Retrieval-based Battery Degradation Prediction for Battery Energy Storage Systems

Yixuan Li*†, Qirui Yang*‡, Hao Wen*, Huiwen Zheng§, Weimin Liu§, Hui Li§, Yuanchun Li*, Yunxin Liu*
*Institute for AI Industry Research (AIR), Tsinghua University, Beijing, China
†Beijing University of Posts and Telecommunications, Beijing, China
‡The Hong Kong University of Science and Technology, Hong Kong, China
§GDS Services Ltd, Shanghai, China
liyixuan0302@bupt.edu.cn,qyangau@cse.ust.hk,wen-h22@mails.tsinghua.edu.cn,
{huiwen.zheng,weimin.liu,hui.li}@gds-services.com,{liyuanchun,liuyunxin}@air.tsinghua.edu.cn

***Abstract*—Long-term battery degradation prediction is an important problem in battery energy storage system (BESS) operations, and the remaining useful life (RUL) is a main indicator that reflects the long-term battery degradation. However, predicting the RUL in an industrial BESS is challenging due to the lack of long-term battery usage data in the target environment, domain difference between BESS environments, and incomplete battery charging/discharging patterns in industrial scenarios. To solve these challenges, we propose a retrieval-based approach, which predicts the RUL of the target battery based on the full-lifetime usage data of reference batteries retrieved from other environments. The basic idea is that the reference batteries with common early-life features are more useful for predicting long-term degradation of the target battery. Based on experiments with both laboratorial datasets and industrial datasets, our method can constantly achieve higher prediction accuracy than state-of-the-art baselines.**

1. Introduction

Battery Energy Storage Systems (BESS) have emerged as an important component in numerous critical industrial scenarios, such as data centers, cellular base stations, and electric stations. They are typically used as uninterruptible power supply (UPS) systems and renewable energy storage systems. A typical BESS is composed of a myriad of interconnected rechargeable batteries, often lithium-ion batteries, which store energy derived from a variety of sources for discharge when necessary. Given the significant financial investment and critical operational role of BESS, the optimal and safe operation of these systems is a pressing concern in the industrial sector.

An intrinsic aspect of BESS operation is the degradation of battery health, a factor which significantly influences the system's overall performance and lifespan. Battery health degradation refers to the gradual decrease in battery capacity and efficiency over time due to various physical and chemical factors. As the main power supply medium in BESS, Lithium-ion batteries encounter inevitable degradation in their performance characteristics, such as battery capacity and output power, throughout their lifespans. Since accelerated battery health degradation not only leads to expedited performance deterioration but can also engender safety concerns [1], it's essential to predict the long-term degradation of battery health to help maintain battery health and provide a basis for timely battery replacement. Meanwhile, a precise prediction of battery degradation can facilitate preemptive measures to extend the battery lifespan and enhance its overall performance [2], [3], [4], [5]. A key metric to reflect long-term battery degradation is the remaining useful life (RUL), which represents the length of time (or number of cycles) a battery is likely to operate before it requires repair or replacement (e.g. when the battery capacity is only 70% of the original capacity).

Existing approaches have studied leveraging early-life battery degradation data for RUL prediction [6], [7], [8], [9]. However, their methods are mostly designed for in-laboratory battery experiments, rather than in-deployment battery operations. We summarize three main challenges for long-term battery degradation prediction in industrial BESS.

1) **Domain difference between BESS.** A primary technical challenge associated with predicting battery health degradation arises from the domain differences between battery systems. Different battery systems exhibit variations in terms of their internal chemistry, design, and operational conditions. These differences result in varying degradation patterns, making it challenging to develop a universal prediction model applicable across all battery domains.
2) **Lack of long-term degradation data.** In a typical BESS, all batteries are used as a whole and there is often only early-life data avaliable for them. This means that there is no full-lifetime reference data to learn the long-term battery degradation pattern.
3) **Incomplete battery charging/discharging.** Unlike in-laboratory experiments that usually fully charge/discharge the batteries to understand the degradation in ideal usage scenarios, the batteries used in BESS are charged/discharged based on the actual workload, making it difficult to adopt the methods that rely on full-cycle features.

Due to these challenges, it is required to predict the long-term battery degradation in the target BESS environment based on early-life data. Obviously, solely relying on the data collected from the target BESS environment is infeasible due to the absense of long-term degradation data for reference. Thus, the problem becomes how to effectively utilize the long-term usage data from other domains (e.g. other BESS or laboratory experiments). Directly learning the degradation pattern from other domains may face the data distribution shift problem [10] and lead to suboptimal performance, due to the domain difference between battery deployment environments. Meanwhile, applying transfer learning or domain adaptation techniques [11] is difficult because of the lack of long-term battery usage data in the target domain as training signals.

To address these challenges, we introduce a retrieval-based method for long-term battery degradation prediction. Our method is inspired by the Retrieval Augmented Generation (RAG) techniques that have demonstrated remarkable effectiveness in various generative tasks [12], [13], [14]. The basic idea of our approach is to retrieve the long-term usage data of source batteries that share similar early-life patterns as the target battery. The long-term usage data serves as the reference for long-term degradation prediction of the target battery. Such a design mitigate the problems of data scarcity in the target BESS and domain difference between batteries.

