive indication of the prediction's performance. It is evident from these results that the proposed EVO-TransformerCNN, which is the one closest to the true value, has obtained the best performance for our private dataset. Despite the naive transformer model, i.e., Nav-Transformer, is also near to the true value, its unstable effect is obvious. The inferior performance of Seq2Seq-LSTM provides evidence that the LSTM model may have limited capability in managing long-term sequences. It is clear that the GPR can successfully capture the degradation trend of the batteries. However, its prediction accuracy is rather low due to the sensitivity of the sliding window method to noise and unexpected drops in the degradation curve. Finally, from the prediction results of the TransformerCNN-1 and TransformerCNN-2, it is easily observed its predicted trajectories are smoother than the trajectories predicted by Nav-Transformer. However, in comparison to the EVO-TransformerCNN, its accuracy is rather low. Based on these findings, we may infer that although the proposed TransformerCNN has the capability to predict smooth trajectories, its architecture and hyper-parameters still require careful tuning in order to fully utilize its capabilities.

The results for specific performance metrics and overall methods on our private dataset are summarized in Table 3, which provides an objective indicator for all methods. The proposed method is always among the top-performing methods except for Cell-8. It is interesting to note that the Nav-Transformer is very competitive from the observation of the objective indicator. However, by examining the predicted degradation trajectories, it is clear that the prediction results are highly unstable. The primary reason could be that noise could easily degrade the performance of the naive transformer despite the fact that it is capable of handling the long-term sequence problem. It is interesting to note that TransformerCNN-1 and TransformerCNN-2 almost have equivalent performances for our private dataset. TransformerCNN-1 obtains a good performance for Cell-4, whereas TransformerCNN-2 performs better on other cells. This indicates that the optimization depends on the problem of the specific battery to be predicted. The ability to obtain a global optimal solution for the evolutionary optimization is well-suited for the present situation.

Fig. 8 plots the predicted degradation trajectories on the testing cells for the Maryland dataset, and Table 4 gives the corresponding performance metrics. The proposed method still gives the highest accuracy on two cells, but Seq2Seq-LSTM instead of Nav-Transformer obtains the next-best performance. In contrast to our private dataset, the Maryland dataset comprises a larger amount of noise, which significantly degrades the performance of the Nav-Transformer. Consequently, we can conclude that in conjunction with the synthetic dataset, the transformer model does perform well in the prediction of degradation trajectories, but its denoising ability is relatively weak. The proposed EVO-TransformerCNN has some redundancy CNN layers

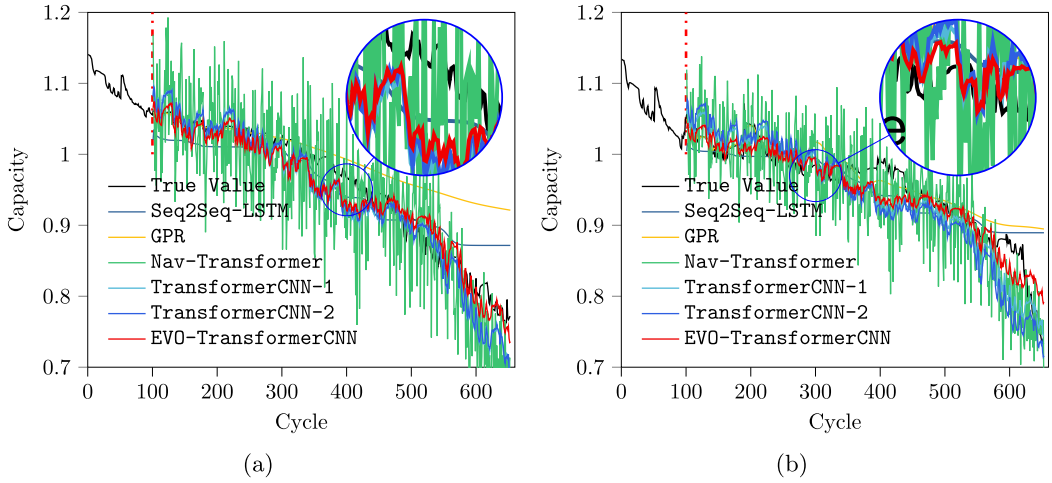


Fig. 8. The results of degradation trajectory prediction for the Maryland dataset. (a)–(b) demonstrate the prediction results for CS2-35 and CS2-37, respectively.

