A SINGLE LEAF HERALDS AUTUMN: PREDICTING BATTERY LIFETIME UNDER STATIONARITY MASKING

Anonymous authors
Paper under double-blind review

ABSTRACT

Early battery lifetime prediction estimates long-term cycle life from limited initial cycles, reducing the time required for full degradation testing. However, early cycles are highly similar, making subtle degradation variations difficult to identify from individual cycle states. Beyond marginal distribution shifts in time-series non-stationarity, we characterize a regime in which near-stationary marginals coexist with variations in conditional transitions between adjacent cycles, which we term *stationarity masking*. To isolate these hidden transition variations from marginal effects, we disentangle marginal and conditional discrepancies and derive a unique transition projection that preserves the marginal distributions of both adjacent cycles while capturing their inter-cycle transition structure. Building on this analysis, we propose LEAF, a simple yet effective framework for early battery lifetime prediction. LEAF learns conditional associations between local states of adjacent cycles, extracts conditional transitions through fixed-marginal projection, and models their sequential evolution for lifetime prediction. Extensive experiments on 16 datasets spanning Li-ion, Zn-ion, and Na-ion batteries validate the effectiveness of LEAF across diverse battery systems and aging conditions. Code is available at `https://anonymous.4open.science/r/Batt_LEAF/`.

1 INTRODUCTION

Long-term evolution in natural systems is often preceded by subtle local variations before pronounced shifts become observable. The idiom *A Single Leaf Heralds Autumn* reflects this notion, suggesting that limited early observations can foreshadow the long-term evolution of a system. Battery lifetime prediction follows a similar principle. Researchers aim to infer the eventual cycle life from only a few initial charge–discharge cycles. This can shorten full-lifetime testing and reduce experimental costs (Tan et al., 2025). However, degradation signals remain weak during the early cycling stage. The high similarity between adjacent cycles obscures weak life-related variations and poses a key challenge to early battery lifetime prediction (Hu et al., 2026; Miao et al., 2026).

The growing demand for reliable early lifetime assessment has driven progress in data-driven battery prognostics. Early statistical learning related degradation signals in a few initial cycles to eventual lifetime, enabling prediction before pronounced capacity degradation (Severson et al., 2019; Attia et al., 2021; Fei et al., 2021; Paulson et al., 2022). With advances in deep learning, recent methods capture degradation dynamics and cross-cycle dependencies from raw cycling sequences, further exploiting weak early-stage signals (Tang et al., 2022; Fei et al., 2022; Hsu et al., 2022; Ma et al., 2022; Tao et al., 2023). Despite continued progress, many models are evaluated through dataset-specific training, while heterogeneous data and evaluation protocols limit unified comparison and cross-system generalization. BatteryML and BatteryLife established unified data-processing and evaluation frameworks for reproducible battery lifetime research (Zhang et al., 2024; Tan et al., 2025). Recent studies further focus on identifying and exploiting weak degradation signals for lifetime prediction from limited cycling observations (Tan et al., 2024; Zhang et al., 2025; Hu et al., 2026; Miao et al., 2026). In practice, battery testing unfolds continuously along the cycle axis, with observations generated sequentially by an evolving internal electrochemical state. Therefore, battery lifetime prediction is inherently a sequential modeling problem with strong temporal dependence. Probability distributions provide a unified statistical basis for characterizing such dynamics and distinguishing the probabilistic structures underlying different forms of temporal variation.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Non-stationarity is a fundamental property of time series, typically reflected by temporal distribution shifts. Existing studies mainly identify non-stationarity through changes in the marginal distribution $p(x_t)$ and develop adaptation methods (Kim et al., 2022; Du et al., 2021; Fan et al., 2023). Early battery cycles often exhibit highly similar distributions, making marginal shifts difficult to identify. Meanwhile, battery aging is driven by cumulative and irreversible electrochemical processes, including solid electrolyte interphase growth, loss of lithium inventory, and loss of active material (Edge et al., 2021; Birkl et al., 2017). Each cycle inherits the accumulated internal state from preceding cycles, rendering the previous cycle a natural conditioning variable for the next-cycle distribution. Consequently, the near invariance of $p(x_t)$ during early life does not imply an invariant $p(x_{t+1} \mid x_t)$, since the latter can vary with the evolving degradation state. We define this regime as *stationarity masking*, where near-stationary marginal distributions conceal degradation-related variations in adjacent-cycle conditional transitions, as illustrated in Figure 1. Accordingly, weak early degradation may be more directly reflected in adjacent-cycle conditional transitions.

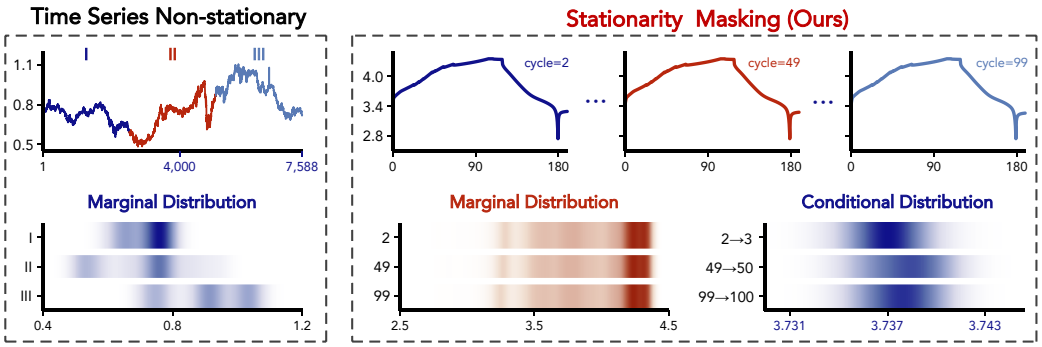


Figure 1: Comparison between conventional time-series non-stationarity and stationarity masking in early battery cycles. Left: the Exchange Rate dataset exhibits clear marginal distribution shifts across temporal regimes. Right: the CALB dataset exhibits near-stationary marginal distributions across early cycles but distinct adjacent-cycle conditional distributions. (More in Appendix A.2.)

Motivated by this observation, we first analyze stationarity masking through the decomposition of adjacent-cycle joint distributions, separating marginal differences from conditional variation under nearly invariant marginals. Building on this insight, we propose LEAF, a simple yet effective framework for early battery lifetime prediction. The central idea is to shift the representation from highly similar individual-cycle states to adjacent-cycle conditional transitions, thereby capturing local differences in cross-cycle evolution. Specifically, LEAF learns conditional associations between local states of consecutive cycles and employs fixed-marginal transition projection to preserve the state compositions of both cycles while obtaining a probabilistically consistent representation of their joint dependence. Our main contributions are summarized as follows:

- We formalize *stationarity masking* in early battery lifetime prediction and use the KL chain rule to explicitly separate marginal variation from conditional discrepancy, showing how adjacent-cycle dependence can vary under nearly invariant marginal distributions.
- We propose LEAF, a simple yet effective conditional-transition modeling framework. LEAF characterizes cross-cycle evolution through adjacent-cycle state associations and employs a fixed-marginal transition projection to preserve cycle-wise state distributions, yielding a consistent joint representation for ordered transition modeling.
- We evaluate LEAF on 16 datasets covering lithium-ion, sodium-ion, and zinc-ion batteries under diverse aging conditions. Results show consistently lower prediction errors across all four battery domains and further validate the overall effectiveness of the proposed conditional-transition representation across diverse battery systems and aging conditions.

2 PRELIMINARIES AND RELATED WORK

2.1 PROBLEM DEFINITION

Battery lifetime is typically determined through controlled cycling degradation tests. Batteries are repeatedly cycled under prescribed charge–discharge protocols in temperature-controlled environments, with each cycle comprising a complete charge and discharge process (Appendix C.1). For the c-th cycle, let $\mathbf{X}_c \in \mathbb{R}^{N\times P}$ denote the resampled multivariate observation, where N is the number of observed variables and P is the number of sampling points within one cycle. The early-cycle sequence is denoted as $\mathbf{X}_{1:C} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_C]$, where C denotes the cutoff index of the early observation window, typically $C = 100$ (Severson et al., 2019; Tan et al., 2025). Following common practice in battery prognostics, the lifetime label y is defined as the first cycle index at which the state of health reaches a prescribed end-of-life threshold. Early battery lifetime prediction learns a mapping $\hat{y} = f_\theta(\mathbf{X}_{1:C})$ from limited early-cycle observations to the eventual cycle life.

2.2 RELATED WORK

Battery lifetime prediction has recently become an active research topic, supported by the growing availability of cycling data and the unified data processing, task settings, and evaluation protocols established by BatteryML and BatteryLife (Zhang et al., 2024; Tan et al., 2025). Early studies mainly related early observations to eventual cycle life through capacity decay, cycling-curve differences, and handcrafted statistical features (Severson et al., 2019; Attia et al., 2021; Weng et al., 2021; O'Kane et al., 2022). With the growth of cycling datasets and deep learning, subsequent studies increasingly learn degradation representations directly from raw current, voltage, and capacity observations, while deep temporal architectures model within-cycle patterns and their evolution over successive cycles (Tang et al., 2022; Hsu et al., 2022; Ma et al., 2022; Fei et al., 2022; Han et al., 2023; He et al., 2024; Wang et al., 2024). Under increasingly standardized benchmarks, recent studies further strengthen representations of complex ageing processes, shifting battery lifetime prediction from dataset-specific fitting toward more stable and transferable degradation modeling (Tan et al., 2024; Eivazi et al., 2024; Zhang et al., 2025; Li et al., 2025; Hu et al., 2026; Miao et al., 2026; Huang et al., 2026; Tan et al., 2026). Despite continued progress, the high similarity of early cycles still masks lifetime-related degradation, while the underlying evolution of conditional distributions has yet to become a primary modeling target in current early-life prediction frameworks.