Specifically, we design an end-to-end architecture, named RetrieveNet, to implement the retrieval-based prediction. The key components of RetrieveNet include a battery feature encoder that converts the early-life usage data of a battery to a embedding vector, a relation analyzer that predicts the relevance between two batteries based on their early-life embeddings, and a retrieval-aggregatation module that selects top-k reference batteries from the source dataset and aggregate them for degradation prediction of the target battery. The incomplete charging/discharging problem is addressed by tailoring the battery feature encoder to rely on common charging intervals between different batteries. These components are jointly trained using the long-term usage data of various source batteries, and applied to the target battery with only early-life data available.

We conduct experiments on both public laboratorial data and private industrial data to evaluate our approach. The results have demonstrated remarkable effectiveness of our method in using small-scale eary-life data for long-term degradation prediction. For example, the accuracy of RUL prediction of our approach is 21.5% higher than state-of-the-art baselines.

The main contributions of this work are as follows:

- We study the problem of battery degradation prediction for BESS operations and highlight three challenges including domain difference, absence of full-lifetime reference data, and incomplete charging/discharging.
- We propose a novel retrieval-based battery degradation prediction method to address the challenges.
- We evaluate our approach on both laboratorial datasets and industrial datasets and demonstrate the effectiveness of our method.

2. Related Work

Degradation prediction of batteries has attracted significant attention in recent years, primarily due to its pivotal role in ensuring accurate estimation and safe operation of batteries. Researchers have devised a lot of methods in a bid to amplify the adaptability and accuracy of these prediction techniques. These methods typically span two categories, model-based techniques and data-driven approaches.

Model-based techniques: Among the model-based techniques, there has been a surge in developing electrochemical models that capture the intrinsic behaviors of lithium-ion batteries, often accompanied by aging mechanism analyses. Torchío et al. have proposed a finite volume method (FVM) within the pseudo-two-dimensional (P2D) Li-ion battery model that offers a simulation for battery pack design and management while reflecting charging and discharging cycles [15]. Li et al. introduced the single-particle model (SP), which not only estimates the state of charge but also monitors the physicochemical reactions occurring within the Li-ion battery. This model particularly stands out due to its inherent noise adaptability from sensors [16]. Recognizing the computational inefficiency inherent to the SP model, an enhanced version has been introduced, which incorporates stress-assisted diffusion. This enhancement ensures improved predictive accuracy in scenarios like changing liquid concentrations during complex discharge processes [17]. Nevertheless, these methods often encounter challenges when adapting to intricate and dynamic environments, being most suitable for specific battery types and their respective charging/discharging settings. The necessity for precise modeling of lithium-ion batteries often demands significant time and effort.

Data-driven approaches: Data-driven methods, particularly those leveraging neural networks, circumvent the need for intricate physical modeling of batteries. They have proven adept at forecasting the RUL of batteries under multifaceted conditions, thus emerging as potent alternatives to traditional modeling techniques. For instance, the SVM algorithm utilizes voltage ascent during the charging phase and changes in voltage derivative (DV) as training features, achieving higher accuracy levels than conventional neural networks [18], [19]. Additionally, the Gaussian process regression (GPR) model [20] leverages Incremental Capacity (IC) which, after wavelet filtering, undergoes hyperparameter optimization via the conjugate gradient method and the multi-island genetic algorithm (MIGA) [21]. These optimizations render predictions in a probabilistic density format, ensuring high precision but also demanding significant computational resources and intricate hyperparameter tuning. Deep learning, especially techniques like multi-scale convolutional neural networks, has demonstrated superior performance in RUL predictions by extracting features at various scales [22]. Specifically, the Multi-scale dilated

convolution network (MsDCN) with its depthwise separable convolution (DSC) has exhibited enhanced prediction accuracy and computational efficiency in comparison to traditional CNN models. While promising, the caveat with most RUL prediction methodologies is their heavy reliance on voluminous battery data, posing challenges for systems with constrained data availability.

In emerging battery scenarios, there may be instances where the battery is in its nascent cycles. Thus, long-term RUL predictions are necessitated based on limited early-cycle features and RUL information. Even prevalent techniques like transfer learning demand comprehensive cycle data from the target domain before predictions via model transfers can be executed. RetrieveNet stands out in this realm, processing, sampling, and encoding source domain features to create a retrieval library of battery characteristics.

In this paper, we introduce an innovative RUL prediction method for BESS batteries based on RetrieveNet. Our proposed methodology seeks to address the limitations inherent in both traditional and data-driven techniques, aiming to enhance the accuracy and speed of predictions for indoor climate control systems, especially under data-restricted conditions.

3. Method

Our work aims to forecast the long-term degradation of batteries by predicting their remaining useful life (RUL) leveraging early-cycle features. Since only early-cycle data is available when the model is launched, it's infeasible to directly train a DNN model with the target-domain data and achieve a satisfying prediction accuracy. Therefore, it is desired to effectively utilize the full-lifecycle battery usage data from other domains.