Table 4
The RMSE and MAE of the trajectory prediction in Maryland dataset.

Method	CS2-35		CS2-37	
	RMSE (%)	MAE (%)	RMSE (%)	MAE (%)
Seq2Seq-LSTM	3.72	3.01	3.49	2.34
GPR	6.60	5.33	3.47	2.06
Nav-Transformer	6.62	5.31	5.42	4.31
TransformerCNN-1	3.01	2.41	3.26	2.57
TransformerCNN-2	2.96	2.32	4.02	3.19
EVO-TransformerCNN	**2.32**	**1.87**	**2.45**	**1.96**

* The best results are highlighted in bold face, with a gray background.

followed by the transformer model, which can effectively suppress the effects of noise. In addition to the carefully adjusted architecture and hyperparameters for the transformer model, the proposed EVO-TransformerCNN shows a relatively high performance on the prediction. Similar to the results of our private dataset, the GPR with sliding-window manner still can only follow the manifold of basic degradation, but there is a significant deviation at the knee point of the degradation trajectories. For the TransformerCNN model without optimization, i.e., TransformerCNN-1 and TransformerCNN-2, they are found to achieve a good performance compared with other peer methods. Consequently, we can conclude that the proposed TransformerCNN is capable of handling the prediction for the battery and has a stable degradation with noise, even without some deliberative configuration optimization.

Fig. 9 plots the predicted degradation trajectories of the testing cells in the on-road vehicle dataset. The main characteristic of this dataset is its origin from practical applications; it exhibits a higher degree of error and data noise, which poses a greater difficulty in accurately predicting. Intuitively, EVO-TransformerCNN can achieve the best performance since its predicted degradation trajectory is most closely aligned with the ground truth. As one of the sliding-window methods, GPR is sensitive to the noise. Although it produces relatively smooth predicted trajectories, its accuracy is still weak. Conversely, the Seq-2-Seq method is resistant to the effects of noise and gains a good performance in most instances. It is important to note that properly adjusting the architecture and hyper-parameters is vital in order to effectively handle noisy situations. Without that, neither TransformerCNN-1 nor TransformerCNN-2 can attain satisfactory performance.

Table 5 gives the detailed performance metrics for the on-road vehicle dataset. From that, we can qualitatively compare the performance of the different methods. From this comparison results, we can clearly see the superior performance of the EVO-TransformerCNN. Occasionally, Seq2Seq-LSTM can perform best among other methods due to its ability to handle the noise in the practical application.

Fig. 10 illustrates the heatmap of RMSE and MAE for the degradation trajectory prediction for the MIT dataset. These heat maps are capable of representing data in a matrix where colors are represented. It can be effective in displaying the spatial distribution of the performance, indicating the stability of the proposed method. The greater the intensity of the color, the higher the performance level. It is clear that the performance of the proposed method is stable on most predictions of degradation trajectories; only a few instances encountered difficulties, such as the third one in cluster one. The reason for this phenomenon in clustering one is the long degradation trajectory, which has limited data available. The restricted data availability hampers the prediction performance. Based on the empirical results, we can infer that the proposed method consistently achieves a stable performance. Next, we will evaluate the performance of the proposed method in a comparative manner.

Figs. 11 and 12 plot the boxplot for the results of prediction on MAE and RMSE, respectively. These figures provide a concise summary of the statistically sound results, including the highest and lowest error values, as well as the average and median values. From the comparison results, we can clearly see that the proposed method obtains the best accuracy in the median values of MAE and RMSE. In addition, the box of the proposed method is relatively narrower than its rivals in most cases, indicating that its performance would be more stable than other methods. It is noted that the blue box of Nav-Transformer is more narrow than the proposed method in the case of cluster 3. This occurrence may be attributed to the over-fitting resulting from the optimization process. This phenomenon occurs occasionally but can

Table 5
The RMSE and MAE of the trajectory prediction for on-road vehicle dataset.