3 PROPOSED METHOD

3.1 MOTIVATION

In this section, we define *stationarity masking* and characterize its probabilistic structure. We use the KL chain rule to decompose adjacent-cycle joint discrepancies into marginal and conditional-transition terms (Theorem 3.1), then characterize the unique Euclidean projection onto the feasible transition set with two prescribed cycle marginals (Theorem 3.2).

Definition 3.1 (Stationarity Masking). *Let $\{X_c\}_{c=1}^{C}$ denote consecutive battery cycle observations on a common measurable space, and let $\mathcal{C}$ index early-life cycles. For two positions $c, c' \in \mathcal{C}$, stationarity masking refers to approximately equal marginal laws together with distinct adjacent-cycle conditional transitions on source states with positive probability:*

$$P(X_c) \approx P(X_{c'}), \qquad P(X_{c+1} \mid X_c) \neq P(X_{c'+1} \mid X_{c'}). \tag{1}$$

To distinguish cycle-wise marginals from cross-cycle transitions, we denote the marginal distribution at cycle c by P_c and the conditional transition from cycle c to cycle $c+1$ by K_c. The joint distribution of two adjacent cycles at position c is then written as:

$$J_c(X_c, X_{c+1}) = P_c(X_c)K_c(X_{c+1} \mid X_c). \tag{2}$$

Theorem 3.1 (Conditional Transition Discrepancy). *For two cycle positions c and c', assume that J_c is absolutely continuous with respect to $J_{c'}$ and the displayed divergences are finite. The KL chain rule decomposes their joint discrepancy into a marginal term and an expected conditional term, separating variation in cycle states from variation in cross-cycle dependence:*

$$D_{\mathrm{KL}}(J_c \| J_{c'}) = D_{\mathrm{KL}}(P_c \| P_{c'}) + \mathbb{E}_{x\sim P_c}\left[D_{\mathrm{KL}}\left(K_c(\cdot \mid x) \| K_{c'}(\cdot \mid x)\right)\right]. \tag{3}$$

where D_{KL} denotes the KL divergence and $\mathbb{E}_{x\sim P_c}$ denotes the expectation over P_c.

Theorem 3.1 shows that the conditional term accounts exactly for the joint discrepancy beyond its marginal component. With identical marginals, the marginal term vanishes, and differing conditionals on positive-probability states yield a positive joint divergence. Under near-invariant marginals, the same decomposition quantifies their respective contributions. The proof and further theoretical analysis are provided in Appendix A.4. We further consider joint transitions under fixed marginal constraints. Let μ and ν denote the probability vectors of the local states of two adjacent cycles. The feasible transition set satisfying both marginal constraints is defined as:

$$\Pi(\mu,\nu)=\left\{Q\geq 0\;\middle|\;\sum_j Q_{ij}=\mu_i,\ \sum_i Q_{ij}=\nu_j\right\}. \tag{4}$$

Theorem 3.2 (Transition Projection under Fixed Marginals). *If the feasible set $\Pi(\mu,\nu)$ is nonempty, then for any reference association matrix G, there exists a unique feasible transition with minimum Euclidean deviation from G among all transitions satisfying the same marginal constraints, defined as:*

$$Q^\star=\arg\min_{Q\in\Pi(\mu,\nu)}\frac{1}{2}\|Q-G\|_F^2. \tag{5}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Equivalently, $Q^\star$ is the unique Euclidean projection of G onto the feasible set $\Pi(\mu,\nu)$ and satisfies:

$$\langle G-Q^\star, Q-Q^\star\rangle_F\leq 0,\qquad \forall Q\in\Pi(\mu,\nu). \tag{6}$$

where $\langle\cdot,\cdot\rangle_F$ denotes the Frobenius inner product.

Theorem 3.2 gives the unique joint transition closest to the association G while preserving both prescribed marginals. Its residual from the independence baseline has zero row and column sums and characterizes the cross-cycle dependence retained under the same state compositions. Fixed-marginal projection therefore preserves conditional association structure without altering the marginal state distributions of either cycle. The complete proof and fixed-marginal dependence characterization are provided in Appendix A.4.

3.2 Model Architecture

We propose LEAF, a battery lifetime prediction model that captures early degradation under stationarity masking through adjacent-cycle conditional transitions. The overall architecture of LEAF is illustrated in Figure 2. LEAF first divides each cycle into consecutive patches and employs a shared linear patch encoder to obtain local cycle states. Associations between adjacent-cycle states are then normalized into conditional transitions and combined with cycle-wise state distributions to construct joint transitions. To preserve the marginal distributions of both adjacent cycles, LEAF projects each joint transition onto a fixed-marginal feasible set and recovers the corresponding conditional transition from the projected joint distribution. Finally, the ordered conditional transitions are modeled by a lightweight transition encoder and fed into a prediction head for battery lifetime prediction.

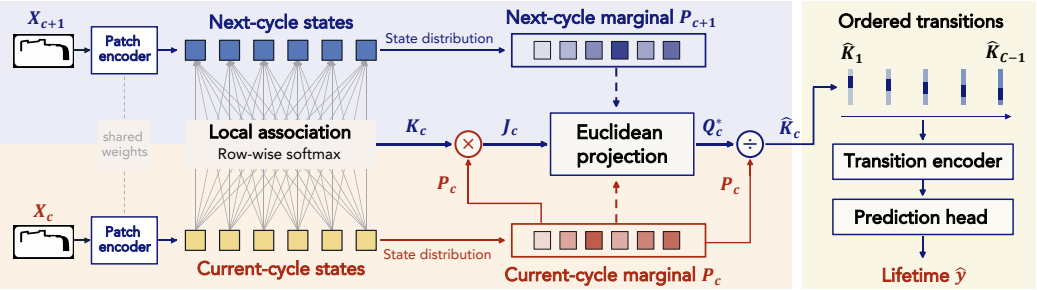


Figure 2: The overall architecture of LEAF.

3.2.1 ADJACENT-CYCLE CONDITIONAL TRANSITION

Given the first C early cycles $\{\mathbf{X}_c\}_{c=1}^{C}$, each cycle contains observations ordered along the charge–discharge process. We divide each cycle into K consecutive patches and employ a shared linear patch encoder to map each patch into a local state $\mathbf{h}_{c,i}$, where $i \in \{1, \ldots, K\}$ denotes its position within the cycle. The current and next cycles share the same encoder parameters, placing their local states in a common feature space. For state i in cycle c and state j in cycle $c+1$, the local association is defined as:

$$A_c(i, j) = s_\theta \left(\mathbf{h}_{c,i}, \mathbf{h}_{c+1,j}\right). \tag{7}$$

where $s_\theta(\cdot, \cdot)$ denotes the learnable local association function. The matrix $A_c \in \mathbb{R}^{K \times K}$ scores correspondences between the ordered local states of two adjacent cycles, allowing each source patch to associate with multiple candidate target patches.

We subsequently apply row-wise normalization over candidate states in the next cycle to obtain the adjacent-cycle conditional transition:

$$K_c(j \mid i) = \frac{\exp A_c(i, j)}{\sum_{j'=1}^{K} \exp A_c(i, j')}. \tag{8}$$

Here, $K_c(j \mid i)$ assigns a normalized association weight to target patch j given source patch i. The matrix defines a conditional distribution over target patch indices in the learned representation. Modeling its ordered evolution across adjacent cycles makes changes in local cross-cycle correspondence available to the lifetime predictor, together with the cycle-wise state compositions.

Meanwhile, the current and next cycles produce normalized patch weights P_c and P_{c+1}, respectively, describing local state composition in the learned representation. Combining the source marginal P_c with the conditional transition K_c defines the joint law over source and target patch indices:

$$J_c(i, j) = P_c(i) K_c(j \mid i). \tag{9}$$

Here, $P_c(i)$ specifies the source patch mass, while $K_c(j \mid i)$ gives its conditional assignment to target patch j in the next cycle. Thus, J_c is a joint probability matrix over source and target patch indices and serves as the joint transition representation for the subsequent fixed-marginal projection.

3.2.2 FIXED-MARGINAL TRANSITION PROJECTION

The joint transition J_c preserves the current-cycle marginal P_c, whereas its next-cycle marginal is not necessarily consistent with the state distribution P_{c+1} obtained directly from $\mathbf{X}_{c+1}$. To preserve the state compositions of both adjacent cycles, we constrain the joint transition within the fixed-marginal feasible set:

$$\Pi(P_c, P_{c+1}) = \left\{ Q \geq 0 \;\middle|\; \sum_{j=1}^{K} Q_{ij} = P_c(i),\ \sum_{i=1}^{K} Q_{ij} = P_{c+1}(j) \right\}. \tag{10}$$

This constraint fixes the current-cycle and next-cycle marginals on the two sides of the transition, allowing the correspondence between local states to vary without altering the state composition of either cycle. Based on Theorem 3.2, we project J_c onto $\Pi(P_c, P_{c+1})$ through Euclidean projection and obtain the fixed-marginal joint transition:

$$Q_c^\star = \underset{Q \in \Pi(P_c, P_{c+1})}{\arg\min} \frac{1}{2} \left\| Q - J_c \right\|_F^2. \tag{11}$$

The resulting $Q_c^\star$ satisfies both the current-cycle marginal P_c and the next-cycle marginal P_{c+1} while minimizing the Euclidean adjustment to J_c. Its deviation from the independence baseline $P_c P_{c+1}^\top$ has zero row and column sums, representing local cross-cycle dependence compatible with the prescribed state compositions. For source states with $P_c(i) > 0$, the corresponding corrected conditional transition is recovered as:

$$\widehat{K}_c(j \mid i) = \frac{Q_c^\star(i, j)}{P_c(i)}. \tag{12}$$

where $\widehat{K}_c(j \mid i)$ denotes the conditional transition after fixed-marginal projection. Repeating this over C cycles yields the ordered transition sequence $\{\widehat{K}_c\}_{c=1}^{C-1}$. LEAF uses a one-dimensional convolution with kernel size 5 as the Transition Encoder to capture local evolution along the ordered transition sequence. The encoded transition features are then fed into the Prediction Head to produce the final lifetime prediction $\hat{y}$. The complete algorithm is provided in Appendix B.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

We adopt the unified evaluation setting used in recent battery lifetime prediction studies (Tan et al., 2025), covering battery chemistries, operating protocols, and testing environments.

