DiffBatt: A Diffusion Model for Battery Degradation Prediction and Synthesis

Hamidreza Eivazi[1,5,*] **André Hebenbrock**[2,5] **Raphael Ginster**[3,5] **Steffen Blömeke**[4,5]
Stefan Wittek[1] **Christoph Herrmann**[4,5] **Thomas S. Spengler**[3,5]
Thomas Turek[2,5] **Andreas Rausch**[1,5]

[1] Institute for Software and Systems Engineering,
Clausthal University of Technology, Germany

[2] Institute of Chemical and Electrochemical Process Engineering,
Clausthal University of Technology, Germany

[3] Institute of Automotive Management and Industrial Production,
Technische Universität Braunschweig, Germany

[4] Institute of Machine Tools and Production Technology,
Technische Universität Braunschweig, Germany

[5] Battery LabFactory Braunschweig (BLB),
Technische Universität Braunschweig, Germany

Abstract

Battery degradation remains a critical challenge in the pursuit of green technologies and sustainable energy solutions. Despite significant research efforts, predicting battery capacity loss accurately remains a formidable task due to its complex nature, influenced by both aging and cycling behaviors. To address this challenge, we introduce a novel general-purpose model for battery degradation prediction and synthesis, DiffBatt. Leveraging an innovative combination of conditional and unconditional diffusion models with classifier-free guidance and transformer architecture, DiffBatt achieves high expressivity and scalability. DiffBatt operates as a probabilistic model to capture uncertainty in aging behaviors and a generative model to simulate battery degradation. The performance of the model excels in prediction tasks while also enabling the generation of synthetic degradation curves, facilitating enhanced model training by data augmentation. In the remaining useful life prediction task, DiffBatt provides accurate results with a mean RMSE of 196 cycles across all datasets, outperforming all other models and demonstrating superior generalizability. This work represents an important step towards developing foundational models for battery degradation.

1 Introduction

1.1 Lithium-ion batteries

Lithium-ion (Li-ion) batteries are key technologies in the field of energy storage, with applications spanning portable electronics and electric vehicles [4]. The prominence of these batteries is largely attributable to their high energy density, which enables substantial energy storage within a compact and lightweight form factor. Moreover, Li-ion batteries demonstrate an extended cycle life compared

*Corresponding author: `he76@tu-clausthal.de` (https://orcid.org/0000-0003-3650-4107)

Foundation Models for Science Workshop,38th Conference on Neural Information Processing Systems (NeurIPS 2024).

to other battery technologies, quantified in terms of charge and discharge cycles, thereby enhancing their cost-effectiveness for long-term usage. The low self-discharge rate of Li-ion batteries ensures minimal energy loss during periods of inactivity, which is a significant advantage over other battery technologies [16, 41]. Nevertheless, several challenges remain. Safety continues to be a major issue, as mechanical damage or improper handling can potentially lead to hazardous events such as thermal runaway [8]. Furthermore, the economic and environmental implications of Li-ion battery production, recycling, and disposal present additional complexities that warrant ongoing investigation [3, 17]. One persistent challenge that continues to impact their long-term performance and reliability is capacity degradation.

1.2 Capacity degradation

The phenomenon of capacity degradation in Li-ion batteries is a multifaceted issue that encompasses both the effects of aging and the effects of cycling. Aging behavior, which is often referred to as calendar aging, pertains to the decline in battery performance over time, irrespective of active usage. Factors such as ambient temperature, state of charge, and storage conditions play a significant role in this degradation mode. In contrast, cycling behavior, also termed cycle aging, is linked to the deterioration that batteries experience during charge and discharge cycles. High charge-discharge rates and frequent cycling result in the accumulation of irreversible changes within the battery's electrochemical structure. This degradation is driven by a number of factors, including the formation of a solid electrolyte interphase layer, electrolyte decomposition, and the growth of lithium plating [7, 13]. Both aging and cycling behaviors collectively result in overall degradation, reducing the battery's capability to store and deliver electric charge [39] over its operational lifespan. In addition, the aging of batteries is very individual depending on, e.g., usage behavior and environmental conditions, which makes a basic understanding difficult and hinders individual battery management. Despite extensive research, accurately predicting the rate and extent of capacity loss remains a formidable challenge [35]. Therefore, advanced modeling techniques, including machine learning, are increasingly being employed to provide more accurate predictions of the battery degradation processes.

1.3 Machine learning in battery life prediction

The degradation of a battery can be quantified by using key performance indicators, including the state of health (SOH) and the remaining useful life (RUL) [28]. The SOH is a measure of the current capacity of a battery relative to its original capacity, expressed as a percentage, and provides insight into the extent of capacity degradation [10, 38]. In contrast, the RUL is a predictive measure that estimates the remaining operational cycles of a battery before it reaches defined performance criteria [31].

Methods for battery degradation modeling can be classified according to Rauf et al. [37] into four domains: i) physics-based models, ii) empirical models, iii) data-driven methods (DDMs), and iv) hybrid methods. Among these various domains, DDMs are emerging as a prominent technique for developing battery degradation models. This is due to the flexibility and independence from specific model assumptions that these approaches offer. In the domain of DDMs, machine learning (ML) methods are widely regarded as one of the most effective approaches for estimating RUL and SOH, due to their ability to address non-linear problems [37]. Since all battery RUL and SOH prediction tasks are effectively regression problems, supervised learning is the most commonly used approach in ML battery studies.

Recent literature reviews [28, 31, 37, 38, 43] indicate that various ML methods are utilized for modeling battery degradation. In the area of artificial neural networks, shallow neural networks can capture nonlinear relationships among an arbitrary number of inputs and outputs, however, they are hindered by slow training processes and a propensity to converge at local minima [28, 38]. In contrast, deep learning algorithms demonstrate superior performance in managing large datasets due to their specialized architectures. They provide higher accuracy and enhanced generalization capabilities but incur significant computational costs [38]. Techniques such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks are commonly employed in this context [31, 37, 43]. Additionally, support vector machines achieve a commendable balance between generalization capability and estimation accuracy. However, they may struggle with scalability on larger datasets [38]. Similarly, relevance vector machines have the disadvantage

of requiring extensive datasets, which results in significant computational complexity. However, they offer the advantage of high accuracy, robust learning capabilities, and the capacity to generate predictions with associated probability distributions [37]. Lastly, Gaussian process regression (GPR) methods are advantageous for their ability to quantify the uncertainty of estimated values, which is particularly valuable in practical applications. Nonetheless, GPR methods typically exhibit lower efficiency in high-dimensional spaces and can be computationally complex [28, 38]. Another recent approach, with a sole focus on the prediction of the SOH, is presented by Luo et al. [32], in which the authors introduce the methodology of diffusion models as a promising avenue for SOH prediction.

Despite the widespread application of ML models in battery degradation analysis, comparing these various approaches presents significant challenges. Many studies utilize different datasets, which are often not publicly available due to confidentiality concerns. To address this issue, BatteryML was developed by Zhang et al. [47], offering a standardized method for data representation that consolidates and harmonizes all accessible public battery datasets. Additionally, BatteryML establishes clear benchmarks for predicting RUL and includes a range of models, such as linear models, tree-based models, and neural networks, tailored for battery degradation prediction. Lastly, BatteryML also introduces the transformer architecture as a novel approach for predicting SOH and RUL [47].

1.4 Contribution to literature

Our contribution represents a threefold advancement in the domain of battery degradation prediction. First, we introduce DiffBatt, a novel application of denoising diffusion probabilistic models (DDPMs) specifically tailored for estimating the SOH and RUL of Li-ion batteries. The presented approach is motivated by the need to handle the stochastic and intricate nature of battery degradation. This innovative approach employs the capabilities of diffusion models to more effectively capture the complex degradation behavior than traditional methods (Section 2.2). Secondly, our results demonstrate a significant improvement over existing benchmarks, showcasing superior accuracy and reliability in predictions (Section 3). Finally, we establish the groundwork for further development by positioning our model as a foundation model, enabling future enhancements and adaptations to diverse energy storage related applications (Section 4). All the codes and pre-trained models are available on GitHub: `https://github.com/HamidrezaEiv/DiffBatt.git`.