Method	Cell-2		Cell-4		Cell-6		Cell-8	
	RMSE (%)	MAE (%)	RMSE (%)	MAE (%)	RMSE (%)	MAE (%)	RMSE (%)	MAE (%)
Seq2Seq-LSTM	8.95	6.39	7.33	6.38	8.34	**6.72**	8.69	7.04
GPR	17.33	14.83	10.03	8.54	11.37	9.54	7.61	6.48
Nav-Transformer	7.73	6.14	8.17	6.83	10.06	8.71	8.73	7.36
TransformerCNN-1	7.01	5.48	6.95	6.04	8.85	7.47	7.65	6.32
TransformerCNN-2	6.74	5.26	7.47	6.21	8.63	7.46	7.43	6.14
EVO-TransformerCNN	**6.45**	**5.15**	**6.82**	**5.86**	**8.32**	7.04	**7.31**	**5.96**

* The best results are highlighted in bold face, with a gray background.

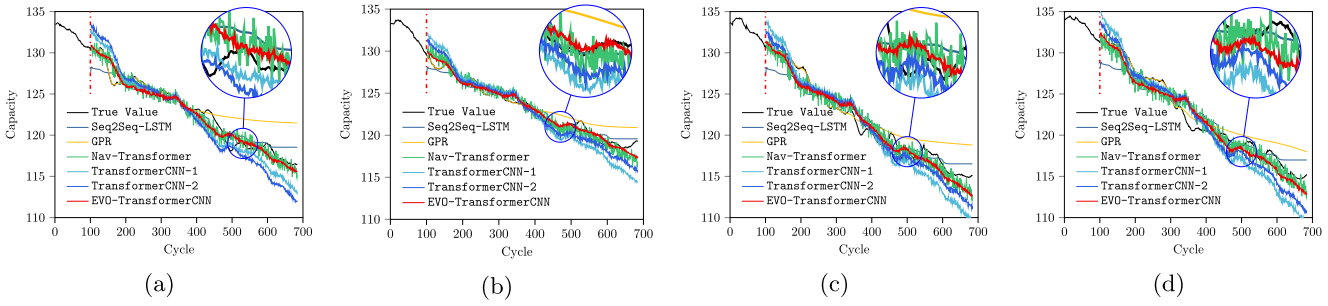


Fig. 9. The results of degradation trajectory prediction on-road vehicle dataset. (a)–(d) demonstrate the prediction results for Cell 2, 4, 6 and 8, respectively.

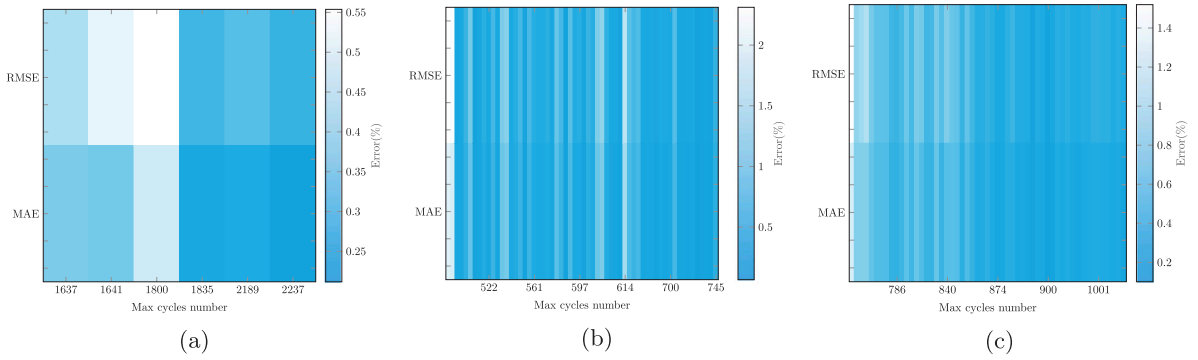


Fig. 10. The prediction results of RMSE and MAE for the degradation trajectory prediction. (a)–(c) represents the prediction results for cluster 1 to cluster 3, respectively.