Inspired by the concept of Retrieval Augmented Generation (RAG), we take a retrieval-based approach to battery degradation prediction. RAG allows the model to be supplemented with existing data, thus reducing the pressure obtaining the effect of a large-scale pre-trained model using only a small-scale model. The effectiveness of RAG has been demonstrated in REALM [12], where the authors empirically showed that REALM's retrieval approach outperforms T5-11b, a model that is two orders of magnitude larger in OpenQA benchmarks, while employing a significantly smaller parameter count. RAG makes it possible to get better results with fewer parameters and less training data. Due to the excellent performance of RAG on NLP tasks, we decided to transfer the method to the problem of battery RUL prediction. To achieve this, we introduce the RetrieveNet, designed to retrieve reference batteries from the source domain with similar early-cycle features as the target battery and aggregate them to predict the RUL target of target battery.

The overall architecture is shown in Figure 1. In a nutshell, RetrieveNet predicts the future trends of a target battery with the assistance of three main components. First, a siamese sequence encoder is used to encode the early-life feature of source and target batteries into the same embedding space. Then, these embeddings are fed into a relation analyzer as source-target pairs to obtain the relevance between each source battery and the target battery. The most relevant (top-k) source batteries are retrieved, and their full-life usage data is aggregated to predict the future trend of the target battery. In this way, RetrieveNet is able to more effectively use the source-domain data despite the data shortage and domain difference issues for the target battery. The whole process is trained end to end using the full-life usage data of multiple batteries. The following sections will introduce the key components in more detail.

3.1. Feature Retrieval Library Construction

To construct the feature retrieval library, we first sample data and extract feature for each battery.

Suppose we want to predict the battery RUL based on l cycles of early-life usage data, we need to obtain the mappings between l-cycle usage data and RUL for many batteries. We adopt a sliding window mechanism to obtain such mappings from the battery sample sequences. Specifically, given a full-life battery usage trace with N cycles (N is the total number of cycles before end-of-life), we sample multiple l-cycle subsequences from it and calculate the RUL for each of them as $N - i - l$, where i is the index of the first cycle in the subsequence. When obtaining the training samples, we sample the windows starting from the 10-th cycle to avoid instability in early cycles, and set the window stride to s to reduce redundancy between different samples. For further illustration, we use x_i to represent the battery usage features at i-th cycle (usually the features contains the normalized voltage and capacity, the diffrential voltage and capacity, and relative capacity etc. in the cycle [23]), and $X_i = \{x_i, x_{i+1}, \cdots, x_{i+l}\}$ represents the sequence of features starting from i-th cycle with window size l.

Repeating the window sliding process, we can obtain a feature sequence pool which contains all sequences sampled from all source-domain batteries. Then we use an encoder to encode the sequences into feature vectors, denoted as $Enc(X_i)$. The encoder comprises two linear layers, each followed by a ReLU activation function. This structure not only enhances the model's non-linearity but also ensures the capture of complex patterns within the input features. Dimension reduction is applied for computational efficiency and to mitigate potential overfitting. Once encoded, the features are transformed into a continuous vector space, capturing essential information for assessing similarity with the source domain batteries and furnishing critical data for subsequent RUL predictions.

After this step, we have a dataset with many <feature embedding, true RUL> pairs. They are used later as the retrieval set for the target battery degradation prediction.

3.2. Relation Analysis and Reference Selection

After encoding, the feature embeddings are normalized and piped into a relation analysis module in RetrieveNet. This module computes the relevance between the encoded

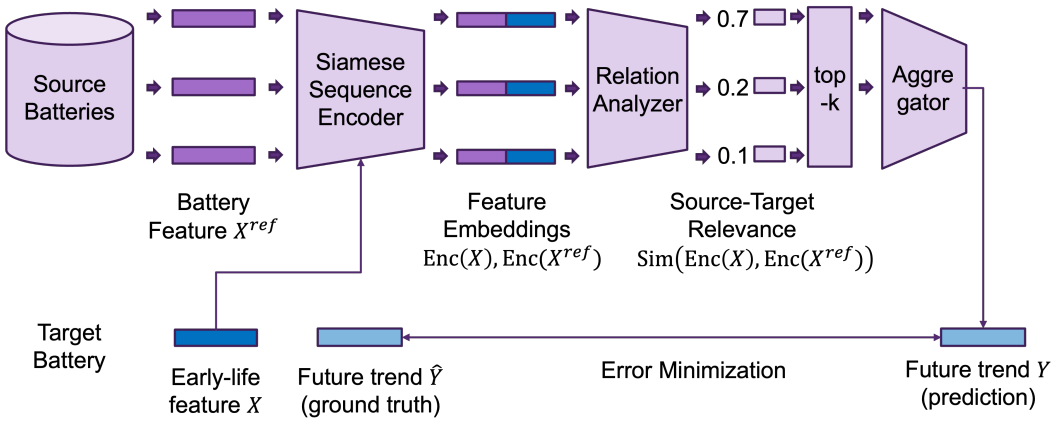

Figure 1: The overall architecture of our approach.

features of two batteries from the source and the target domains. The relation analyzer calculates the relevance between the two batteries by simply computing the dot product between their feature embeddings.