2 Methodology

DDPMs [21] are an expressive and flexible family of generative models that utilize a parameterized Markov chain to produce high-quality samples that match the characteristics of the training data. The key idea behind DDPMs is to learn a reverse process, also known as the reverse diffusion or denoising process, which gradually removes noise from a sequence of noisy samples until it reveals the original signal. The forward process of a DDPM is a Markov chain that progressively adds noise to the data in the opposite direction of sampling. This process continues until the signal is completely obscured by noise. The model's objective is to learn a set of transformations that can effectively undo this noise accumulation and recover the original signal.

The training of DDPMs involves using variational inference to optimize the model parameters, with the goal of minimizing the difference between the generated samples and the true data distribution. This is achieved by estimating the lower bound of the loss function along a large number of diffusion steps, which are computed iteratively during the training process. The resulting model can generate new, diverse samples that resemble the original data, making it useful for applications such as image and audio synthesis [12, 22, 45], data augmentation [33], and more [2, 14, 29, 46].

We follow the principles proposed by Ho and Salimans [20] for classifier-free diffusion guidance to increase sample quality while decreasing sample diversity in diffusion models. Classifier-free guidance serves to achieve similar objectives for performing truncated or low-temperature sampling in certain generative models, such as generative adversarial networks (GANs) and flow-based models. The intended outcome is a decrease in sample diversity accompanied by an increase in individual sample quality. Examples are truncation in BigGAN [6] and low-temperature sampling in Glow [26], which lead to a trade-off curve between the Fréchet inception distance (FID) score and the inception score. This enables the flexibility to generate high-quality or more diverse samples when predicting or synthesizing battery degradation, respectively.

DiffBatt is based on a DDPM enhanced with transformer models and utilizes classifier-free diffusion guidance for conditional generative modeling. In the following, we provide a brief discussion on the methodological aspects of DDPMs and classifier-free guidance.

2.1 Background

Denoising diffusion models learn to systematically transform a sample of a simple prior, typically a unit Gaussian, to a sample from an unknown data distribution $q(\boldsymbol{x})$. In a fixed *forward process*, a given data sample $\boldsymbol{x}_0 \sim q(\boldsymbol{x})$ is corrupted by Gaussian noise according to a variance schedule $\{\beta_t \in (0,1)\}_{t=1}^T$ over the course of T timesteps

$$q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0) = \prod_{t=1}^{T} q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}), \quad q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = \mathcal{N}(\boldsymbol{x}_t; \sqrt{1-\beta_t}\boldsymbol{x}_{t-1}, \beta_t \boldsymbol{I}). \tag{1}$$

The so-called *reverse process* is considered to generate new samples

$$q(\boldsymbol{x}_{0:T}) = p(\boldsymbol{x}_T) \prod_{t=1}^{T} q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t), \quad q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) = \mathcal{N}(\boldsymbol{x}_{t-1}; \boldsymbol{\mu}(\boldsymbol{x}_t, t), \boldsymbol{\Sigma}(\boldsymbol{x}_t, t)), \tag{2}$$

in which the unknown true inverse conditional distribution $q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$ is approximated by a neural network $p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$ parameterized by θ. The model seeks to learn an estimator for the mean parameter $\boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t)$, under the constraint that the covariance remains unchanged as

$$\boldsymbol{\Sigma}(\boldsymbol{x}_t, t) = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \boldsymbol{I} = \Sigma_t \boldsymbol{I} \tag{3}$$

with $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$, $\alpha_t = 1 - \beta_t$. The mean $\boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t)$ is parameterized as

$$\boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)\right), \tag{4}$$

where $\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t$ for $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, $\boldsymbol{\epsilon}_\theta$ is a function approximator intended to predict $\boldsymbol{\epsilon}_t$ from $\boldsymbol{x}_t$ and $\boldsymbol{\epsilon}_t$ indicates Gaussian noise to diffuse $\boldsymbol{x}_0$ to $\boldsymbol{x}_t$. The reverse process is trained to approximate the joint distribution of the forward process by optimizing the evidence lower bound. With the parameterizations suggested in [22] the loss simplifies to

$$L_{\text{simple}}(\theta) := \mathbb{E}_{t,\boldsymbol{x}_0,\boldsymbol{\epsilon}}\left[\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)\|^2\right], \tag{5}$$

resembling denoising score matching over multiple noise scales. To summarize, we can train the reverse process to predict $\boldsymbol{\epsilon}_t$. To learn a conditional model $p_\theta(\boldsymbol{x}_0|\boldsymbol{c})$ the diffusion model is extended by incorporating the conditioning variable $\boldsymbol{c}$ into the reverse process

$$\begin{aligned} p_\theta(\boldsymbol{x}_{0:T}|\boldsymbol{c}) &= p(\boldsymbol{x}_T) \prod_{t=1}^{T} q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{c}), \\ p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{c}) &= \mathcal{N}(\boldsymbol{x}_{t-1}; \boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t, \boldsymbol{c}), \boldsymbol{\Sigma}_\theta(\boldsymbol{x}_t, t, \boldsymbol{c})). \end{aligned} \tag{6}$$

Following *classifier-free guidance* [20], we train an unconditional DDPM alongside the conditional one by randomly setting the conditioning information $\boldsymbol{c}$ to a null token with probability p_{uncond}, set as a hyperparameter. To generate samples, we combine the scores from the conditional and unconditional models

$$\tilde{\boldsymbol{\epsilon}}_\theta(\boldsymbol{x}_t, t, \boldsymbol{c}) = (1+w)\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t, \boldsymbol{c}) - w\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t), \tag{7}$$

where w is the guidance strength.

2.2 Architecture

Similar to the current state-of-the-art architectures for image and audio diffusion models [9, 45], DiffBatt is based on a U-Net architecture (see Fig. 1a) and employs diffusion processes to generate SOH curves similar to a time series generation task. Conditioning for battery information, e.g., the capacity matrix, or a diffusion timestep t, is provided by adding embeddings into intermediate

layers of the network [21]. The model consists of residual blocks with one-dimensional convolutions, attention modules, and pooling and up-sampling layers. In the reverse process, the model takes a one-dimensional Gaussian noise as input and generates a sample of SOH degradation. Notably, SOH typically exhibits a decreasing trend with progressive cycling. To better capture this physical behavior, we append a positional encoding to the output of the first convolution, allowing the denoising process to incorporate knowledge of the cycle number. Figure 1 depicts a schematic view of the model architecture.

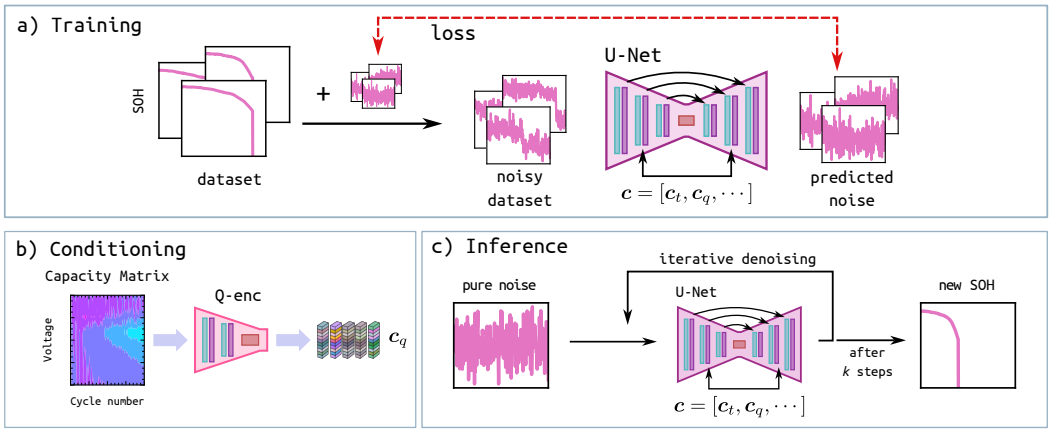


Figure 1: Schematic view of the model architecture. Adapted and modified from the work by Fürrutter et al. [14], with permission from the authors. Modifications include context-specific changes.