Datasets. We evaluate LEAF on 16 battery datasets organized into four domains. (1) **Li-ion** contains 13 public datasets covering diverse chemistries, temperatures, and charge and discharge protocols (Xing et al., 2013; Devie et al., 2018; Severson et al., 2019; Juarez-Robles et al., 2020; Preger et al., 2020; Mohtat et al., 2021; Weng et al., 2021; Li et al., 2021; Ma et al., 2022; Zhu et al., 2022; Cui et al., 2024; Wang et al., 2024; Li et al., 2024). (2) **CALB** contains industrially tested large-format commercial Li-ion batteries (Tan et al., 2025). (3) **Na-ion** contains commercial sodium-ion batteries under diverse charge and discharge protocols (Tan et al., 2025). (4) **Zn-ion** contains laboratory-tested zinc-ion batteries with varied electrolytes, cell formats, and temperatures (Tan et al., 2025). Detailed dataset statistics and data processing details are provided in Appendix C.2.

Evaluation Metrics. Following prior battery lifetime prediction studies (Tan et al., 2025; Miao et al., 2026), lifetime prediction uses mean absolute percentage error (MAPE) and 15%-Acc. The former measures relative error between predicted and ground-truth lifetimes, and the latter denotes the proportion of test prefix samples with errors within 15%. Formulas are in Appendix C.3.

Baselines. We compare LEAF with eight representative methods, including iTransformer (Liu et al., 2024), DiffBatt (Eivazi et al., 2024), BatLiNet (Zhang et al., 2025), LiPM (Li et al., 2025), CPMLP, CPTransformer, CPGRU (Tan et al., 2025), and DITING (Miao et al., 2026). These baselines span general time-series modeling, generative modeling, inter-cell learning, pretraining, and cross-cycle architectures, providing broad comparisons for LEAF. Details are provided in Appendix C.4.

Training Details. All models are trained with Adam using mean squared error as the objective. Model configurations are selected by validation MAPE. Each experiment is repeated three times, and the mean and standard deviation are reported. Further details can be found in Appendix C.5.

4.2 OVERALL PERFORMANCE

As shown in Table 1, LEAF achieves the lowest mean MAPE and highest mean 15%-ACC across all four domains. Relative to the strongest baseline in each domain, MAPE decreases by 13.0%, 18.7%, 19.5%, and 16.3%, respectively. These comparisons support the predictive value of adjacent-cycle conditional representations across the different battery systems and aging conditions.

Table 1: Overall battery lifetime prediction performance. The best results are highlighted in red and bold, while the second-best results are highlighted in blue and underlined.

	Li-ion		CALB		Na-ion		Zn-ion	
Method	MAPE	15%-ACC	MAPE	15%-ACC	MAPE	15%-ACC	MAPE	15%-ACC
iTransformer	$0.245\pm.011$	$0.401_{\pm.019}$	$0.179_{\pm.026}$	$0.625_{\pm.077}$	$0.432_{\pm.056}$	$0.219_{\pm.080}$	$0.969_{\pm.051}$	$0.297_{\pm.069}$
DiffBatt	$0.226_{\pm.007}$	$0.401_{\pm.011}$	$0.211_{\pm.025}$	$0.496_{\pm.135}$	$0.405_{\pm.039}$	$0.304_{\pm.073}$	$0.623_{\pm.124}$	$0.197_{\pm.085}$
BatLiNet	$0.190_{\pm.005}$	$0.452_{\pm.015}$	$0.205_{\pm.017}$	$0.482_{\pm.074}$	$0.396_{\pm.050}$	$0.290_{\pm.088}$	$0.578_{\pm.089}$	$0.234_{\pm.091}$
LiPM	$0.183_{\pm.004}$	$0.586_{\pm.009}$	$0.179_{\pm.010}$	$0.559_{\pm.088}$	$0.313_{\pm.041}$	$0.325_{\pm.086}$	$0.596_{\pm.027}$	$0.217_{\pm.066}$
CPGRU	$0.180_{\pm.007}$	$0.573_{\pm.011}$	$\underline{0.139}_{\pm.015}$	$\underline{0.690}_{\pm.125}$	$0.296_{\pm.054}$	$0.245_{\pm.097}$	$0.578_{\pm.034}$	$0.271_{\pm.085}$
CPTransformer	$0.188_{\pm.005}$	$0.559_{\pm.012}$	$0.153_{\pm.004}$	$0.671_{\pm.123}$	$\underline{0.251}_{\pm.039}$	$\underline{0.394}_{\pm.078}$	$0.557_{\pm.038}$	$0.302_{\pm.059}$
CPMLP	$0.175_{\pm.004}$	$0.627_{\pm.006}$	$0.142_{\pm.011}$	$0.687_{\pm.049}$	$0.261_{\pm.048}$	$0.355_{\pm.041}$	$\underline{0.523}_{\pm.046}$	$\underline{0.311}_{\pm.060}$
DITING	$\underline{0.169}_{\pm.003}$	$\underline{0.643}_{\pm.007}$	$0.148_{\pm.014}$	$0.652_{\pm.060}$	$0.289_{\pm.034}$	$0.364_{\pm.053}$	$0.595_{\pm.077}$	$0.274_{\pm.082}$
LEAF	$\mathbf{0.147}_{\pm.005}$	$\mathbf{0.694}_{\pm.007}$	$\mathbf{0.113}_{\pm.008}$	$\mathbf{0.778}_{\pm.042}$	$\mathbf{0.202}_{\pm.039}$	$\mathbf{0.444}_{\pm.045}$	$\mathbf{0.438}_{\pm.031}$	$\mathbf{0.351}_{\pm.063}$

4.3 ABLATION STUDY

To systematically assess the contribution of each key design in LEAF, we construct four representative ablation variants. **w/o CT** removes adjacent-cycle conditional transitions and uses only cycle-wise marginal states for lifetime prediction, evaluating the contribution of the conditional-transition representation. **w/o FP** removes fixed-marginal projection and directly uses the initial conditional

transitions K_c, examining the effect of marginal constraints. **w/o TE** removes the transition encoder and replaces it with simple aggregation, evaluating the contribution of local cross-cycle evolution modeling. **w/o TO** randomly permutes the transition sequence while preserving all transition observations, examining the predictive role of ordered transition evolution.

Table 2: Ablation study of LEAF across four battery domains.

	Li-ion		CALB		Na-ion		Zn-ion	
Method	MAPE	15%-ACC	MAPE	15%-ACC	MAPE	15%-ACC	MAPE	15%-ACC
LEAF	$\mathbf{0.147}_{\pm.005}$	$\mathbf{0.694}_{\pm.007}$	$\mathbf{0.113}_{\pm.008}$	$\mathbf{0.778}_{\pm.042}$	$\mathbf{0.202}_{\pm.039}$	$\mathbf{0.444}_{\pm.045}$	$\mathbf{0.438}_{\pm.031}$	$\mathbf{0.351}_{\pm.063}$
w/o CT	$0.184_{\pm.009}$	$0.598_{\pm.018}$	$0.139_{\pm.014}$	$0.681_{\pm.063}$	$0.251_{\pm.051}$	$0.351_{\pm.067}$	$0.497_{\pm.058}$	$0.306_{\pm.077}$
w/o FP	$0.162_{\pm.006}$	$0.655_{\pm.012}$	$0.128_{\pm.010}$	$0.717_{\pm.049}$	$0.267_{\pm.046}$	$0.329_{\pm.074}$	$0.472_{\pm.041}$	$0.331_{\pm.071}$
w/o TE	$0.171_{\pm.011}$	$0.641_{\pm.015}$	$0.145_{\pm.018}$	$0.664_{\pm.082}$	$0.231_{\pm.044}$	$0.386_{\pm.058}$	$0.516_{\pm.063}$	$0.287_{\pm.083}$
w/o TO	$0.158_{\pm.007}$	$0.672_{\pm.014}$	$0.123_{\pm.009}$	$0.731_{\pm.055}$	$0.242_{\pm.059}$	$0.366_{\pm.062}$	$0.481_{\pm.052}$	$0.319_{\pm.098}$

The four ablation variants examine complementary aspects of the learned representation in LEAF. **w/o CT** uses only cycle-wise marginal states and increases MAPE in every domain, supporting the predictive value of the conditional-transition representation. **w/o FP** retains cross-cycle associations but removes their marginal constraints; the resulting error increases support the fixed-marginal correction within this architecture. For sequential modeling, **w/o TE** replaces learned transition encoding with aggregation, whereas **w/o TO** preserves the observed transitions but disrupts temporal order. Both variants increase mean MAPE across all four evaluated battery domains. Together, these comparisons associate the reported gains with the joint use of conditional representations, marginal consistency, and ordered transition encoding for early lifetime prediction.