When predicting the RUL for the target battery, we first extract the feature sequence X and obtain feature embedding $Enc(X)$ following the process in Section 3.1. The relation analyzer calculates the relevance score between the target feature embedding and each feature embedding in the retrieval set. A higher relevance score indicates a higher similarity between the early-life features of the target battery and the corresponding source battery, which implies that the source batteries are more valuable to assistant the RUL prediction of the target battery. Considering the balance between accuracy and efficiency, we selects top k source batteries with the highest relevance scores to form the reference set. This strategy reduces the computational complexity and minimizes the noise from low-relevance features.

3.3. Feature Aggregation and Prediction

After obtaining the reference batteries, RetrieveNet will aggregate the information from the retrieved batteries from the source domains to predict the RUL of the target battery. We define the encoding function as Enc, the encoded feature embedding of the target battery as $Enc(X^{target})$, the retrieved feature embeddings of source batteries as $\mathcal{X}_{enc}^{source} = \{Enc(X^{source-1}), Enc(X^{source-2}), \cdots, Enc(X^{source-k})\}$, and the ground-truth RULs of the retrieved source batteries as $\mathcal{Y}^{source} = \{Y^{source-1}, Y^{source-2}, \cdots, Y^{source-k}\}$. RetrieveNet concatenates all these vectors and values to form a single vector V, which is used as the input of the final predictor. The predictor is a multi-layer perceptron (MLP), and its output is the prediction of the RUL of the target battery. Our aim for this network is to capture the non-linear relation between the target features and the retrieval features, utilizing the ground truth of the retrieved source batteries to accurately estimate the target RUL.

3.4. End-to-end Inference and Training

The steps described above are elaborated in Algorithm 1. In the algorithm, lines 2 to 4 utilize the sliding window to construct the encoded feature embeddings and store them in the library. Lines 6 and 7 randomly select a reference set and mask the vectors originating from the same battery source as the target vector. Lines 9 and 10 calculate the similarity between the target vector and the retrieved vectors, then sort and pick the top k vectors with the highest similarity, and retrieve the corresponding RUL from the library. Lines 11 to 13 aggregate the target feature vector, retrieved feature vectors, and retrieved RULs as the input to the MLP, and the output result of the MLP is taken as the predicted RUL. For the training stage, we use mean square error (MSE) as the loss function since it can intuitively elaborate the difference between predicted RUL and the ground truth.

When training the model with the source batteries, we randomly select a battery as the target battery in each step and follow the above process to obtain a predicted RUL. The loss is calculated as the mean squared error (MSE) between the predicted RUL and the actual RUL.

3.5. Source-Target Feature Alignment

An important issue that worth noted is the feature (mis)alignment between the source and target batteries in real-world BESS. Specifically, the batteries in the target BESS may have significantly different charging/discharging pattern as the source BESS (e.g. more/less frequent charging, different loads, etc.). We use data augmentation and charging interval alignment to address this issue.

Identification of Common Charging Interval for Feature Extraction. Many existing approaches assumes that the batteries are used with periodic full charges and discharges, but real-world BESS often embraces incomplete charge/discharge cycles. Although such incomplete patterns lack clear start or end points for charging phases, they're usually still regular. For instance, many BESS would charge the batteries to 100% in every night to ensure maximum

Algorithm 1 Training process of RUL prediction algorithm in RetrieveNet

Input: The feature sequences of target battery $\mathcal{X}^{target} = \{X^{target-1}, X^{target-2}, \cdots, X^{target-m}\}$, the ground truth of target battery $\mathcal{Y}^{target} = \{\hat{Y}^{target-1}, \hat{Y}^{target-2}, \cdots, \hat{Y}^{target-m}\}$, the feature sequence pool $\mathcal{L}$, the size of randomly selected retrieval set n, the number of most similar source feature embeddings k

for $i = 1$ **to** m **do**
  $\mathcal{X}^{source} \leftarrow$ RandomSelect$(\mathcal{L}, n)$
  $\mathcal{X}^{source} \leftarrow$ Mask$(\mathcal{X}^{source}, X^{target-i})$
  for $j = 1$ **to** n **do**
    $\mathcal{S} \leftarrow sim(Enc(X^{source-j}), Enc(X^{target-i}))$
  end for
  $\mathcal{S}^{sorted} \leftarrow \mathbf{Sort}(\mathcal{S})$ *by descending order*
  for $p = 1$ **to** k **do**
    $q \leftarrow GetIndex(\mathcal{S}^{sorted}, p)$
    $\mathcal{X}_{enc}^{source} \leftarrow Enc(X^{source-q})$
    $\mathcal{Y}^{source} \leftarrow Y^{source-q}$
  end for
  $\mathcal{V} \leftarrow \{X^{target}, \mathcal{X}_{enc}^{source}, \mathcal{Y}^{source}\}$
  $Y_{target-i} \leftarrow$ MLP$(\mathcal{V})$
  $loss =$ MSE$(Y^{target-i}, \hat{Y}^{target-i})$
  Update parameters via gradient descent
end for

capacity available in the next day. Building on this observation, we propose to identify a periodic charging interval (e.g. charging from 50% to 90%) that is shared between source and target batteries, and extract features from this interval to ensure feature alighment. The extracted features can be directly used in Algorithm 1 in the same way as fully charged/discharged batteries.