DiffBatt can transform Gaussian noise into a new SOH curve through the reverse diffusion process, as illustrated in Fig. 1c and Fig. 2. For this study, we employ the concept of the *capacity matrix* ($\boldsymbol{Q}$), as introduced by Attia et al. [1], as an additional condition for the diffusion process. The capacity matrix serves as a compact representation of battery electrochemical cycling data, incorporating a series of feature representations. Consistent with prior research on machine learning for predicting battery degradation [1, 40, 47], we utilize the capacity matrix corresponding to the first 100 cycles. This choice is driven by the high costs, time, and effort associated with long-term battery testing. Our goal is to leverage early life performance data to predict battery degradation and minimize resource expenditure. To encode $\boldsymbol{Q}$ into an embedding ($\boldsymbol{c}_q$), we utilize a transformer encoder (see Fig. 1b). This allows DiffBatt to generate SOH curves, from which the RUL can be derived by calculating the number of cycles until the SOH drops below a specified threshold, such as 80% of the nominal capacity.

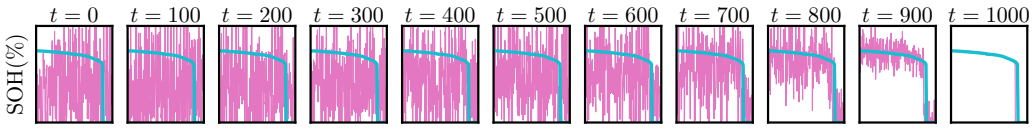


Figure 2: Denoising steps for one test sample of the MATR dataset.

2.3 Data

We conduct a comprehensive evaluation based on several publicly accessible battery datasets, as detailed in BatteryML [47]. The datasets included are CALCE [18, 44], HUST [34], MATR [23, 40], RWTH [30], SNL [36], and UL_PUR [24, 25]. These datasets encompass various battery chemistries including lithium iron phosphate (LFP), lithium cobalt oxide (LCO), nickel manganese cobalt oxide (NMC), nickel cobalt aluminum oxide (NCA), and a combination of NMC and LCO (NMC_LCO). The datasets provide diverse information on battery materials, capacities, voltages, temperatures, and state of charge (SOC) and RUL ranges. We combine several datasets as recommended in the BatteryML [47] study: CRUH (combining CALCE, RWTH, UL_PUR, and HNEI), CRUSH (combining CALCE, RWTH, UL_PUR, SNL, and HNEI), and MIX (including all datasets except SNL). In addition, we

utilize the data splits provided by BatteryML to maintain consistency and comparability, and we benchmark our results against those reported in their study. We refer to Appendix A.1 and Zhang et al. [47] for more detailed information on the data.

3 Results and discussion

Our model can be utilized for RUL prediction, SOH estimation, and SOH synthesis. It should be noted that we refer to generating new SOH curves for data augmentation as SOH synthesis. SOH and RUL prediction are two common tasks in managing inevitable capacity fade, widely discussed in the literature. While both tasks aim to predict capacity fade, they utilize different data representations. In this section, we conduct experiments based on the BatteryML benchmark tests [47] and compare the results. Furthermore, we explore the potential of our model to generate synthetic data, thereby augmenting limited battery degradation datasets for ML tasks. Further detailed analysis and results are included in Appendices A.2 and A.3.

3.1 RUL prediction

We train DiffBatt with p_{uncond} of 0.2 on different datasets and compare it with the models from BatteryML [47]. For the prediction tasks, we generate ten SOH samples from ten different input noises for each sample of the capacity matrix. Then, we select the sample that best fits the first 100 cycles as the final prediction. The RUL is further computed from the predicted SOH curve.

The results for the RUL prediction task are summarized in Table 1. This table illustrates the performance of the models in the RUL task using root-mean-squared error (RMSE) as the evaluation metric. RMSE is suitable for the RUL task since it represents the error based on an average number of cycles in which the predicted RUL differs from the reference. We train the DiffBatt model with ten different initialization seeds for each test and report the mean error along with the standard deviation, indicated as a subscript. DiffBatt shows notable performance improvements on multiple datasets. Specifically, DiffBatt achieved the lowest RMSE on the `MATR1`, `SNL`, and `CRUSH` datasets, with RMSE values of 88 ± 4, 125 ± 11, and 294 ± 18 respectively. These results highlight DiffBatt's robustness and precision in predicting RUL across different battery compositions and operational conditions. For instance, in the `SNL` dataset, DiffBatt's RMSE of 125 ± 11 outperforms the best benchmark model, PCR, which has an RMSE of 200 and is significantly lower than that of the CNN model, 924 ± 267, indicating a substantial improvement in predictive accuracy.

Furthermore, the comparative analysis shows that DiffBatt consistently performs better than other advanced models across various datasets. On the `MIX` dataset, DiffBatt achieved a mean RMSE of 202 ± 6, which is slightly higher than the best-performing model's RMSE of 197 but significantly lower than those of the deep learning models. Although DiffBatt did not achieve the lowest RMSE on the `MATR2` dataset (235 ± 16), it remains competitive compared to other deep learning models. Importantly, DiffBatt exhibits a mean RMSE of 196 across all datasets, outperforming all other models and demonstrating superior generalizability. These results illustrate DiffBatt's efficacy in learning and generalizing from diverse data sources. By utilizing the data splits and benchmarking against results from BatteryML [47], we ensure that our comparisons are both fair and indicative of DiffBatt's capabilities.

In Fig. 3, we present the results for the RUL task for each test sample from the `MIX` dataset, obtained using the DiffBatt model. Additionally, the figure includes SOH curves that correspond to the predictions with the lowest and highest uncertainty. DiffBatt is capable of quantifying the uncertainty in its predictions, which is reported here as the standard deviation of the RUL, computed from ten generated samples. The data reveals that samples with higher prediction errors generally tend to exhibit larger deviations in RUL.

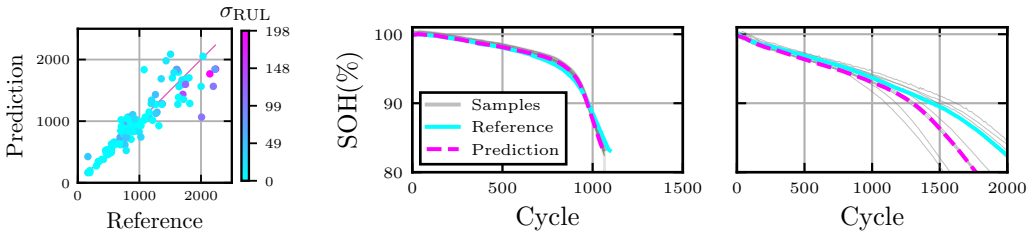
3.2 SOH estimation

In practical applications, estimating a battery's SOH requires predicting the current discharge capacity under standardized conditions using reference performance tests (RPTs) and historical cycling data. However, the discrepancy between real-world battery usage and these standardized conditions poses significant challenges in obtaining precise ground-truth labels, complicating accurate SOH estimation. Consequently, developing a robust benchmark test for SOH estimation remains an

Table 1: Results obtained from DiffBatt for RUL prediction against benchmark results reported in [47]. For models sensitive to initialization, values in the table correspond to the mean error and the standard deviation, mean_{std}, across ten seeds. Values that are underlined represent the best results from the benchmarks in BatteryML, while values that are **bold** indicate the overall best results.