4.4 Model Analysis

In this section, we examine the following questions about the learned transition representation: **Q1:** How do marginal and conditional discrepancies compare across early cycles? **Q2:** What dependence remains after fixed-marginal projection? **Q3:** How is prediction evaluated as the observation budget increases? **Q4:** Which local transitions receive the largest contribution scores? We show results on four representative battery datasets, HNEI, Stanford, Na-ion, and Zn-ion, while complete results are provided in the Appendix D.1- D.3.

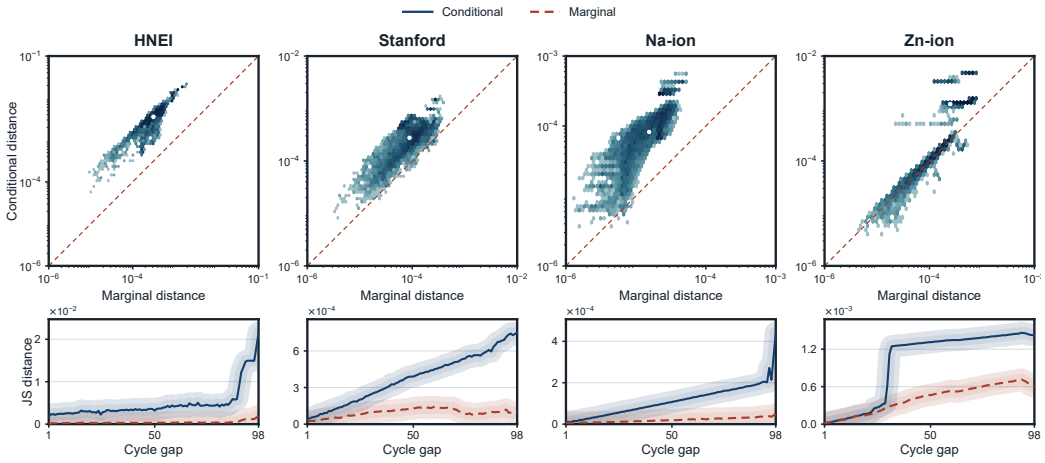


Figure 3: Marginal and conditional-transition discrepancies across early battery cycles. In the upper part, each point compares the marginal and conditional-transition distances for the same pair of cycle positions. The lower part shows how the JS distances change with increasing cycle gaps.

Conditional transitions (Q1). We visualize marginal and conditional-transition discrepancies across different early-cycle positions. In Figure 3, conditional discrepancies exceed marginal discrepancies for many pairs of early cycle positions across all four datasets, and their average sepa-

ration grows with the cycle gap. The plotted distributions therefore expose variation in cross-cycle dependence alongside the smaller changes in marginal cycle states.

Fixed-marginal projection (Q2). After fixed-marginal projection, clear and non-uniform conditional dependencies remain among local states. Figure 4 shows enhanced and suppressed associations relative to the independence baseline, forming distinct local source–target association patterns. Subtracting the outer product of the prescribed marginals gives a residual with zero row and column sums. These patterns therefore describe dependence within each fixed-marginal joint distribution.

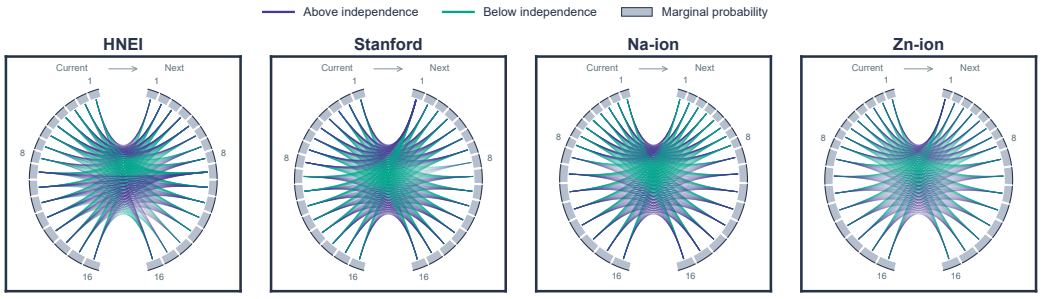


Figure 4: Fixed-marginal conditional-transition structures across four battery domains. Each circular plot connects source groups in the current cycle to target groups in the next cycle. Purple and green connections denote transitions enhanced and suppressed relative to the independence baseline, respectively, while the outer bars indicate marginal probabilities preserved during projection.

Early distinguishability (Q3). Figure 5 shows that LEAF consistently yields lower validation MAPE than CPMLP across all observation budgets and domains. Its advantage appears with limited early-cycle observations and persists as more cycles become available.

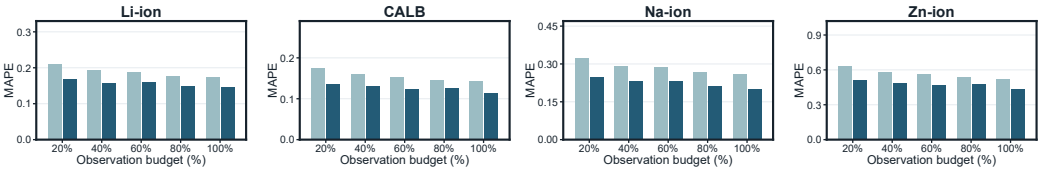


Figure 5: Validation MAPE of LEAF and CPMLP with increasing observation budgets.

Local contributions (Q4). Local conditional transitions receive highly non-uniform contribution scores. In Figure 6, larger scores concentrate on particular local source–target state pairs, and their locations differ across battery domains. These patterns identify the local associations emphasized by the fitted lifetime predictor. Grouping source and target supports further shows how the local contribution profile varies along the charge–discharge trajectory. LEAF models these local correspondences jointly with their ordered cross-cycle evolution for lifetime prediction.

4.5 Cross-Domain Transferability

In practice, new battery systems often have limited lifetime data, making cross-domain transferability important for early lifetime prediction. We pretrain LEAF on Li-ion and transfer it to Zn-ion, Na-ion, and CALB. The target-only baseline uses all target training batteries, while FT and DA use half. **Frozen** directly evaluates the pretrained model, **FT** fine-tunes all parameters, and **DA** additionally aligns source and target representations. Table 3 shows large errors for direct transfer. FT improves over Frozen in every domain, while DA gives the lowest MAPE among transfer variants in all targets and outperforms the target-only baseline on Zn-ion and Na-ion, whereas the baseline remains lowest for CALB. These results indicate that target adaptation is important, while source pretraining alone does not ensure improved target performance. Appendix D.4 details the protocol and compares models using identical target training subsets within each adaptation setting. We evaluate generalization under unseen aging conditions in Appendix D.5.

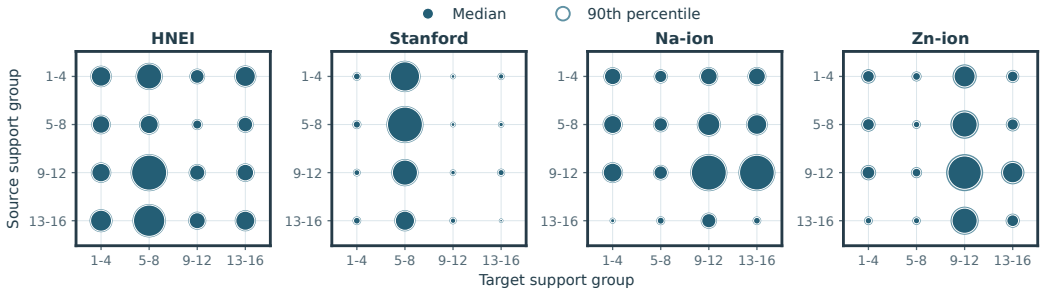


Figure 6: Local conditional-transition contributions to lifetime prediction across four battery domains, where rows and columns denote source and target support groups, and marker sizes indicate contribution magnitude summarized by the median and 90th percentile.

Table 3: Cross-domain transfer performance under different adaptation settings.

Methods	Metrics	Baseline	Frozen	FT	DA
Zn-ion	MAPE	$0.438_{\pm.031}$	$3.012_{\pm.164}$	$0.421_{\pm.057}$	$\mathbf{0.374}_{\pm.028}$
	15%-ACC	$0.351_{\pm.063}$	$0.061_{\pm.031}$	$0.369_{\pm.071}$	$\mathbf{0.423}_{\pm.047}$
Na-ion	MAPE	$0.202_{\pm.039}$	$3.864_{\pm.151}$	$0.217_{\pm.033}$	$\mathbf{0.181}_{\pm.024}$
	15%-ACC	$\mathbf{0.444}_{\pm.045}$	$0.018_{\pm.012}$	$0.417_{\pm.057}$	$0.432_{\pm.049}$
CALB	MAPE	$\mathbf{0.113}_{\pm.008}$	$1.428_{\pm.132}$	$0.128_{\pm.021}$	$0.119_{\pm.015}$
	15%-ACC	$0.778_{\pm.042}$	$0.236_{\pm.094}$	$0.751_{\pm.061}$	$\mathbf{0.802}_{\pm.048}$

4.6 Efficiency Analysis

Figure 7 compares training time and GPU memory consumption across representative baselines on the CALB dataset, with bubble size indicating the number of trainable parameters. LEAF has computational costs comparable to lightweight predictive baselines and substantially lower than more complex models such as DiffBatt and LiPM. Together with its prediction performance, LEAF achieves a favorable balance between accuracy and efficiency.