Extending the Retrieval Set with Data Augmentation. Another problem is that the source and target batteries may not be used at the same pace. For example, some batteries are fully charged once in every week and some may be charged every day. This may lead to insufficient reference samples in the retrieval set that share similar early-life feature with the target battery. To address this problem, we propose to create more samples by scaling the battery usage data in the time domain. Specifically, we introduce an spline interpolation-based data augmentation approach. Spline interpolation is popular since the interpolation error can be very small even with using low-degree polynomials for the spline, and only require few computational cost. To apply spline interpolation for data augmentation, we first create new sampling cycles based on given scaling rates. Then create interpolation splines between original data points and calculate values of the polynomials at those sampling cycles to get the new data points. The degree of the polynomials can be various according to the complexity of the original data. In the battery charging and discharging scenarios, we use the cubic polynomials as the spline function.

4. Evaluation

4.1. Experimental Setup

Datasets. We evaluate our method on two publicly accessible laboratorial datasets and a industrial dataset collected from an IDC. Each dataset is characterized by diverse charge-discharge regimes, ambient thermal conditions, and battery chemical compositions.

The first public dataset, referred to as the "EES dataset", was derived from the work of Ma et al. [23] published in Energy & Environmental Science. This dataset aggregates data from 77 LFP/graphite cells. Each cell underwent a distinct multi-stage discharge protocol, albeit they uniformly conformed to a singular rapid charging strategy. This uniformity accentuates the intricacies of the discharge process. Another public dataset, termed the "NE dataset", was introduced in Severson et al. [24] published in Nature Energy. It assimilates information from over 100 LFP/graphite cells, governed by 72 diverse charging protocols. Notwithstanding the heterogeneity in charging routines, a consistent discharge protocol is prevalent across the cells, thus underscoring the variability inherent in the charging dynamics.

The industrial dataset is collected from an Internet Datacenter (IDC), henceforth denoted as the "IDC dataset". The datacenter is equipped with two Battery Energy Storage Systems (BESS), wherein each BESS contains 8 battery modules. Every individual module is constituted of 20 serially connected lithium iron phosphate batteries. Primarily designed to act as auxiliary power reserves, these batteries ensure the continuity of operations without interruptions. Periodic charging/discharging assessments are conducted monthly to validate their operational efficacy. In our experiments, one BESS was designated as the source while the other was treated as the target.

Baselines. We compared our method with the methods used in Ma et al. [23] and Severson et al. [24] (abbreviated as EES and NE respectively), three supervised training approaches using different models (MLP, RNN, and LSTM), an autoregressive timeseries forecasting method (ARIMA). All methods except for ARIMA use the same training data to train the prediction model. ARIMA use the early-life capacity values of the test battery to iteratively predict the next-cycle capacity.

Metrics. In order to provide a rigorous quantitative evaluation of the RUL prediction models, three universally recognized metrics were utilized: RMSE, R^2, and MAPE.

1) **Root Mean Square Error (RMSE)**: RMSE offers a measure of the standard deviation of the residuals, providing insight into the discrepancies between the predicted and observed RULs. It is mathematically represented as RMSE $= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y^i - \hat{y}^i)^2}$, where $\hat{y}^i$is the RUL predicted at the ith cycle, y^i denotes the actual observed RUL, and n corresponds to the total cycles associated with a particular cell.
2) **Coefficient of Determination (R^2)**: The R^2 value delineates the fraction of variance in the depen-

dent variable that can be ascertained from the independent variable(s). It can be articulated as $R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}^i - \bar{y})^2}{\sum_{i=1}^{n}(y^i - \bar{y})^2}$, with $\bar{y}$ being the average RUL, determined by $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y^i$.

3) **Mean Absolute Percentage Error (MAPE)**: MAPE quantifies the prediction accuracy in a percentage format, offering the advantage of rendering errors in a consistent unit scale. The equation for MAPE is given by: MAPE $= \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y^i - \hat{y}^i}{y}\right|$. In this context, y symbolizes the initial cycle life of the cell during RUL prediction. Utilization of this metric assists in counteracting the disproportionate impact of errors observed towards a battery's later stages, where absolute variations can greatly exceed authentic RUL values.

For an encompassing assessment over diverse test cells, the arithmetic mean of these metrics was determined across all test cells for subsequent comparison.

4.2. Results

EES Dataset. On the EES dataset, 55 batteries was employed to formulate the training set and establish the retrieval library. The RUL prediction accuracy is measured on 22 other test batteries. The overall comparison with the baseline methods is shown in Table 1. The EES and MLP methods exhibited remarkable prediction accuracy across all baselines, with an RMSE of 186 cycles and 182 cycles respectively Yet, RetrieveNet demonstrated a global RMSE of 146 cycles, a MAPE of 6.19%, and an R^2 of 0.881, outperforming all of the baselines, coupled with amplified confidence.

The detailed prediction results of RetrieveNet is illustrated in Figure 2. The initial predictive outcomes exhibited minor deviations around the genuine RUL. Conversely, with the progression of battery age, these forecasts steadily converged with the factual values, culminating in a nuanced approximation of the RUL, underscoring the viability of the adopted methodology. It is worth noting that a predominant portion of the relative discrepancies resided within the range of [-200 cycles, 200 cycles], implying that our model offers a precise prediction of the RUL for every battery charge-discharge cycle, demonstrating its reliability.