Models	MATR1	MATR2	HUST	SNL	CLO	CRUH	CRUSH	MIX	Mean
"Variance" model	136	211	398	360	179	118	506	521	304
"Discharge" model	329	**149**	**322**	267	143	76	>1000	>1000	411
"Full" model	167	>1000	335	433	**138**	93	>1000	331	437
Ridge regression	116	184	>1000	242	169	65	>1000	372	268
PCR	90	187	435	200	197	68	560	376	264
PLSR	104	181	431	242	176	**60**	535	383	264
Gaussian process	154	224	>1000	251	204	115	>1000	573	440
XGBoost	334	799	395	547	215	119	330	205	368
Random forest	168_{9}	233_{7}	368_{7}	532_{25}	192_{2}	81_{1}	416_{5}	$\mathbf{197}_{0}$	273
MLP	149_{3}	275_{27}	459_{9}	370_{81}	146_{5}	103_{4}	565_{9}	451_{42}	263
CNN	102_{94}	228_{104}	465_{75}	924_{267}	>1000	174_{92}	545_{11}	272_{101}	464
LSTM	119_{11}	219_{33}	443_{29}	539_{40}	222_{12}	105_{10}	519_{39}	268_{9}	304
Transformer	135_{13}	364_{25}	391_{11}	424_{23}	187_{14}	81_{8}	550_{21}	271_{16}	300
DiffBatt (ours)	$\mathbf{88}_{4}$	235_{16}	368_{23}	$\mathbf{125}_{11}$	140_{14}	119_{6}	$\mathbf{294}_{18}$	202_{6}	**196**

ongoing effort within the research community. Therefore, we benchmark our SOH estimation results by approximating the SOH under actual operating conditions. Results are reported in Table 2. We employ the same datasets in our RUL prediction tasks for the SOH estimation experiments, which allows for reproducibility due to the clear data splits. For each degradation curve, we calculate the error as the RMSE between the reference and predicted curves up to an SOH representing the end of life (EOL). The reference SOH is zero-padded to match the length of the predicted curve if necessary. Results are reported as the mean RMSE across test samples for different EOLs. To ensure reproducibility, a detailed discussion on experimental setup and error metrics is included in Appendix A.2.



(a) RUL prediction

(b) SOH prediction

Figure 3: Results obtained from DiffBatt for RUL prediction and SOH estimation on the MIX test datasets. (a) Predicted RUL against the reference, colored by the standard deviation σ_{RUL}. (b) Generated samples and selected predictions based on the best fit to the first 100 cycles, compared against the reference SOH for test samples with the lowest (left) and highest (right) uncertainty in the predictions.

DiffBatt demonstrates strong performance in SOH estimation, achieving low RMSE values across various datasets. When considering an EOL of 80%, our model exhibits the highest precision with the CRUH dataset, yielding the lowest RMSE of 1.26 ± 0.04, and maintains consistent accuracy with RMSE values of 1.68 ± 0.08 and 1.89 ± 0.03 on the MATR1 and SNL datasets, respectively. DiffBatt has a higher RMSE with the HUST dataset at 2.75 ± 0.15, and results in an RMSE of 2.17 ± 0.07 with the MATR2 dataset. The MIX dataset, aggregating diverse data sources, results in an RMSE of 1.98 ± 0.03, indicating the model's ability to generalize effectively. Consistent RMSE values across the CLO (2.32 ± 0.07) and CRUSH (2.35 ± 0.05) datasets further highlight DiffBatt's robustness across different battery chemistries and operating conditions. To account for the varying EOL requirements

across different battery applications, such as electric vehicles and storage systems, we also evaluated DiffBatt's performance for EOL values of 90%, 70%, and 60%. At an EOL of 90%, DiffBatt maintains strong precision with low RMSE values across all datasets. Despite the increased difficulty, DiffBatt demonstrates robust performance with a minor increase in RMSE values across different EOL thresholds, resulting in mean RMSE values of 1.19, 2.05, 2.59, and 3.16 for EOLs of 90%, 80%, 70%, and 60%, respectively. These results underscore DiffBatt's adaptability for reliable SOH estimations tailored to specific battery application requirements.

Table 2: Results obtained from DiffBatt for SOH estimation task considering different end of life (EOL) values. Reported results correspond to the mean RMSE and the standard deviation, mean_{std}, across ten seeds.

EOL	MATR1	MATR2	HUST	SNL	CLO	CRUH	CRUSH	MIX	Mean
90%	$1.00_{0.07}$	$1.36_{0.07}$	$1.46_{0.1}$	$0.98_{0.05}$	$1.47_{0.06}$	$0.68_{0.02}$	$1.39_{0.03}$	$1.21_{0.03}$	1.19
80%	$1.68_{0.08}$	$2.17_{0.07}$	$2.75_{0.15}$	$1.89_{0.03}$	$2.32_{0.07}$	$1.26_{0.04}$	$2.35_{0.05}$	$1.98_{0.03}$	2.05
70%	$1.86_{0.08}$	$2.83_{0.07}$	$3.13_{0.16}$	$2.40_{0.03}$	$2.77_{0.07}$	$2.00_{0.06}$	$3.25_{0.09}$	$2.51_{0.03}$	2.59
60%	$2.23_{0.08}$	$3.97_{0.07}$	$3.43_{0.16}$	$2.95_{0.04}$	$3.16_{0.09}$	$2.43_{0.07}$	$3.97_{0.12}$	$3.13_{0.02}$	3.16

3.3 Battery degradation synthesis

Conducting comprehensive battery degradation experiments in a real-world setting is often time-consuming and resource-intensive, requiring years to complete [42]. Synthesizing battery degradation data offers significant advantages by reducing the time and cost of long-term experiments and ensuring consistent, high-quality data. Our DiffBatt model can generate high-quality battery degradation curves, enabling the creation of large and diverse datasets that can enhance model training and robustness. To evaluate the effectiveness of our approach, we utilize the DiffBatt model trained on the MATR1 dataset to generate 10 synthetic SOH curves for each sample in the training set, using varying rates for classifier-free diffusion guidance. We then leverage these synthesized SOH curves to train a random forest (RF) model for predicting RUL, while introducing 1% Gaussian noise into the input capacity matrix for data augmentation. Each model is trained with ten different initialization seeds. To assess the quality and diversity of our generated samples, we use FID [19], which provides insight into the similarity between real and synthetic images, Recall [27] as the main metric for measuring diversity and the RMSE of the random forest model on the test dataset, serving as a measure of its predictive accuracy. We also report Precision [27] for completeness. Note that FID is computed using principal component analysis (PCA) to map the data into a latent space instead of the Inception model. Results are summarized in Table 3 against those reported in [47] for a random forest model trained without synthetic data. The synthetic data generated with varying guidance strengths are depicted in Fig. 6 in the appendix.

On the one hand, the consistently low range of FID scores indicates the high quality of the synthetic data. The configuration with no guidance, $w = 0$, achieves the lowest FID score of 0.405, indicating that this setting produces synthetic data most similar to the real data. However, as w increases, the FID score rises, reaching 0.413 at $w = 6.0$, indicating a slight decrease in the similarity between synthetic and real data.

On the other hand, Recall represents the diversity by measuring the fraction of the training data manifold covered by the generated samples. The highest Recall is obtained with $w = 0.0$ and, as expected, sampling with a high guidance strength reduces diversity, leading to a Recall of 0.398 for $w = 6.0$. This reduction suggests that while higher guidance strength might improve certain aspects of the synthetic data, it compromises the overall diversity.