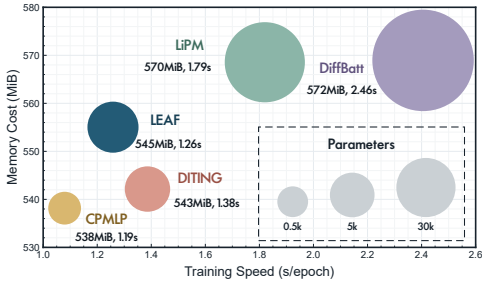


Figure 7: Training efficiency of baseline comparisons.

5 Conclusion

This paper studies early battery lifetime prediction through stationarity masking, where nearly invariant marginal distributions coexist with changes in adjacent-cycle conditional dependence. Our analysis separates marginal and conditional discrepancies and characterizes a unique joint projection with prescribed marginals. Building on this insight, LEAF combines local cross-cycle associations, fixed-marginal projection, and ordered transition modeling. The reported comparisons across 16 heterogeneous battery datasets yield lower mean prediction errors in four battery domains, while component ablations support the combined representation. Future work will examine longer-range transitions and broader industrial testing conditions for early battery lifetime prediction.

AI USE STATEMENT

Generative AI was used to assist with writing and language polishing, literature retrieval and related-work discovery, and the initial drafting and refinement of selected sections of the manuscript. Its use was limited to supporting the presentation, organization, and clarity of the paper, while the research ideas, methodology, experimental design, result analysis, and final conclusions were independently determined, checked, and verified by the authors.

ETHICS STATEMENT

This work focuses on machine learning for battery lifetime prediction and uses only battery cycling and degradation data. It does not involve human subjects, personally identifiable information, or sensitive attributes. The proposed method is intended to improve early lifetime assessment and battery degradation modeling and does not introduce applications with apparent ethical risks. We follow relevant data-use and research practices and are not aware of specific ethical, privacy, fairness, or security concerns that require further discussion.

REPRODUCIBILITY STATEMENT

We provide detailed theoretical derivations and proofs, the complete algorithm and projection solver, and comprehensive experimental details, including data preprocessing, model configurations, training procedures, and evaluation protocols. Additional experiments and analyses are also reported, and code and implementation details are provided to facilitate reproduction of the reported results.

REFERENCES

Peter M. Attia, Aditya Grover, Norman Jin, Kristen A. Severson, Todor M. Markov, Yang-Hung Liao, Michael H. Chen, Bryan Cheong, Nicholas Perkins, Zi Yang, et al. Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature*, 578(7795):397–402, 2020.

Peter M. Attia, Kristen A. Severson, and Jeremy 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.

Christoph R. Birkl, Matthew R. Roberts, Euan McTurk, Peter G. Bruce, and David A. Howey. Degradation diagnostics for lithium ion cells. *Journal of Power Sources*, 341:373–386, 2017. doi: 10.1016/j.jpowsour.2016.12.011.

Xiao Cui, Stephen Dongmin Kang, Sunny Wang, Justin A. Rose, Huada Lian, Alexis Geslin, Steven B. Torrisi, Martin Z. Bazant, Shijing Sun, and William C. Chueh. Data-driven analysis of battery formation reveals the role of electrode utilization in extending cycle life. *Joule*, 8 (11):3072–3087, 2024. doi: 10.1016/j.joule.2024.07.024.

Arnaud Devie, George Baure, and Matthieu Dubarry. Intrinsic variability in the degradation of a batch of commercial 18650 lithium-ion cells. *Energies*, 11(5):1031, 2018. doi: 10.3390/en11051031.

Yuntao Du, Jindong Wang, Wenjie Feng, Sinno Pan, Tao Qin, Renjun Xu, and Chongjun Wang. AdaRNN: Adaptive learning and forecasting of time series. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, pp. 402–411, 2021. doi: 10.1145/3459637.3482315.

Jacqueline S. Edge, Simon O'Kane, Ryan Prosser, Niall D. Kirkaldy, Anisha N. Patel, Alastair Hales, Abir Ghosh, Weilong Ai, Jingyi Chen, Jiang Yang, Shen Li, Mei-Chin Pang, Laura Bravo Diaz, Anna Tomaszewska, M. Waseem Marzook, Karthik N. Radhakrishnan, Huizhi Wang, Yatish Patel, Billy Wu, and Gregory J. Offer. Lithium ion battery degradation: What you need to know. *Physical Chemistry Chemical Physics*, 23(14):8200–8221, 2021. doi: 10.1039/D1CP00359C.

Hamidreza Eivazi, André Hebenbrock, Raphael Ginster, Steffen Blömeke, Stefan Wittek, Christoph Herrmann, Thomas S. Spengler, Thomas Turek, and Andreas Rausch. Diffbatt: A diffusion model for battery degradation prediction and synthesis. In *Foundation Models for Science Workshop at the 38th Conference on Neural Information Processing Systems*, 2024.

Wei Fan, Pengyang Wang, Dongkun Wang, Dongjie Wang, Yuanchun Zhou, and Yanjie Fu. Dish-TS: A general paradigm for alleviating distribution shift in time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 7522–7529, 2023. doi: 10.1609/aaai.v37i6.25914.

Zicheng Fei, Fangfang Yang, Kwok-Leung Tsui, Lishuai Li, and Zijun Zhang. Early prediction of battery lifetime via a machine learning based framework. *Energy*, 225:120205, 2021. doi: 10.1016/j.energy.2021.120205.

Zicheng Fei, Zijun Zhang, Fangfang Yang, Kwok-Leung Tsui, and Lishuai Li. Early-stage lifetime prediction for lithium-ion batteries: A deep learning framework jointly considering machine-learned and handcrafted data features. *Journal of Energy Storage*, 52:104936, 2022. doi: 10.1016/j.est.2022.104936.

Y. Han, C. Li, L. Zheng, G. Lei, and L. Li. Remaining useful life prediction of lithium-ion batteries by using a denoising transformer-based neural network. *Energies*, 16(17):6328, 2023.

N. He, Q. Wang, Z. Lu, Y. Chai, and F. Yang. Early prediction of battery lifetime based on graphical features and convolutional neural networks. *Applied Energy*, 353:122048, 2024.

Chia-Wei Hsu, Rui Xiong, Nan-Yow Chen, Ju Li, and Nien-Ti Tsou. Deep neural network battery life and voltage prediction by using data of one cycle only. *Applied Energy*, 306:118134, 2022. doi: 10.1016/j.apenergy.2021.118134.

Jincheng Hu, Pengyu Fu, Zhongbao Wei, Yanjun Huang, Juliana Early, Ashley Fly, and Yuanjian Zhang. Early prediction of lithium-ion battery degradation with a generative pre-trained transformer. *Nature Communications*, 17(1):126, 2026. doi: 10.1038/s41467-025-66819-0.

Xinghao Huang, Chen Liang, Shengyu Tao, Yunhong Che, Ningyu Bian, Jiale Zhang, Runhua Wang, Yuqi Zhang, Bizhong Xia, and Xuan Zhang. IC^{2}ML: Unified battery state-of-health, degradation trajectory and remaining useful life prediction via intra-cycle and inter-cycle enhanced machine learning. *Journal of Power Sources*, 666:239148, 2026. doi: 10.1016/j.jpowsour.2025.239148.

Daniel Juarez-Robles, Judith A. Jeevarajan, and Partha P. Mukherjee. Degradation-safety analytics in lithium-ion cells: Part i. aging under charge/discharge cycling. *Journal of The Electrochemical Society*, 167(16):160510, 2020.

Daniel Juarez-Robles, Saad Azam, Judith A. Jeevarajan, and Partha 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, 2021. doi: 10.1149/1945-7111/ac30af.

Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International Conference on Learning Representations*, 2022.

Juren Li, Yang Yang, Hanchen Su, Jiayu Liu, Youmin Chen, Jianfeng Zhang, and Lujia Pan. Lipm: Foundation model for lithium-ion battery analysis. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, pp. 1412–1423, 2025. doi: 10.1145/3711896.3737027.

Tingkai Li, Zihao Zhou, Adam Thelen, David A. Howey, and Chao Hu. Predicting battery lifetime under varying usage conditions from early aging data. *Cell Reports Physical Science*, 5(4):101891, 2024. doi: 10.1016/j.xcrp.2024.101891.

Weihan Li, Neil Sengupta, Philipp Dechent, David Howey, Anuradha Annaswamy, and Dirk Uwe Sauer. One-shot battery degradation trajectory prediction with deep learning. *Journal of Power Sources*, 506:230024, 2021. doi: 10.1016/j.jpowsour.2021.230024.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024.

Guijun Ma, Songpei Xu, Benben Jiang, Cheng Cheng, Xin Yang, Yue Shen, Tao Yang, Yunhui Huang, Han Ding, and Ye Yuan. Real-time personalized health status prediction of lithium-ion batteries using deep transfer learning. *Energy & Environmental Science*, 15(10):4083–4094, 2022. doi: 10.1039/D2EE01676A.

Hao Miao, Ni Zhang, Zefei Ning, and Li Wang. DITING: A weak degradation listener for battery lifetime early prediction. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306 of *Proceedings of Machine Learning Research*. PMLR, 2026.

Peyman Mohtat, Suhak Lee, Jason B. Siegel, and Anna G. Stefanopoulou. Reversible and irreversible expansion of lithium-ion batteries under a wide range of stress factors. *Journal of The Electrochemical Society*, 168(10):100520, 2021.

Simon E. J. O'Kane, Weilong Ai, Ganesh Madabattula, Diego Alonso-Alvarez, Robert Timms, Valentin Sulzer, Jacqueline S. Edge, Billy Wu, Gregory J. Offer, and Monica Marinescu. Lithium-ion battery degradation: How to model it. *Physical Chemistry Chemical Physics*, 24:7909–7922, 2022. doi: 10.1039/D2CP00417H.