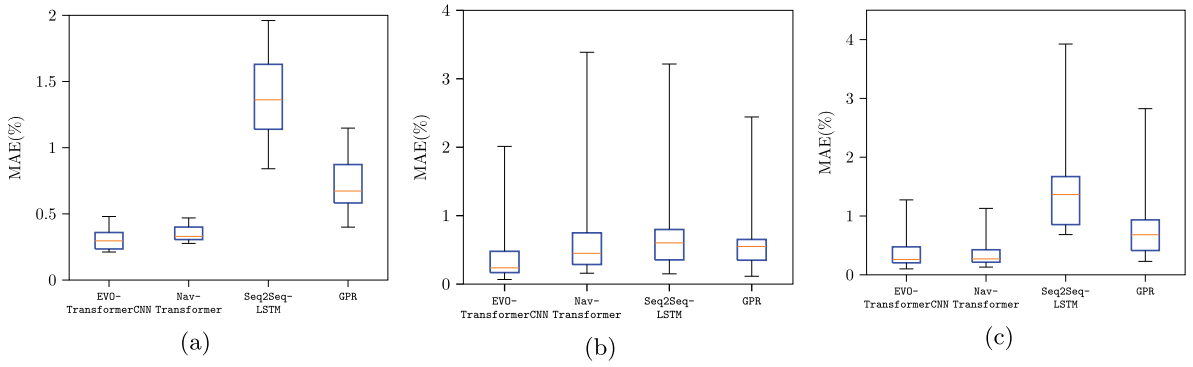


Fig. 11. A comparison of the prediction error on MAE for three methods. (a)–(c) represents the prediction results for cluster 1 to cluster 3, respectively.

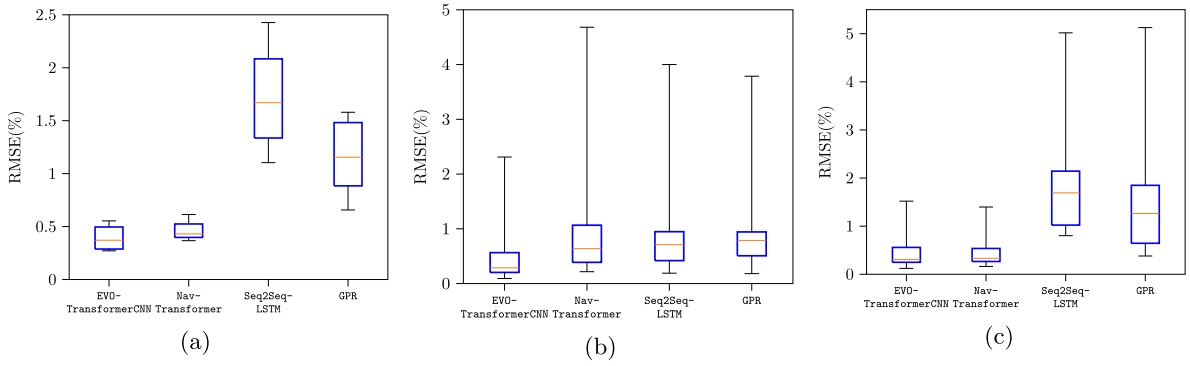


Fig. 12. A comparison of the prediction error on RMSE for three methods. (a)–(c) represents the prediction results for cluster 1 to cluster 3, respectively.

be easily remedied by implementing specific procedures. For example, in some cases, some suboptimal solutions rather than optimal solutions are selected to construct the predictor and enhance the ability of generalization. Furthermore, it is evident that every method has exhibited relatively weak results in the second clustering. More precisely, the upper and lower boundaries of the boxplot are wider in comparison to the boundaries of other clusters. One possible explanation is that the test dataset in cluster 2 is rather large, which makes predicting the degradation trajectory more challenging compared to other clusters. Nevertheless, by examining the median value, lower quartile value, and upper quartile value, it is evident that the proposed `Nav-Transformer` continues to deliver exceptional performance.