The RMSE values provide insight into the predictive performance of the random forest model. The lowest RMSE of 104 is achieved with $w = 2.0$, although the FID score is not the lowest in this case. This indicates that while the synthetic data generated at higher w values diverges more from the real data (as per FID), it still offers a good balance between quality and diversity to enhance the model's predictive accuracy. Conversely, the RMSE for the random forest model trained on the real data reported in [47] was 168, significantly higher than the RMSE values observed for models trained on synthetic data.

Table 3: Results for the MATR1 dataset obtained using random forests trained on synthetic data generated with different diffusion guidance w, against results reported in [47] obtained from a random forest trained without synthetic data.

		w				
Models	RF [47]	0.0	1.0	2.0	4.0	6.0
FID (↓)	NA	**0.405**	0.408	0.409	0.411	0.413
Precision (↑)	NA	**0.998**	0.99	0.985	0.963	0.968
Recall (↑)	NA	**0.663**	0.663	0.614	0.446	0.398
RMSE	168_9	109_6	107_6	$\mathbf{104}_6$	105_7	106_8

4 Towards a foundation model for battery degradation

A foundation model is a large-scale, general-purpose artificial intelligence model pre-trained on extensive datasets, designed to be fine-tuned for various specific tasks with minimal additional training [5]. By training on several battery datasets and demonstrating strong generalizability and robustness across various tasks, DiffBatt offers a promising pathway toward creating such a foundation model for battery degradation. Inspired by the principles outlined in [5], we explore how DiffBatt can be developed into a comprehensive foundation model for battery degradation.

Expressivity. DiffBatt is trained on the largest publicly available battery degradation datasets, encompassing a wide range of battery chemistries and operational conditions. This comprehensive training equips DiffBatt with a robust understanding of battery degradation patterns, making it an ideal candidate for foundational modeling. Our results demonstrate DiffBatt's high *expressivity* and its ability to capture the complex dynamics involved in battery degradation.

Multimodality. In DiffBatt, the input conditions are encoded using a transformer encoder that can be fine-tuned with relatively small amounts of new data from different cell chemistries. This ensures the model's quick adaptation to new battery technologies with minimal retraining. Furthermore, additional data modalities such as temperature, current profiles, and environmental conditions can be encoded and added to the condition vector, enhancing *multimodality*. This multimodal approach ensures all relevant factors influencing battery health are considered, improving the model's accuracy and adaptability.

Scalability and memory. DiffBatt's flexible architecture, leveraging diffusion models and transformers, ensures efficient *scalability* and *memory* usage. Its design allows it to handle extensive volumes of data and complex battery degradation scenarios effectively, making it capable of scaling with increasing data inputs without compromising performance. Additionally, DiffBatt's ability to synthesize battery degradation curves furthers its versatility by enabling the creation of large, diverse datasets. By generating synthetic degradation data, DiffBatt can augment existing datasets from cell chemistries with limited samples, enhancing the generalization capabilities of downstream models.

Compositionality. DiffBatt exemplifies compositionality through its probabilistic model structure, which integrates conditional/unconditional diffusion models with transformer encoders. This architectural composition allows DiffBatt to leverage the strengths of both approaches, enabling flexible adaptation to a wide range of battery degradation tasks. The diffusion models handle probabilistic data generation, while the transformer encoders effectively capture and encode various input conditions. This compositional framework facilitates the transfer of learned parameters to new models, ensuring the efficient application of foundational knowledge to novel contexts and tasks.

By leveraging pre-training on diverse datasets and advanced architectural frameworks, DiffBatt provides a versatile and adaptable solution, capable of addressing a wide range of battery health prediction tasks. Our model can facilitate more reliable, efficient, and scalable battery technology advancements, ultimately contributing to the development of longer-lasting and more efficient energy storage solutions.

5 Conclusions and outlook

Tackling battery degradation is a major hurdle in advancing green technologies and sustainable energy solutions. Accurately predicting battery capacity loss remains particularly challenging due to its intricate and complex nature. To address this issue, we present DiffBatt, a novel general-purpose model for predicting and synthesizing battery degradation patterns based on diffusion models with classifier-free guidance and transformer encoders. A key innovation is the integration of conditional and unconditional diffusion models, enabling the robust generation of high-quality degradation curves. DiffBatt functions as both a probabilistic model to capture the inherent uncertainties in aging processes and a generative model to simulate and predict battery degradation over time.

We evaluate the performance of DiffBatt across three different tasks, i.e., RUL prediction, SOH estimation, and SOH synthesis. In the RUL prediction task, DiffBatt achieved the lowest RMSE on the `MATR1`, `SNL`, and `CRUSH` datasets, with RMSE values of 88 ± 4, 125 ± 11, and 294 ± 18 respectively. Notably, DiffBatt results in a mean RMSE of 196 across all datasets, significantly outperforming all other competing models. These results illustrate DiffBatt's efficacy in learning from and generalizing across diverse data sources. Consistently low RMSE values in the SOH prediction task further highlight DiffBatt's robustness and reliability across different battery chemistries. Moreover, we showcase the broad applicability of DiffBatt to generate high-quality battery degradation curves. We show that augmenting battery datasets with synthetic data can lead to a better and more accurate performance of downstream ML models, e.g., for RUL prediction.

We believe that by training on several diverse battery datasets and demonstrating strong generalizability and robustness across various tasks, DiffBatt offers a promising pathway toward developing a foundational model for battery degradation. However, to support a deep understanding of degradation mechanisms and derive counter measures in battery design and or battery production as well as formation the data-driven DiffBatt can be linked to physical-based models or needs to be extended with regard to the variation of battery design and process parameters.

Acknowledgement

This research was conducted within the Research Training Group CircularLIB, supported by the Ministry of Science and Culture of Lower Saxony with funds from the program zukunft.niedersachsen of the Volkswagen Foundation (MWK | ZN3678). We would like to thank the anonymous reviewers who contributed to the improvement of our manuscript with their valuable suggestions and comments. We acknowledge that Fig. 1 is adapted and modified from the work by Fürrutter et al. [14], available on arXiv [15] under a CC BY-SA 4.0 license. The modifications include context-specific changes.

References

[1] P. M. Attia, K. A. Severson, and J. D. Witmer. Statistical learning for accurate and interpretable battery lifetime prediction. *Journal of The Electrochemical Society*, 168(9):090547, 2021. doi: 10.1149/1945-7111/ac2704. URL https://dx.doi.org/10.1149/1945-7111/ac2704.

[2] J.-H. Bastek, W. Sun, and D. M. Kochmann. Physics-informed diffusion models. *arXiv preprint arXiv:2403.14404*, 2024. URL https://arxiv.org/abs/2403.14404.

[3] S. Blömeke, C. Scheller, F. Cerdas, C. Thies, R. Hachenberger, M. Gonter, C. Herrmann, and T. S. Spengler. Material and energy flow analysis for environmental and economic impact assessment of industrial recycling routes for lithium-ion traction batteries. *Journal of Cleaner Production*, 377:134344, 12 2022. ISSN 09596526. doi: 10.1016/j.jclepro.2022.134344. URL https://linkinghub.elsevier.com/retrieve/pii/S0959652622039166.

[4] G. E. Blomgren. The development and future of lithium ion batteries. *Journal of The Electrochemical Society*, 164(1):A5019–A5025, 12 2017. ISSN 0013-4651. doi: 10.1149/2.0251701jes. URL https://iopscience.iop.org/article/10.1149/2.0251701jes.