Noah Paulson, Joseph Kubal, Logan Ward, Saurabh Saxena, Wenquan Lu, and Susan Babinec. Feature engineering for machine learning enabled early prediction of battery lifetime. *Journal of Power Sources*, 527:231127, 2022. doi: 10.1016/j.jpowsour.2022.231127.

Yuliya Preger, Heather M. Barkholtz, Armando Fresquez, Daniel L. Campbell, Benjamin W. Juba, Jessica Roman-Kustas, Summer R. Ferreira, and Babu Chalamala. Degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions. *Journal of The Electrochemical Society*, 167(12):120532, 2020. doi: 10.1149/1945-7111/abae37.

Kristen A. Severson, Peter M. Attia, Norman Jin, Nicholas Perkins, Benben Jiang, Zi Yang, Michael H. Chen, Muratahan Aykol, Patrick K. Herring, Dimitrios Fraggedakis, Martin Z. Bazant, Stephen J. Harris, William C. Chueh, and Richard D. Braatz. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4:383–391, 2019. doi: 10.1038/s41560-019-0356-8.

Ruifeng Tan, Xibin Lu, Minhao Cheng, Jia Li, Jiaqiang Huang, and Tong-Yi Zhang. Forecasting battery degradation trajectory under domain shift with domain generalization. *Energy Storage Materials*, 72:103725, 2024. doi: 10.1016/j.ensm.2024.103725.

Ruifeng Tan, Weixiang Hong, Jiayue Tang, Xibin Lu, Ruijun Ma, Xiang Zheng, Jia Li, Jiaqiang Huang, and Tong-Yi Zhang. BatteryLife: A comprehensive dataset and benchmark for battery life prediction. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 5789–5800. Association for Computing Machinery, 2025. doi: 10.1145/3711896.3737372.

Ruifeng Tan, Jintao Dong, Weixiang Hong, Jia Li, Jiaqiang Huang, and Tong-Yi Zhang. Batterymformer: Multi-level learning for battery degradation trajectory forecasting. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, 2026. doi: 10.1145/3770855.3818948.

Yugui Tang, Kuo Yang, Haoran Zheng, Shujing Zhang, and Zhen Zhang. Early prediction of lithium-ion battery lifetime via a hybrid deep learning model. *Measurement*, 199:111530, 2022. doi: 10.1016/j.measurement.2022.111530.

Shengyu Tao, Chongbo Sun, Shiyi Fu, Yu Wang, Ruifei Ma, Zhiyuan Han, Yaojie Sun, Yang Li, Guodan Wei, Xuan Zhang, Guangmin Zhou, and Hongbin Sun. Battery cross-operation-condition lifetime prediction via interpretable feature engineering assisted adaptive machine learning. *ACS Energy Letters*, 8(8):3269–3279, 2023. doi: 10.1021/acsenergylett.3c01012.

Fujin Wang, Zhi Zhai, Zhibin Zhao, Yi Di, and Xuefeng Chen. Physics-informed neural network for lithium-ion battery degradation stable modeling and prognosis. *Nature Communications*, 15 (1):4332, 2024.

Andrew Weng, Peyman Mohtat, Peter M. Attia, Valentin Sulzer, Suhak Lee, Gregory Less, and Anna Stefanopoulou. Predicting the impact of formation protocols on battery lifetime immediately after manufacturing. *Joule*, 5(11):2971–2992, 2021.

Yinjiao Xing, Eden W. M. Ma, Kwok-Leung Tsui, and Michael Pecht. An ensemble model for predicting the remaining useful performance of lithium-ion batteries. *Microelectronics Reliability*, 53(6):811–820, 2013. doi: 10.1016/j.microrel.2012.12.003.

Han Zhang, Xiaofan Gui, Shun Zheng, Ziheng Lu, Yuqi Li, and Jiang Bian. BatteryML: An open-source platform for machine learning on battery degradation. In *International Conference on Learning Representations*, 2024.

Han Zhang, Yuqi Li, Shun Zheng, Ziheng Lu, Xiaofan Gui, Wei Xu, and Jiang Bian. Battery lifetime prediction across diverse ageing conditions with inter-cell deep learning. *Nature Machine Intelligence*, 7:270–277, 2025. doi: 10.1038/s42256-024-00972-x.

Jiangong Zhu, Yixiu Wang, Yuan Huang, R. Bhushan Gopaluni, Yankai Cao, Michael Heere, Martin J. Mühlbauer, Liuda Mereacre, Haifeng Dai, Xinhua Liu, et al. Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation. *Nature Communications*, 13(1): 2261, 2022.

A Detailed Motivation and Theoretical Analysis

A.1 Localized Early-Cycle Variations

As an illustrative example, we examine voltage variations within early battery cycles. Figure 8 visualizes the normalized voltage variations along the within-cycle axis for all datasets, where the first and second halves correspond to the charge and discharge segments, respectively. Although the magnitude and location of these variations differ substantially across battery systems, larger changes are often concentrated in specific local regions, while substantial portions of each cycle remain relatively stable. The percentile ranges further show that the locations and magnitudes of these local variations can differ across batteries within the same dataset. This example illustrates that early-cycle variations can exhibit pronounced localized structure rather than being distributed uniformly along the cycling trajectory, motivating local cross-cycle modeling.

A.2 Stationarity Masking across More Datasets

We extend the distribution analysis of stationarity masking to all battery datasets. As shown in Figure 9, one battery is randomly selected from each dataset to visualize the marginal voltage distributions at cycles 2, 49, and 99, together with the corresponding conditional distributions of $2 \to 3$, $49 \to 50$, and $99 \to 100$. Across the selected batteries, the marginal voltage distributions remain highly similar across cycle positions, whereas adjacent-cycle conditional distributions show clearer variations. These observation-level visualizations illustrate the contrast between marginal similarity and conditional variation across different battery systems.

A.3 Further Motivation of Stationarity Masking

Battery aging accumulates across cycles, while its observable effects can remain weak during early life (Edge et al., 2021; Birkl et al., 2017). From a probabilistic perspective, the joint distribution of two adjacent cycles can be factorized as $J_c = P_c K_c$, where P_c describes the marginal state distribution of the current cycle and K_c describes the conditional transition to the next cycle. This factorization shows that similar marginals do not uniquely determine the adjacent-cycle joint law: even when cycle-wise state compositions change only slightly, their cross-cycle dependence can still differ. Consequently, marginal cycle states alone may provide a limited description of subtle early evolution, whereas adjacent-cycle conditional associations offer a complementary and more structured statistical view. LEAF builds on this distinction by learning local associations between adjacent cycles and projecting their joint representation while preserving both cycle marginals, thereby separating state composition from cross-cycle dependence within the same marginal constraints.

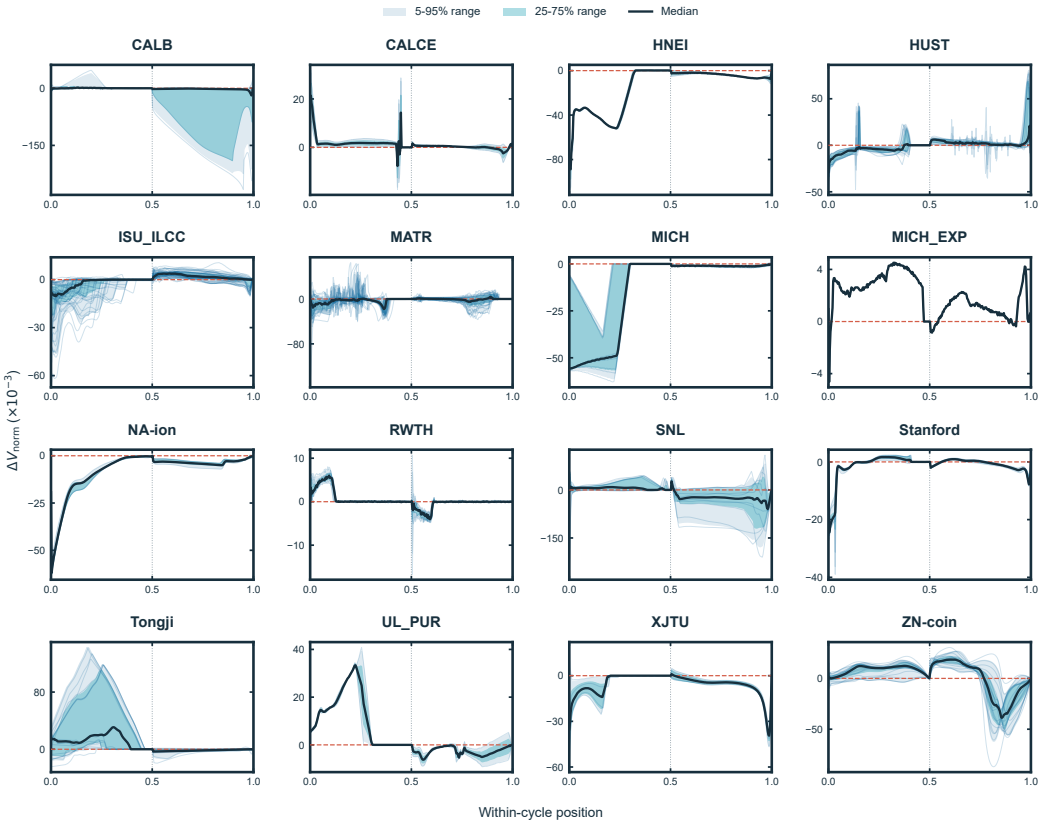


Figure 8: Normalized voltage variations across within-cycle positions, summarized by the median and 25–75% and 5–95% ranges.