7. Conclusion

This research presents a novel evolutionary transformer-CNN model for accurately predicting batteries' early degradation. Initially, we propose a method for generating battery capacity degradation curves by employing CGAN. The entire process of data synthesis is controlled by the corresponding one-hot code, which serves as the identifier for the specific clustering. In order to alleviate the issue of duplicative curves, a selection mechanism is proposed to eliminate redundant synthetic data that have similarities. We furthermore present an evaluation methodology for evaluating the quality of the generated data. The use of synthetic data enables the prediction of the degradation of capacities with limited capacities in the early phase. A predictor, termed evolutionary transformer-CNN model, where the hyperparameters and key architecture of the CNN part are optimized by the evolutionary process. The validation results on a large dataset have demonstrated the superior efficacy of the proposed method, achieving an MAE and RMSE of less than 3% for most cases. In addition, the synthetic dataset generation results were given, offering a better understanding of data dependency for the data-driven method.

One major future work is to investigate the proposed method in more batteries with different characteristics further. In addition, since the data-driven system strongly relies on the quantity and quality of the data used, acquiring the batteries' data was too slow and costly. How to generate synthetic data that is closer to the realistic one can be an interesting direction to explore in the future.

CRediT authorship contribution statement

Haiyan Jin: Methodology. **Rui Ru:** Writing – original draft, Formal analysis, Data curation. **Lei Cai:** Conceptualization. **Jinhao Meng:** Visualization, Validation. **Bin Wang:** Software, Resources. **Jichang Peng:** Writing – original draft, Resources. **Shengxiang Yang:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is financially supported by the National Natural Science Foundation of China under Grant 62372369, 62272383 and 52107229, The Key Research and Development Program of Shaanxi Province under Grant 2024GX-YBXM-442.

Data availability

Data will be made available on request.

References

[1] von Bülow F, Meisen T. A review on methods for state of health forecasting of lithium-ion batteries applicable in real-world operational conditions. J Energy Storage 2023;57:105978.
[2] Edge JS, O'Kane S, Prosser R, Kirkaldy ND, Patel AN, Hales A, et al. Lithium ion battery degradation: What you need to know. Phys Chem Chem Phys 2021;23(14):8200–21.
[3] Xie J, Lu Y-C. A retrospective on lithium-ion batteries. Nature Commun 2020;11(1):2499.
[4] Landi M, Gross G. Measurement techniques for online battery state of health estimation in vehicle-to-grid applications. IEEE Trans Instrum Meas 2014;63(5):1224–34.
[5] Fairley P. Energy storage: Power revolution. Nature 2015;526(7575):S102–4.
[6] Che Y, Hu X, Lin X, Guo J, Teodorescu R. Health prognostics for lithium-ion batteries: Mechanisms, methods, and prospects. Energy Environ Sci 2023. http://dx.doi.org/10.1039/D2EE03019E.
[7] Attia PM, Bills A, Brosa Planella F, Dechent P, dos Reis G, Dubarry M, et al. Review—Knees in Lithium-ion battery aging trajectories. J Electrochem Soc 2022;169(6):060517.
[8] Ling C. A review of the recent progress in battery informatics. npj Comput Mater 2022;8(1):1–22.
[9] Xiong R, Sun X, Meng X, Shen W, Sun F. Advancing fault diagnosis in next-generation smart battery with multidimensional sensors. Appl Energy 2024;364:123202.
[10] Li W, Rentemeister M, Badeda J, Jöst D, Schulte D, Sauer DU. Digital twin for battery systems: Cloud battery management system with online state-of-charge and state-of-health estimation. J Energy Storage 2020;30:101557.
[11] Shahjalal M, Roy PK, Shams T, Fly A, Chowdhury JI, Ahmed MR, et al. A review on second-life of Li-ion batteries: Prospects, challenges, and issues. Energy 2022;241:122881.
[12] Liu K, Tang X, Teodorescu R, Gao F, Meng J. Future ageing trajectory prediction for lithium-ion battery considering the knee point effect. IEEE Trans Energy Convers 2022;37(2):1282–91.
[13] Gou B, Xu Y, Feng X. State-of-health estimation and remaining-useful-life prediction for lithium-ion battery using a hybrid data-driven method. IEEE Trans Veh Technol 2020;69(10):10854–67.