[5] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, E. Brynjolfsson, S. Buch, D. Card, R. Castellon, N. Chatterji, A. Chen, K. Creel, J. Q. Davis, D. Demszky, C. Donahue, M. Doumbouya, E. Durmus, S. Ermon, J. Etchemendy, K. Ethayarajh, L. Fei-Fei, C. Finn, T. Gale, L. Gillespie, K. Goel, N. Goodman, S. Grossman, N. Guha, T. Hashimoto, P. Henderson, J. Hewitt, D. E. Ho, J. Hong, K. Hsu, J. Huang, T. Icard, S. Jain, D. Jurafsky, P. Kalluri, S. Karamcheti, G. Keeling, F. Khani, O. Khattab, P. W. Koh, M. Krass, R. Krishna, R. Kuditipudi, A. Kumar,

F. Ladhak, M. Lee, T. Lee, J. Leskovec, I. Levent, X. L. Li, X. Li, T. Ma, A. Malik, C. D. Manning, S. Mirchandani, E. Mitchell, Z. Munyikwa, S. Nair, A. Narayan, D. Narayanan, B. Newman, A. Nie, J. C. Niebles, H. Nilforoshan, J. Nyarko, G. Ogut, L. Orr, I. Papadimitriou, J. S. Park, C. Piech, E. Portelance, C. Potts, A. Raghunathan, R. Reich, H. Ren, F. Rong, Y. Roohani, C. Ruiz, J. Ryan, C. Ré, D. Sadigh, S. Sagawa, K. Santhanam, A. Shih, K. Srinivasan, A. Tamkin, R. Taori, A. W. Thomas, F. Tramèr, R. E. Wang, W. Wang, B. Wu, J. Wu, Y. Wu, S. M. Xie, M. Yasunaga, J. You, M. Zaharia, M. Zhang, T. Zhang, X. Zhang, Y. Zhang, L. Zheng, K. Zhou, and P. Liang. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2022. URL https://arxiv.org/abs/2108.07258.

[6] A. Brock, J. Donahue, and K. Simonyan. Large scale GAN training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2019. URL https://arxiv.org/abs/1809.11096.

[7] M. Broussely, P. Biensan, F. Bonhomme, P. Blanchard, S. Herreyre, K. Nechev, and R. Staniewicz. Main aging mechanisms in Li-ion batteries. *Journal of Power Sources*, 146(1-2):90–96, 8 2005. ISSN 03787753. doi: 10.1016/j.jpowsour.2005.03.172. URL https://linkinghub.elsevier.com/retrieve/pii/S0378775305005082.

[8] P. V. Chombo and Y. Laoonual. A review of safety strategies of a Li-ion battery. *Journal of Power Sources*, 478:228649, 12 2020. ISSN 03787753. doi: 10.1016/j.jpowsour.2020.228649. URL https://linkinghub.elsevier.com/retrieve/pii/S0378775320309538.

[9] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10850–10869, 2023. doi: 10.1109/TPAMI.2023.3261988. URL https://ieeexplore.ieee.org/document/10081412.

[10] Z. Cui, L. Wang, Q. Li, and K. Wang. A comprehensive review on the state of charge estimation for lithium-ion battery based on neural network. *International Journal of Energy Research*, 46(5):5423–5440, 4 2022. ISSN 0363-907X. doi: 10.1002/er.7545. URL https://onlinelibrary.wiley.com/doi/10.1002/er.7545.

[11] A. Devie, G. Baure, and M. Dubarry. Intrinsic variability in the degradation of a batch of commercial 18650 lithium-ion cells. *Energies*, 11(5), 2018. ISSN 1996-1073. doi: 10.3390/en11051031. URL https://www.mdpi.com/1996-1073/11/5/1031.

[12] P. Dhariwal and A. Q. Nichol. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=AAWuCvzaVt.

[13] J. S. Edge, S. O'Kane, R. Prosser, N. D. Kirkaldy, A. N. Patel, A. Hales, A. Ghosh, W. Ai, J. Chen, J. Yang, S. Li, M.-C. Pang, L. Bravo Diaz, A. Tomaszewska, M. W. Marzook, K. N. Radhakrishnan, H. Wang, Y. Patel, B. Wu, and G. J. Offer. Lithium ion battery degradation: what you need to know. *Physical Chemistry Chemical Physics*, 23(14):8200–8221, 4 2021. ISSN 1463-9076. doi: 10.1039/D1CP00359C. URL https://xlink.rsc.org/?DOI=D1CP00359C.

[14] F. Fürrutter, G. Muñoz-Gil, and H. J. Briegel. Quantum circuit synthesis with diffusion models. *Nature Machine Intelligence*, 6(5):515–524, May 2024. ISSN 2522-5839. doi: 10.1038/s42256-024-00831-9. URL http://dx.doi.org/10.1038/s42256-024-00831-9.

[15] F. Fürrutter, G. Muñoz-Gil, and H. J. Briegel. Quantum circuit synthesis with diffusion models. *arXiv preprint arXiv:2311.02041*, 2024. URL https://arxiv.org/abs/2311.02041.

[16] M. Galeotti, L. Cinà, C. Giammanco, S. Cordiner, and A. Di Carlo. Performance analysis and SOH (state of health) evaluation of lithium polymer batteries through electrochemical impedance spectroscopy. *Energy*, 89:678–686, 9 2015. ISSN 03605442. doi: 10.1016/j.energy.2015.05.148. URL https://linkinghub.elsevier.com/retrieve/pii/S0360544215007756.

[17] R. Ginster, S. Blömeke, J. Popien, C. Scheller, F. Cerdas, C. Herrmann, and T. S. Spengler. Circular battery production in the EU: Insights from integrating life cycle assessment into system dynamics modeling on recycled content and environmental impacts. *Journal of Industrial Ecology*, 28(5):1165–1182, 10 2024. ISSN 1088-1980. doi: 10.1111/jiec.13527. URL https://onlinelibrary.wiley.com/doi/10.1111/jiec.13527.

[18] 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*, 196(23):10314–10321, 2011. doi: https://doi.org/10.1016/j.jpowsour.2011.08.040. URL https://www.sciencedirect.com/science/article/pii/S0378775311015400.

[19] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/8a1d694707eb0fefe65871369074926d-Paper.pdf.

[20] J. Ho and T. Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022. URL https://arxiv.org/abs/2207.12598.

[21] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.

[22] J. Ho, C. Saharia, W. Chan, D. J. Fleet, M. Norouzi, and T. Salimans. Cascaded diffusion models for high fidelity image generation. *Journal of Machine Learning Research*, 23(47):1–33, 2022. URL http://jmlr.org/papers/v23/21-0635.html.

[23] J. Hong, D. Lee, E.-R. Jeong, and Y. Yi. Towards the swift prediction of the remaining useful life of lithium-ion batteries with end-to-end deep learning. *Applied Energy*, 278:115646, 2020. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2020.115646. URL https://www.sciencedirect.com/science/article/pii/S0306261920311429.

[24] D. Juarez-Robles, J. A. Jeevarajan, and P. P. Mukherjee. Degradation-safety analytics in lithium-ion cells: Part I. aging under charge/discharge cycling. *Journal of The Electrochemical Society*, 167(16):160510, nov 2020. doi: 10.1149/1945-7111/abc8c0. URL https://dx.doi.org/10.1149/1945-7111/abc8c0.

[25] D. Juarez-Robles, S. Azam, J. A. Jeevarajan, and P. P. Mukherjee. Degradation-safety analytics in lithium-ion cells and modules: Part III. aging and safety of pouch format cells. *Journal of The Electrochemical Society*, 168(11):110501, nov 2021. doi: 10.1149/1945-7111/ac30af. URL https://dx.doi.org/10.1149/1945-7111/ac30af.

[26] D. P. Kingma and P. Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/d139db6a236200b21cc7f752979132d0-Paper.pdf.

[27] T. Kynkäänniemi, T. Karras, S. Laine, J. Lehtinen, and T. Aila. Improved precision and recall metric for assessing generative models. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/0234c510bc6d908b28c70ff313743079-Abstract.html.