NE Dataset. On the NE public dataset, the prediction models are trained with the data of 82 batteries and applied to estimate the RUL of 38 test batteries. A subset of four batteries was omitted from the analysis due to an elevated level of measurement noise. The RetrieveNet method showcased an aggregate RMSE of 73 cycles, a MAPE of 7.3%, and an R^2 of 0.767, as detailed in Table 1. In contrast with the conventional NE model, there was a decrement of 65.9% in the prediction RMSE, denoting a substantial augmentation in prediction precision. Employing the analogous methodology, the MLP Model, RNN Model, and LSTM Model posted RMSE metrics of 82, 79, and 78 cycles respectively on the NE dataset, corroborating the effectiveness of the established framework in RUL prognostication.

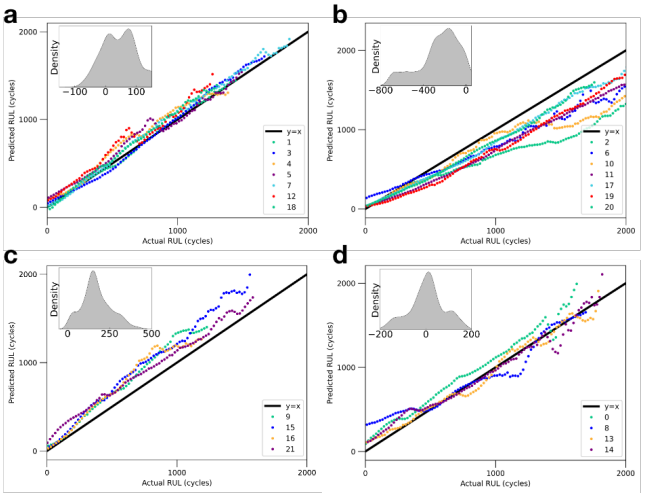

Figure 2: Predicted results for the 22 EES test batteries. Subfigures (a), (b), (c), and (d) depict the predicted RUL versus the actual RUL for each test battery, where the solid line represents the ideal outcome when predicted RUL consistently matches the actual RUL. The figure also presents the density distribution of the RUL prediction error. For clarity in representation, RUL is shown every 20 cycles.

As delineated in Figure 3, a predominant segment of batteries, during the forecasting phase, adeptly anticipated the authentic RUL. Moreover, with an escalation in the enumerated initial battery cycles, symbolized as K, the metrics of both RMSE and MAPE observed a consistent decrement, congruent with the projected outcomes. As the volume of anticipated battery feature cycles in the targeted domain amplifies, a richer congruence emerges within the feature retrieval library, culminating in enhanced precision and a contraction in prediction discrepancies. In pragmatic manufacturing requisites, contingent on the extant feature magnitude of the target battery in focus, the congruent feature cycle span can be judiciously modulated, thereby realizing a dynamic and adaptable RUL prediction.

IDC Dataset. In the evaluation of the IDC dataset, which encompasses real industrial batteries, it was discerned that these batteries experience fractional charge and discharge cycles contingent upon manufacturing stipulations. Given that the degradation in battery capacity hadn't yet surpassed the demarcated threshold and reached the end of life, it became infeasible to obtain the groundtruth RUL of the batteries. Therefore, when applying our prediction method to the IDC dataset, we let the model predict the battery SOH (state of health) after certain number of days (e.g. 50, 100, 200 days), rather the predicting the remaining days to reach certain capacity (i.e. RUL). These two goals are equivalent from the prespective of degradation prediction algorithms.

The charge-discharge dynamics of IDC batteries presented a heterogeneous pattern, revealing habitual yet non-absolute cycles in pragmatic deployment. Temporal days were designated as the metric intervals for battery attributes and the corresponding SOH. To bolster data heterogene-

TABLE 1: The RUL prediction results of different methods on two public datasets.

Metric	Method	EES Dataset			NE Dataset		
RMSE (Cycles) (lower is better)	EES	/	/	186	/	/	/
	NE	/	/	/	214	/	/
	MLP	190	205	182	82	79	68
	RNN	246	218	235	89	87	69
	LSTM	180	128	522	78	67	65
	ARIMA	741	773	726	1150	685	646
	RetrieveNet	173	166	146	73	59.63	58
R2 (higher is better)	EES	/	/	0.804	/	/	/
	NE	/	/	/	/	/	/
	MLP	0.819	0.737	0.761	0.744	0.682	0.693
	RNN	0.676	0.743	0.602	0.613	0.482	0.304
	LSTM	0.814	0.92	-0.497	0.723	0.769	0.371
	ARIMA	-1.53	-1.9	-1.81	-144	-90.1	-108
	RetrieveNet	0.856	0.857	0.881	0.767	0.763	0.772
MAPE (%) (lower is better)	EES	/	/	8.72	/	/	/
	NE	/	/	/	10.72	/	/
	MLP	8.4	8.92	7.93	9.91	9.12	7.91
	RNN	12	9.51	11.1	8.34	10.55	8.49
	LSTM	7.51	5.27	23.8	8.42	7.31	7.49
	ARIMA	33.7	35.9	33.4	219	122	113
	RetrieveNet	7.7	7.08	6.19	7.35	6.04	5.89