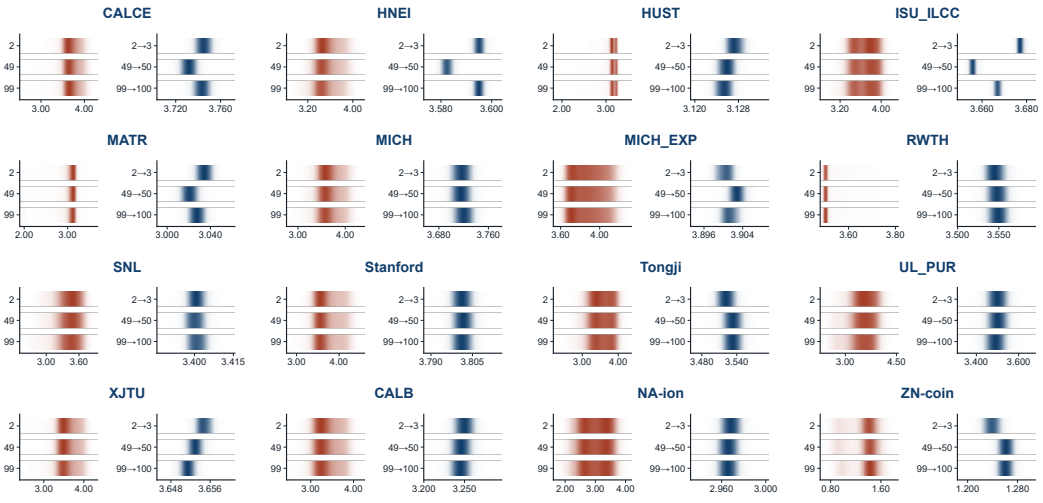


Figure 9: Visualization of stationarity masking across different battery datasets. One battery is randomly selected from each dataset. The left panels show the marginal voltage distributions at cycles 2, 49, and 99, while the right panels show the corresponding conditional distributions of $2 \to 3$, $49 \to 50$, and $99 \to 100$, revealing clear transition variations.

A.4 Theoretical Analysis

Proof of Theorem 3.1. Under the absolute-continuity and finiteness assumptions in the theorem, substituting $J_c(x, y) = P_c(x)K_c(y \mid x)$ into the KL divergence and separating the marginal and conditional terms gives:

$$\begin{aligned} D_{\mathrm{KL}}(J_c \| J_{c'}) &= \mathbb{E}_{(x,y)\sim J_c}\left[\log \frac{P_c(x)}{P_{c'}(x)} + \log \frac{K_c(y \mid x)}{K_{c'}(y \mid x)}\right] \\ &= D_{\mathrm{KL}}(P_c \| P_{c'}) + \mathbb{E}_{x\sim P_c} D_{\mathrm{KL}}\left(K_c(\cdot \mid x) \| K_{c'}(\cdot \mid x)\right). \end{aligned} \tag{13}$$

When $P_c = P_{c'}$, the marginal term vanishes. If the conditional transitions differ on a set of positive source probability, the conditional term is strictly positive, yielding distinct adjacent-cycle joint laws despite identical marginals. For nearly invariant marginals, the same decomposition explicitly separates marginal variation from conditional-transition variation. This decomposition characterizes stationarity masking as a regime in which weak cycle-wise distribution changes coexist with stronger variation in cross-cycle dependence.

Representation-level conditional transition. The marginal–conditional decomposition naturally extends from the observation space to the learned representation space. Let

$$Z_c = \phi_\theta(X_c), \tag{14}$$

where ϕ_θ is a shared representation map applied to cycle observations. The adjacent-cycle representation pair (Z_c, Z_{c+1}) induces the joint law

$$J_c^Z(z, z') = P_c^Z(z)K_c^Z(z' \mid z), \tag{15}$$

where P_c^Z denotes the cycle-wise representation distribution and K_c^Z denotes the corresponding adjacent-cycle conditional transition.

For two cycle positions c and c', the same KL decomposition gives

$$\begin{aligned} D_{\mathrm{KL}}\left(J_c^Z \| J_{c'}^Z\right) &= D_{\mathrm{KL}}\left(P_c^Z \| P_{c'}^Z\right) \\ &\quad + \mathbb{E}_{z\sim P_c^Z} D_{\mathrm{KL}}\left(K_c^Z(\cdot \mid z) \| K_{c'}^Z(\cdot \mid z)\right). \end{aligned} \tag{16}$$

Moreover, since (Z_c, Z_{c+1}) is obtained from (X_c, X_{c+1}) through a deterministic measurable mapping, the data-processing inequality gives

$$D_{\mathrm{KL}}\left(J_c^Z \| J_{c'}^Z\right) \le D_{\mathrm{KL}}\left(J_c \| J_{c'}\right). \tag{17}$$

Thus, the learned representation provides a compressed space in which cross-cycle distributional variation can be organized into cycle-wise marginal variation and adjacent-cycle conditional variation.

LEAF instantiates this representation using local patch states. For an adjacent-cycle pair $(\mathbf{X}_c, \mathbf{X}_{c+1})$, the shared patch encoder produces $\{h_{c,i}\}_{i=1}^K$ and $\{h_{c+1,j}\}_{j=1}^K$. The local association matrix

$$A_c(i, j) = s_\theta(h_{c,i}, h_{c+1,j}) \tag{18}$$

measures source–target correspondence between local states. Row-wise normalization then defines

$$K_c(j \mid i) = \frac{\exp A_c(i, j)}{\sum_{j'=1}^K \exp A_c(i, j')}, \tag{19}$$

which describes the conditional allocation of target states given each source state. Combined with the current-cycle state distribution P_c, this yields the patch-level joint transition

$$J_c(i, j) = P_c(i)K_c(j \mid i). \tag{20}$$

This construction transfers the marginal–conditional factorization to a finite local-state space. Cycle-wise state composition is represented by P_c, while cross-cycle dependence is represented by K_c. Their separation provides a structured representation of early-cycle evolution under stationarity masking.

Proof of Theorem 3.2. For probability vectors μ and ν, the outer product $\mu\nu^\top$ satisfies both marginal constraints and therefore belongs to $\Pi(\mu,\nu)$, so the feasible set is nonempty. The affine equality constraints and nonnegativity constraints make $\Pi(\mu,\nu)$ closed and convex. Moreover, for every $Q \in \Pi(\mu,\nu)$,

$$0 \le Q_{ij} \le \mu_i,$$

so the feasible set is bounded and hence compact.

Define

$$\phi(Q) = \frac{1}{2}\|Q - G\|_F^2. \tag{21}$$

Since ϕ is continuous and strictly convex, it admits a unique minimizer $Q^\star$ over $\Pi(\mu,\nu)$.

For any $Q \in \Pi(\mu,\nu)$, convexity gives

$$Q^\star + t(Q - Q^\star) \in \Pi(\mu,\nu), \qquad 0 \le t \le 1.$$

Taking the one-sided derivative of $\phi(Q^\star + t(Q - Q^\star))$ at $t = 0$ yields

$$\langle Q^\star - G, Q - Q^\star \rangle_F \ge 0. \tag{22}$$

Conversely,

$$\begin{aligned}\phi(Q) - \phi(Q^\star) &= \langle Q^\star - G, Q - Q^\star \rangle_F + \frac{1}{2}\|Q - Q^\star\|_F^2 \\ &\ge 0.\end{aligned} \tag{23}$$

For $Q \neq Q^\star$, the second term is strictly positive. Therefore, $Q^\star$ is the unique Euclidean projection of G onto the fixed-marginal feasible set $\Pi(\mu,\nu)$:

$$Q^\star = \arg\min_{Q \in \Pi(\mu,\nu)} \frac{1}{2}\|Q - G\|_F^2. \tag{24}$$

Hence, among all joint transitions sharing the prescribed source and target marginals, $Q^\star$ preserves the reference association structure with minimum Euclidean adjustment.

Dependence structure under fixed marginals. For any $Q \in \Pi(p,q)$, the independent joint distribution with the same marginals is $pq^\top$. Define the residual

$$R = Q - pq^\top. \tag{25}$$

Since both Q and $pq^\top$ share the same row and column marginals,

$$R\mathbf{1} = 0, \qquad R^\top \mathbf{1} = 0. \tag{26}$$

Therefore, R characterizes the degrees of freedom in the joint distribution that remain after both marginal state compositions have been fixed.

For source states with $p_i > 0$, the corresponding conditional transition is

$$\widehat{K} = \mathrm{diag}(p)^{-1}Q = \mathbf{1}q^\top + \mathrm{diag}(p)^{-1}R. \tag{27}$$

The first term, $\mathbf{1}q^\top$, is the conditional transition induced by source–target independence, whereas $\mathrm{diag}(p)^{-1}R$ describes the source-dependent departure from this baseline.

This decomposition separates cycle-wise state composition from cross-cycle association. The fixed marginals p and q determine the state composition of the two cycles, while R captures the remaining dependence between source and target states. Consequently, different adjacent-cycle transition structures can be represented under identical marginal state compositions.

Applied to LEAF, the projection

$$Q_c^\star = \mathrm{Proj}_{\Pi(P_c, P_{c+1})}(J_c) \tag{28}$$

retains the prescribed state distributions of both adjacent cycles while extracting a marginal-consistent cross-cycle dependence structure. The corrected conditional transition

$$\widehat{K}_c = \mathrm{diag}(P_c)^{-1}Q_c^\star \tag{29}$$

therefore provides a representation in which changes in local source–target association can be modeled independently of changes in cycle-wise state composition. Its ordered evolution across early cycles forms the transition sequence used by LEAF for lifetime prediction.