[28] A. G. Li, A. C. West, and M. Preindl. Towards unified machine learning characterization of lithium-ion battery degradation across multiple levels: A critical review. *Applied Energy*, 316:119030, 6 2022. ISSN 03062619. doi: 10.1016/j.apenergy.2022.119030. URL https://linkinghub.elsevier.com/retrieve/pii/S0306261922004354.

[29] T. Li, L. Biferale, F. Bonaccorso, M. A. Scarpolini, and M. Buzzicotti. Synthetic Lagrangian turbulence by generative diffusion models. *Nature Machine Intelligence*, 6(4):393–403, Apr 2024. ISSN 2522-5839. doi: 10.1038/s42256-024-00810-0. URL https://doi.org/10.1038/s42256-024-00810-0.

[30] W. Li, N. Sengupta, P. A. Dechent, D. Howey, A. Annaswamy, and D. U. Sauer. One-shot battery degradation trajectory prediction with deep learning. *Journal of Power Sources*, page 230024, 2021. ISSN 0378-7753. doi: 10.1016/j.jpowsour.2021.230024. URL https://publications.rwth-aachen.de/record/820366.

[31] X. Li, D. Yu, V. Søren Byg, and S. Daniel Ioan. The development of machine learning-based remaining useful life prediction for lithium-ion batteries. *Journal of Energy Chemistry*, 82:103–121, 7 2023. ISSN 20954956. doi: 10.1016/j.jechem.2023.03.026. URL https://linkinghub.elsevier.com/retrieve/pii/S2095495623001870.

[32] C. Luo, Z. Zhang, S. Zhu, and Y. Li. State-of-health prediction of lithium-ion batteries based on diffusion model with transfer learning. *Energies*, 16(9):3815, 4 2023. ISSN 1996-1073. doi: 10.3390/en16093815. URL https://www.mdpi.com/1996-1073/16/9/3815.

[33] L. Luzi, P. M. Mayer, J. Casco-Rodriguez, A. Siahkoohi, and R. G. Baraniuk. Boomerang: Local sampling on image manifolds using diffusion models. *arXiv preprint arXiv:2210.12100*, 2024. URL https://arxiv.org/abs/2210.12100.

[34] 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*, 2022. doi: 10.1039/D2EE01676A. URL http://dx.doi.org/10.1039/D2EE01676A.

[35] S. E. J. O'Kane, W. Ai, G. Madabattula, D. Alonso-Alvarez, R. Timms, V. Sulzer, J. S. Edge, B. Wu, G. J. Offer, and M. Marinescu. Lithium-ion battery degradation: how to model it. *Physical Chemistry Chemical Physics*, 24(13):7909–7922, 3 2022. ISSN 1463-9076. doi: 10.1039/D2CP00417H. URL https://xlink.rsc.org/?DOI=D2CP00417H.

[36] Y. Preger, H. M. Barkholtz, A. Fresquez, D. L. Campbell, B. W. Juba, J. Romàn-Kustas, S. R. Ferreira, and B. Chalamala. Degradation of commercial lithium-ion cells as a function of chemistry and cycling

conditions. *Journal of The Electrochemical Society*, 167(12):120532, sep 2020. doi: 10.1149/1945-7111/abae37. URL https://dx.doi.org/10.1149/1945-7111/abae37.

[37] H. Rauf, M. Khalid, and N. Arshad. Machine learning in state of health and remaining useful life estimation: Theoretical and technological development in battery degradation modelling. *Renewable and Sustainable Energy Reviews*, 156:111903, 3 2022. ISSN 13640321. doi: 10.1016/j.rser.2021.111903. URL https://linkinghub.elsevier.com/retrieve/pii/S1364032121011692.

[38] Z. Ren and C. Du. A review of machine learning state-of-charge and state-of-health estimation algorithms for lithium-ion batteries. *Energy Reports*, 9:2993–3021, 12 2023. ISSN 23524847. doi: 10.1016/j.egyr.2023.01.108. URL https://linkinghub.elsevier.com/retrieve/pii/S235248472300118X.

[39] H. Rubenbauer and S. Henninger. Definitions and reference values for battery systems in electrical power grids. *Journal of Energy Storage*, 12:87–107, 8 2017. ISSN 2352152X. doi: 10.1016/j.est.2017.04.004. URL https://linkinghub.elsevier.com/retrieve/pii/S2352152X16301797.

[40] K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis, M. Z. Bazant, S. J. Harris, W. C. Chueh, and R. D. Braatz. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4(5):383–391, 3 2019. ISSN 2058-7546. doi: 10.1038/s41560-019-0356-8. URL https://www.nature.com/articles/s41560-019-0356-8.

[41] J. Vetter, P. Novák, M. Wagner, C. Veit, K.-C. Möller, J. Besenhard, M. Winter, M. Wohlfahrt-Mehrens, C. Vogler, and A. Hammouche. Ageing mechanisms in lithium-ion batteries. *Journal of Power Sources*, 147(1-2):269–281, 9 2005. ISSN 03787753. doi: 10.1016/j.jpowsour.2005.01.006. URL https://linkinghub.elsevier.com/retrieve/pii/S0378775305000832.

[42] V. I. voor Technologisch Onderzoek (VITO), C. à l'Energie Atomique et aux Energies Alternatives (CEA), Siemens, T. U. M. (TUM), T. S. B. Testing, ALGOLiON, R. A. University, L. Smart, T. U. Eindhoven, Voltia, and V. E. T. Solutions. Everlasting: Electric vehicle enhanced range, lifetime and safety through ingenious battery management, 2021. URL https://data.4tu.nl/collections/EVERLASTING_Electric_Vehicle_Enhanced_Range_Lifetime_And_Safety_Through_INGenious_battery_management_/5065445/11.

[43] S. Wang, S. Jin, D. Bai, Y. Fan, H. Shi, and C. Fernandez. A critical review of improved deep learning methods for the remaining useful life prediction of lithium-ion batteries. *Energy Reports*, 7:5562–5574, 2021. ISSN 23524847. doi: 10.1016/j.egyr.2021.08.182. URL files/1917/Acriticalreviewofimproveddeeplearningmethodsfortheremaining.pdfhttps://doi.org/10.1016/j.egyr.2021.08.182.

[44] Y. Xing, E. W. Ma, K.-L. Tsui, and M. Pecht. An ensemble model for predicting the remaining useful performance of lithium-ion batteries. *Microelectronics Reliability*, 53(6):811–820, 2013. doi: https://doi.org/10.1016/j.microrel.2012.12.003. URL https://www.sciencedirect.com/science/article/pii/S0026271412005227.

[45] Y. Yang, M. Jin, H. Wen, C. Zhang, Y. Liang, L. Ma, Y. Wang, C. Liu, B. Yang, Z. Xu, J. Bian, S. Pan, and Q. Wen. A survey on diffusion models for time series and spatio-temporal data. *arXiv preprint arXiv:2404.18886*, 2024. URL https://arxiv.org/abs/2404.18886.

[46] B. Zhang, P. Xu, X. Chen, and Q. Zhuang. Generative quantum machine learning via denoising diffusion probabilistic models. *Physical Review Letters*, 132:100602, Mar 2024. doi: 10.1103/PhysRevLett.132.100602. URL https://link.aps.org/doi/10.1103/PhysRevLett.132.100602.

[47] H. Zhang, X. Gui, S. Zheng, Z. Lu, Y. Li, and J. Bian. BatteryML: An open-source platform for machine learning on battery degradation. In *The Twelfth International Conference on Learning Representations*, 2024. URL http://arxiv.org/abs/2310.14714.

A Appendix

A.1 Datasets

Battery degradation curves utilized for training and testing the models are depicted in Fig. 4 for each cell chemistry. A brief summary of the datasets included in BatteryML [47] is presented here.