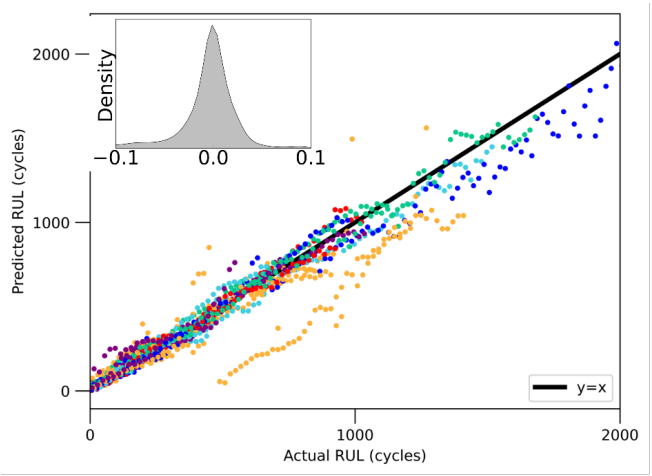

Figure 3: Predicted results for the 38 NE test batteries. The figure depicts the predicted RUL versus the actual RUL for each test battery, where the solid line represents the ideal outcome when predicted RUL consistently matches the actual RUL. The figure also presents the density distribution of the RUL prediction error. For clarity in representation, RUL is shown every 20 cycles.

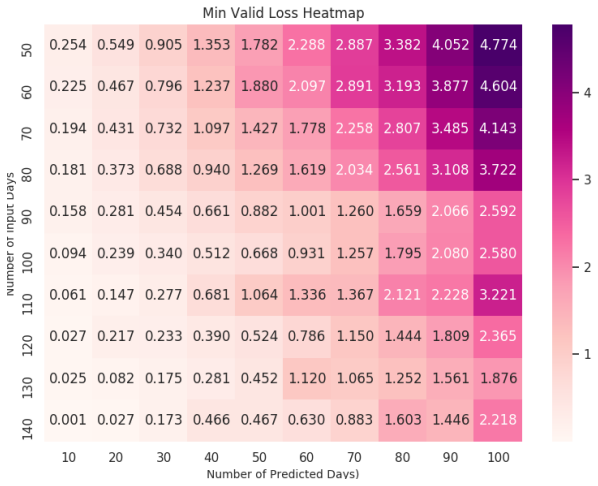

Figure 4: The prediction error heatmap on the IDC dataset. The vertical i represents the number of days of the input SOH data, and the horizontal j denotes the number of days for the predicted SOH. The values represent the loss between the predicted SOH and the actual SOH.

ity, retrospective SOH data from batteries situated in the IDC's northern region underwent augmentation via spline interpolation-centric data enhancement methodologies. Pursuant to this enrichment, sampling intervals materialized at delineated magnification ratios, culminating in the formation of a battery health retrieval compendium.

During the model's training phase, the inaugural SOH metric from the prevailing i-cycle sequence in one group was harnessed as a benchmark. Concomitant trajectories were discerned within the retrieval repository, forecasting the cycle count j' requisite to attain the SOH post the j-th cycle. The topmost k congruent trajectories were earmarked as benchmarks. Their forecasted metrics were weighted vis-à-vis their congruence scores, aggregated, and the discrepancy between the forecasted j' cycle and the factual j cycle was quantified as the loss. An analogous computational strategy was appropriated during the evaluation phase, with the results explicated in Figure 4.

The RetrieveNet model manifested remarkable precision on this empirical battery dataset. Utilizing 50 days of accessible early-life SOH records and forecasting the SOH

after 100 days incurred a deviation of 4.77%. Conversely, when 100 days of antecedent SOH records were utilized, the deviation for prognosticating the forthcoming 100 days' SOH diminished to 2.58%, proficiently addressing tangible battery health forecasting exigencies. The reasonable pattern in the prediction error heatmap also demonstrates the consistency of the RetrieveNet model across different usage settings.

5. Conclusion

We introduce a battery degradation prediction method tailored for BESS. Specifically, we design a retrieval-based method to deal with the data scarcity, domain difference, and incomplete charging/discharging characteristics of batteries in BESS. The effectiveness of our method is demonstrated with experiments on both laboratorial and industrial datasets. A possible threat the validity of our method is that the shortage of full-life battery usage data and the huge domain difference between BESS domains may make it harder to retrieve meaningful references for degradation prediction. Our future work plans to further test and improve our method on more large-scale real industrial datasets with more complex battery dynamics.

References

[1] E. Martinez-Laserna, E. Sarasketa-Zabala, I. V. Sarria, D.-I. Stroe, M. Swierczynski, A. Warnecke, J.-M. Timmermans, S. Goutam, N. Omar, and P. Rodriguez, "Technical viability of battery second life: A study from the ageing perspective," *IEEE Transactions on Industry Applications*, vol. 54, no. 3, pp. 2703–2713, 2018.

[2] R. R. Richardson, M. A. Osborne, and D. A. Howey, "Battery health prediction under generalized conditions using a gaussian process transition model," *Journal of Energy Storage*, vol. 23, pp. 320–328, 2019.