[14] Zhao B, Zhang W, Zhang Y, Zhang C, Zhang C, Zhang J. Research on the remaining useful life prediction method for lithium-ion batteries by fusion of feature engineering and deep learning. Appl Energy 2024;358:122325.
[15] Jia X, Zhang C, Li Y, Zou C, Wang LY, Cai X. Knee-point-conscious battery aging trajectory prediction based on physics-guided machine learning. IEEE Trans Transp Electr 2023;1.
[16] Saldaña G, Martín JIS, Zamora I, Asensio FJ, Oñederra O, González-Pérez M. Empirical calendar ageing model for electric vehicles and energy storage systems batteries. J Energy Storage 2022;55:105676.
[17] Zhao R, Kollmeyer PJ, Lorenz RD, Jahns TM. A compact methodology via a recurrent neural network for accurate equivalent circuit type modeling of lithium-ion batteries. IEEE Trans Ind Appl 2019;55(2):1922–31.
[18] Tian N, Wang Y, Chen J, Fang H. One-shot parameter identification of the Thevenin's model for batteries: Methods and validation. J Energy Storage 2020;29:101282.
[19] Liu K, Gao Y, Zhu C, Li K, Fei M, Peng C, et al. Electrochemical modeling and parameterization towards control-oriented management of lithium-ion batteries. Control Eng Pract 2022;124:105176.
[20] Cui B, Wang H, Li R, Xiang L, Zhao H, Xiao R, et al. Ultra-early prediction of lithium-ion battery performance using mechanism and data-driven fusion model. Appl Energy 2024;353:122080.
[21] Micea MV, Ungurean L, Cârstoiu GN, Groza V. Online state-of-health assessment for battery management systems. IEEE Trans Instrum Meas 2011;60(6):1997–2006.
[22] Ye L-H, Chen S-J, Shi Y-F, Peng D-H, Shi A-P. Remaining useful life prediction of lithium-ion battery based on chaotic particle swarm optimization and particle filter. Int J Electrochem Sci 2023;18(5):100122.
[23] Hu X, Yang X, Feng F, Liu K, Lin X. A particle filter and long short-term memory fusion technique for lithium-ion battery remaining useful life prediction. J Dyn Syst Meas Control 2021;143(6):061001.
[24] Chen D, Meng J, Huang H, Wu J, Liu P, Lu J, et al. An Empirical-Data Hybrid Driven Approach for Remaining Useful Life prediction of lithium-ion batteries considering capacity diving. Energy 2022;245:123222.
[25] Chen L, Bao X, Lopes AM, Xu C, Wu X, Kong H, et al. State of health estimation of lithium-ion batteries based on equivalent circuit model and data-driven method. J Energy Storage 2023;73:109195.
[26] Hu X, Xu L, Lin X, Pecht M. Battery lifetime prognostics. Joule 2020;4(2):310–46.
[27] Luo K, Chen X, Zheng H, Shi Z. A review of deep learning approach to predicting the state of health and state of charge of lithium-ion batteries. J Energy Chem 2022;74:159–73.
[28] Peng J, Meng J, Du X, Cai L, Stroe D-I. A fast impedance measurement method for lithium-ion battery using power spectrum property. IEEE Trans Ind Inf 2023;19(7):8253–61.
[29] Xu L, Deng Z, Xie Y, Lin X, Hu X. A novel hybrid physics-based and data-driven approach for degradation trajectory prediction in Li-ion batteries. IEEE Trans Transp Electr 2022;1.
[30] Severson KA, Attia PM, Jin N, Perkins N, Jiang B, Yang Z, et al. Data-driven prediction of battery cycle life before capacity degradation. Nat Energy 2019;4(5):383–91.