The CALCE dataset includes full lifecycle data from 13 batteries with an LCO cathode. Each battery has a nominal capacity of 1100 mAh. They were all charged using a constant current/constant voltage protocol: 0.5C current until reaching 4.2V, maintaining 4.2V until the current dropped below 0.05A, and a cutoff voltage of 2.7V [44, 18].

The MATR dataset, provided by Severson et al. [40] and Hong et al. [23], is one of the largest public datasets containing 180 commercial 18650 LFP batteries. These batteries, cycled at a forced convection temperature chamber of 30°C, have a nominal capacity of 1.1 Ah and a nominal voltage of 3.3V. The dataset comprises three subsets: MATR1, MATR2 [40], and CLO [23], all categorized due to distinct measurement batches.

The HUST dataset includes 77 LFP batteries, similar to those in the MATR dataset. These batteries followed an identical charging protocol with varying multi-stage discharge protocols, all conducted at a constant temperature of 30°C [34].

The HNEI dataset contains 14 commercial 18650 cells with a graphite anode and a blended NMC and LCO cathode. These cells were cycled at 1.5C to 100% depth of discharge for over 1000 cycles at room temperature [11].

The SNL dataset includes 61 commercial 18650 cells (NCA, NMC, and LFP), cycled to 80% capacity. The study evaluates the impact of temperature, depth of discharge, and discharge current on long-term degradation [36].

The UL_PUR dataset comprises 10 commercial pouch cells with a graphite negative electrode and an NCA cathode. These cells were cycled at 1C between 2.7V and 4.2V, equivalent to 0-100% state of charge (SOC), at room temperature until reaching 10-20% capacity fade. Additionally, modules were cycled at C/2 between 13.7V and 21.0V until 20% capacity fade [24, 25].

The RWTH dataset contains data from 48 lithium-ion battery cells aged under identical conditions. These cells feature a carbon anode and an NMC cathode [30]. The cells were cycled at a constant ambient temperature of 25°C. Each cycle involved a 30-minute discharge phase down to 3.5V and a 30-minute charge phase up to 3.9V, with the currents capped at a maximum of 4A. This resulted in cycles between approximately 20% and 80% state of charge.

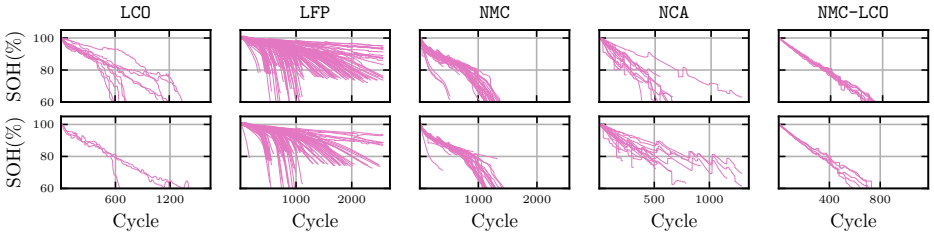


Figure 4: Train (up) and test (bottom) samples for each cell chemistry. The data is scaled using the SOH of the first cycle.

A.2 SOH prediction

For prediction tasks, i.e. we employ a guidance strength of $w = 0.0$ and generate ten samples for each input capacity matrix. The capacity matrix is constructed from the first 100 cycles. We further select the final prediction based on the best fit to the SOH of the first 100 cycles. The RMSE for an SOH sample j is computed as

$$\text{RMSE}_j = \sqrt{\frac{1}{n_j} \sum_{i=1}^{n_j} (\tilde{y}_i - y_i)^2} \tag{8}$$

where $\tilde{y}$ and y represent the predicted and the reference SOH in percentage, respectively, i denotes the cycle number and n_j is the cycle number at which the predicted SOH reaches the EOL. Further, we report the mean RMSE across all the test samples as the RMSE for the dataset.

Figure 5 illustrates the predicted SOH versus the reference SOH for all test samples in the `MIX` dataset. The results demonstrate that DiffBatt effectively captures various degradation dynamics and accurately predicts SOH for the majority of test samples and highlights DiffBatt's ability to generalize across different battery chemistries and operational conditions present in the `MIX` dataset. This capability is essential for developing reliable battery health monitoring systems that can adapt to diverse usage patterns and environmental factors.

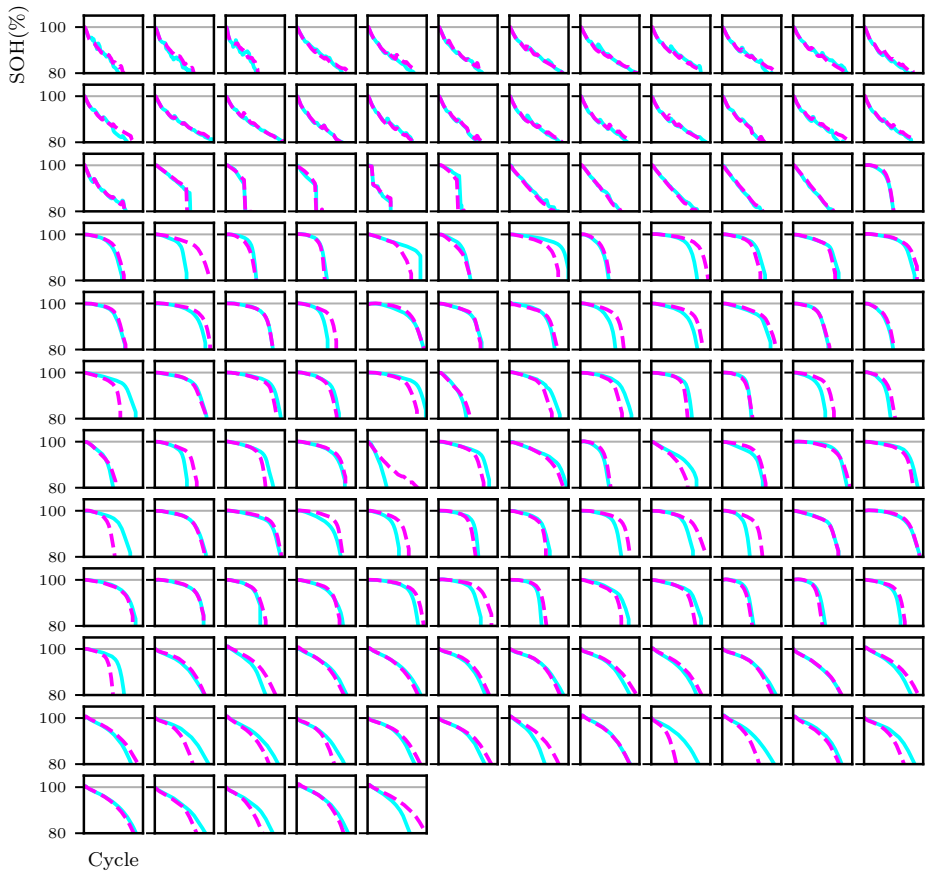


Figure 5: SOH predictions against reference for all the test samples of `MIX` dataset. The pink dashed line shows the prediction and the cyan solid line shows the reference.

A.3 SOH synthesis

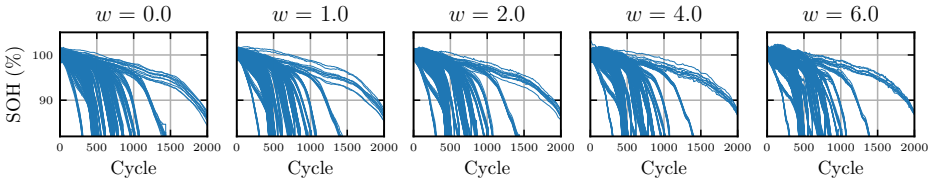


Figure 6: `MATR1` synthetic data generated by DiffBatt with different guidance strengths.