[3] S. Greenbank and D. Howey, "Automated feature extraction and selection for data-driven models of rapid battery capacity fade and end of life," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 5, pp. 2965–2973, 2021.

[4] L. Liao and F. Köttig, "A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction," *Applied Soft Computing*, vol. 44, pp. 191–199, 2016.

[5] Y. Chen, Y. He, Z. Li, L. Chen, and C. Zhang, "Remaining useful life prediction and state of health diagnosis of lithium-ion battery based on second-order central difference particle filter," *IEEE Access*, vol. 8, pp. 37 305–37 313, 2020.

[6] W. He, N. Williard, M. Osterman, and M. Pecht, "Prognostics of lithium-ion batteries based on dempster–shafer theory and the bayesian monte carlo method," *Journal of Power Sources*, vol. 196, no. 23, pp. 10 314–10 321, 2011.

[7] F. Yang, D. Wang, Y. Xing, and K.-L. Tsui, "Prognostics of li (nimnco) o2-based lithium-ion batteries using a novel battery degradation model," *Microelectronics Reliability*, vol. 70, pp. 70–78, 2017.

[8] A. Singh, C. Feltner, J. Peck, and K. I. Kuhn, "Data driven prediction of battery cycle life before capacity degradation," *arXiv preprint arXiv:2110.09687*, 2021.

[9] N. H. Paulson, J. Kubal, L. Ward, S. Saxena, W. Lu, and S. J. Babinec, "Feature engineering for machine learning enabled early prediction of battery lifetime," *Journal of Power Sources*, vol. 527, p. 231127, 2022.

[10] R. Taori, A. Dave, V. Shankar, N. Carlini, B. Recht, and L. Schmidt, "Measuring robustness to natural distribution shifts in image classification," *Advances in Neural Information Processing Systems*, vol. 33, pp. 18 583–18 599, 2020.

[11] A. Farahani, S. Voghoei, K. Rasheed, and H. R. Arabnia, "A brief review of domain adaptation," *Advances in data science and information engineering: proceedings from ICDATA 2020 and IKE 2020*, pp. 877–894, 2021.

[12] K. Guu, K. Lee, Z. Tung, P. Pasupat, and M. Chang, "Retrieval augmented language model pre-training," in *International conference on machine learning*. PMLR, 2020, pp. 3929–3938.

[13] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel *et al.*, "Retrieval-augmented generation for knowledge-intensive nlp tasks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 9459–9474, 2020.

[14] Y. Mao, P. He, X. Liu, Y. Shen, J. Gao, J. Han, and W. Chen, "Generation-augmented retrieval for open-domain question answering," *arXiv preprint arXiv:2009.08553*, 2020.

[15] M. Torchio, L. Magni, R. B. Gopaluni, R. D. Braatz, and D. M. Raimondo, "Lionsimba: a matlab framework based on a finite volume model suitable for li-ion battery design, simulation, and control," *Journal of The Electrochemical Society*, vol. 163, no. 7, p. A1192, 2016.

[16] W. Li, Y. Fan, F. Ringbeck, D. Jöst, X. Han, M. Ouyang, and D. U. Sauer, "Electrochemical model-based state estimation for lithium-ion batteries with adaptive unscented kalman filter," *Journal of Power Sources*, vol. 476, p. 228534, 2020.

[17] J. Li, N. Lotfi, R. G. Landers, and J. Park, "A single particle model for lithium-ion batteries with electrolyte and stress-enhanced diffusion physics," *Journal of The Electrochemical Society*, vol. 164, no. 4, p. A874, 2017.

[18] X. Li, X. Shu, J. Shen, R. Xiao, W. Yan, and Z. Chen, "An on-board remaining useful life estimation algorithm for lithium-ion batteries of electric vehicles," *Energies*, vol. 10, no. 5, p. 691, 2017.

[19] Z. Deng, L. Yang, Y. Cai, H. Deng, and L. Sun, "Online available capacity prediction and state of charge estimation based on advanced data-driven algorithms for lithium iron phosphate battery," *Energy*, vol. 112, pp. 469–480, 2016.

[20] D. Yang, X. Zhang, R. Pan, Y. Wang, and Z. Chen, "A novel gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve," *Journal of Power Sources*, vol. 384, pp. 387–395, 2018.

[21] Z. Wang, J. Ma, and L. Zhang, "State-of-health estimation for lithium-ion batteries based on the multi-island genetic algorithm and the gaussian process regression," *Ieee Access*, vol. 5, pp. 21 286–21 295, 2017.

[22] F. Deng, Y. Bi, Y. Liu, and S. Yang, "Deep-learning-based remaining useful life prediction based on a multi-scale dilated convolution network," *Mathematics*, vol. 9, no. 23, p. 3035, 2021.

[23] G. Ma, S. Xu, B. Jiang, C. Cheng, X. Yang, Y. Shen, T. Yang, Y. Huang, H. Ding, and Y. Yuan, "Real-time personalized health status prediction of lithium-ion batteries using deep transfer learning," *Energy & Environmental Science*, vol. 15, no. 10, pp. 4083–4094, 2022.

[24] K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis *et al.*, "Data-driven prediction of battery cycle life before capacity degradation," *Nature Energy*, vol. 4, no. 5, pp. 383–391, 2019.