[31] Sohn S, Byun H-E, Lee JH. Two-stage deep learning for online prediction of knee-point in Li-ion battery capacity degradation. Appl Energy 2022;328:120204.
[32] Liu K, Shang Y, Ouyang Q, Widanage WD. A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery. IEEE Trans Ind Electron 2020;1.
[33] Qian C, Xu B, Xia Q, Ren Y, Sun B, Wang Z. SOH prediction for Lithium-Ion batteries by using historical state and future load information with an AM-seq2seq model. Appl Energy 2023;336:120793.
[34] Guo N, Chen S, Tao J, Liu Y, Wan J, Li X. Semi-supervised learning for explainable few-shot battery lifetime prediction. Joule 2024;S254243512400103X.
[35] Che Y, Stroe D-I, Hu X, Teodorescu R. Semi-supervised self-learning-based lifetime prediction for batteries. IEEE Trans Ind Inf 2022;1–10.
[36] van Engelen JE, Hoos HH. A survey on semi-supervised learning. Mach Learn 2020;109(2):373–440.
[37] Ma G, Xu S, Jiang B, Cheng C, Yang X, Shen Y, et al. Real-time personalized health status prediction of lithium-ion batteries using deep transfer learning. Energy Environ Sci 2022;15(10):4083–94.
[38] Liu K, Peng Q, Che Y, Zheng Y, Li K, Teodorescu R, et al. Transfer learning for battery smarter state estimation and ageing prognostics: Recent progress, challenges, and prospects. Adv Appl Energy 2023;9:100117.
[39] 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.
[40] Lin M, You Y, Meng J, Wang W, Wu J, Stroe D-I. Lithium-ion battery degradation trajectory early prediction with synthetic dataset and deep learning. J Energy Chem 2023;85:534–46.
[41] Qiu X, Wang S, Chen K. A conditional generative adversarial network-based synthetic data augmentation technique for battery state-of-charge estimation. Appl Soft Comput 2023;142:110281.
[42] Li Y, Wu H. A clustering method based on K-means algorithm. Physics Procedia 2012;25:1104–9.
[43] Müller M, editor. Dynamic time warping, in information retrieval for music and motion. Berlin, Heidelberg: Springer; 2007, p. 69–84.
[44] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is All you need. In: Advances in neural information processing systems, vol. 30, Curran Associates, Inc.; 2017.
[45] Cai L, Cui N, Jin H, Meng J, Yang S, Peng J, et al. A unified deep learning optimization paradigm for lithium-ion battery state-of-health estimation. IEEE Trans Energy Convers 2023;1–12.
[46] Meng J, Cai L, Stroe D-I, Luo G, Sui X, Teodorescu R. Lithium-ion battery state-of-health estimation in electric vehicle using optimized partial charging voltage profiles. Energy 2019;185:1054–62.
[47] Michael P. CALCE Battery Group, https://web.calce.umd.edu/batteries/data.htm.
[48] Deng Z, Xu L, Liu H, Hu X, Duan Z, Xu Y. Prognostics of battery capacity based on charging data and data-driven methods for on-road vehicles. Appl Energy 2023;339:120954.
[49] Liu K, Li Y, Hu X, Lucu M, Widanage WD. Gaussian process regression with automatic relevance determination Kernel for calendar aging prediction of lithium-ion batteries. IEEE Trans Ind Inf 2020;16(6):3767–77.
[50] Liu D, Pang J, Zhou J, Peng Y, Pecht M. Prognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression. Microelectron Reliab 2013;53(6):832–9.