A.5 Relation to Existing Battery Lifetime Prediction Methods

Early battery lifetime prediction has progressed from handcrafted degradation descriptors to learned representations of temporal evolution. Early statistical approaches relate lifetime to capacity trajectories, cycling-curve differences, and engineered features (Severson et al., 2019; Attia et al., 2021; Fei et al., 2021; Paulson et al., 2022), while subsequent deep models learn degradation representations directly from current, voltage, and capacity observations and model their evolution across cycles (Tang et al., 2022; Fei et al., 2022; Hsu et al., 2022; Ma et al., 2022). Recent methods further improve representation learning from complementary directions. BatLiNet exploits relative degradation across batteries through inter-cell learning (Zhang et al., 2025), LiPM learns transferable degradation knowledge through large-scale pretraining (Li et al., 2025), and DITING identifies weak degradation through deviations from an early-cycle health reference (Miao et al., 2026). CyclePatch-based models, including CPMLP, CPGRU, and CPTransformer, encode individual cycles into cycle-level representations and then model their evolution along the cycle axis (Tan et al., 2025). Despite these advances, their primary modeling objects remain cycle-wise representations, reference deviations, inter-cell differences, or temporal dependencies among cycle features.

LEAF instead takes the conditional relationship between adjacent cycles as the primary representation of early degradation. Under stationarity masking, marginal cycle distributions may remain highly similar while cross-cycle dependence continues to vary. LEAF therefore learns local associations between adjacent-cycle states, constructs conditional transitions, and projects the resulting joint transitions onto a fixed-marginal feasible set, preserving the state compositions of both cycles while isolating their dependence structure. This also distinguishes LEAF from direct cycle differences, which summarize changes in observed values without explicitly representing source-dependent transitions between local states. The contribution of LEAF is thus not merely the use of adjacent cycles or a more complex temporal backbone, but an explicit probabilistic representation that separates cycle-wise state composition from adjacent-cycle conditional dependence and models its ordered evolution over early life.

B The Algorithm of LEAF

Let s_y denote the mean lifetime of the training batteries.

Algorithm 1 LEAF – Overall Architecture

Input: early-cycle observations $\{\mathbf{X}_c\}_{c=1}^{C}$, lifetime label y
Output: predicted lifetime $\hat{y}$
for $c = 1$ to C **do**
  divide $\mathbf{X}_c$ into K consecutive patches
  $\{\mathbf{h}_{c,i}\}_{i=1}^{K} \leftarrow \mathrm{LinearPatchEncoder}(\mathbf{X}_c)$
  compute normalized state distribution P_c
end for
for $c = 1$ to $C-1$ **do**
  $A_c(i,j) \leftarrow s_\theta(\mathbf{h}_{c,i}, \mathbf{h}_{c+1,j})$
  $K_c(j \mid i) \leftarrow \mathrm{Softmax}_j(A_c(i,j))$
  $J_c(i,j) \leftarrow P_c(i) K_c(j \mid i)$
  $Q_c^\star \leftarrow \arg\min_{Q \in \Pi(P_c, P_{c+1})} \frac{1}{2}\|Q - J_c\|_F^2$
  $\widehat{K}_c(j \mid i) \leftarrow Q_c^\star(i,j)/P_c(i)$
end for
$\mathbf{Z} \leftarrow \mathrm{Conv1D}_{k=5}([\widehat{K}_1, \ldots, \widehat{K}_{C-1}])$
$\hat{y} \leftarrow \mathrm{PredictionHead}(\mathbf{Z})$
minimize $\mathcal{L} = (\hat{y}/s_y - y/s_y)^2$ and update model parameters

Projection solver. For $p = P_c$ and $q = P_{c+1}$, we solve the fixed-marginal Euclidean projection using Dykstra's algorithm. For $X \in \mathbb{R}^{K \times K}$, the affine projection onto the set with row marginal p and column marginal q is:

$$\mathcal{A}_{p,q}(X) = X + \frac{(p - X\mathbf{1})\mathbf{1}^\top}{K} + \frac{\mathbf{1}(q - X^\top\mathbf{1})^\top}{K} - \frac{1 - \mathbf{1}^\top X \mathbf{1}}{K^2}\mathbf{1}\mathbf{1}^\top. \tag{30}$$

Dykstra's iteration is initialized with $X = J_c$ and $U = V = 0$, and alternates between the affine marginal constraints and nonnegativity:

$$Y = \mathcal{A}_{p,q}(X + U), \qquad U \leftarrow X + U - Y, \tag{31}$$

$$X' = \max(Y + V, 0), \qquad V \leftarrow Y + V - X', \qquad X \leftarrow X'. \tag{32}$$

The iteration continues until both the marginal residuals and successive iterate changes satisfy the prescribed convergence tolerances. The resulting joint transition is then converted to the corresponding conditional transition using the source marginal.

C Experimental Details

C.1 Degradation Test

Battery lifetime is commonly obtained through controlled cycling degradation tests (Tan et al., 2025). Batteries are operated repeatedly under predefined charge and discharge protocols while current and voltage are recorded throughout the test. Each cycle consists of charging and discharging stages, forming a complete degradation process as cycling progresses. For the i-th cycle, the discharge capacity is computed by integrating the measured current over the discharge interval:

$$Q_i = \int_{t_i^s}^{t_i^e} |I(t)|\, dt, \tag{33}$$

where t_i^s and t_i^e denote the beginning and end of the discharge stage. The cycle-level state of health is defined as

$$\mathrm{SOH}_i = \frac{Q_i}{Q_0}, \tag{34}$$

where Q_0 denotes the reference capacity.

For Li-ion, Na-ion, and Zn-ion batteries, the nominal capacity is used as Q_0, and battery lifetime is defined by the first cycle at which SOH reaches 80% (Severson et al., 2019; Tan et al., 2025). For CALB batteries, the first-cycle discharge capacity is used as the reference capacity and the lifetime threshold is set to 90% SOH (Tan et al., 2025). Different chemical systems, operating temperatures, and charge/discharge protocols together form diverse degradation conditions.

C.2 Dataset Details and Preprocessing

The evaluation covers 16 battery datasets, including 13 public Li-ion datasets together with the CALB, Na-ion, and Zn-ion domains. These datasets differ substantially in cell format, electrode chemistry, nominal capacity, operating temperature, and cycling protocol, providing diverse degradation conditions for battery lifetime prediction. Table 4 summarizes their major characteristics.

Table 4: Statistics of the four battery domains.

Domain	Batteries	Formats	Chemistries	Temperatures	Protocols	Test Setting
Li-ion	845	5	15	8	631	Laboratory
CALB	27	1	1	4	2	Industrial
Na-ion	31	1	1	2	12	Laboratory
Zn-ion	95	3	63	3	1	Laboratory

CALCE. The CALCE dataset contains prismatic Li-ion cells with $LiCoO_2$ cathodes and graphite anodes. The cells have a nominal capacity of 1.1 Ah and were cycled at 25°C under two operating protocols (Xing et al., 2013).

HNEI. HNEI consists of commercial 18650 cells with a mixed $LiCoO_2$/Ni–Co–Mn oxide cathode and a graphite anode. The cells have a nominal capacity of 2.8 Ah and were evaluated at 25°C using two cycling protocols (Devie et al., 2018).

MATR. MATR provides a large collection of commercial 18650 $LiFePO_4$/graphite cells. A broad range of fast-charging strategies is included while the discharge procedure remains largely shared across cells, making the dataset particularly diverse in charging conditions (Severson et al., 2019; Attia et al., 2020).

UL_PUR. UL_PUR contains 18650 cells using $LiNi_{0.8}Co_{0.15}Al_{0.05}O_2$ cathodes and graphite anodes. Their nominal capacity is 3.4 Ah, and the cells were aged under a common cycling procedure at approximately 23°C (Juarez-Robles et al., 2020; 2021).

SNL. SNL combines several commercial 18650 cell chemistries, including $LiFePO_4$, NCA, and NMC cathodes. The experiments span multiple cycling protocols and temperatures of 15°C, 25°C, and 35°C, introducing substantial variation in both chemistry and operating conditions (Preger et al., 2020).

MICH_EXP. MICH_EXP consists of NMC111/graphite pouch cells with a nominal capacity of 5.0 Ah. Cycling experiments were conducted under several protocols at -5°C, 25°C, and 45°C, providing degradation records under different thermal stresses (Mohtat et al., 2021).

MICH. MICH contains NMC111/graphite pouch cells with a nominal capacity of 2.36 Ah. The batteries were tested at 25°C and 45°C under two cycling settings designed to investigate the influence of formation and operating conditions on subsequent aging (Weng et al., 2021).

RWTH. RWTH comprises commercial 18650 cells using an NMC cathode and a carbon-based anode. The cells have a nominal capacity of 3 Ah and were cycled under a common protocol at 25°C (Li et al., 2021).

HUST. HUST contains 18650 $LiFePO_4$/graphite cells with a nominal capacity of 1.1 Ah. The charging procedure is shared across cells, whereas the discharge conditions vary through a large collection of multi-stage protocols, producing substantial diversity in usage conditions (Ma et al., 2022).

Tongji. Tongji includes commercial 18650 cells spanning multiple electrode combinations, nominal capacities, operating temperatures, and cycling protocols. These variations provide heterogeneous degradation patterns across both battery chemistry and operating conditions (Zhu et al., 2022).

Stanford. Stanford consists of small-capacity pouch cells with NMC532 cathodes and artificial-graphite anodes. Different formation procedures are applied before the cells enter a common cycling protocol at 25°C, enabling analysis of the relationship between early formation and long-term lifetime (Cui et al., 2024).

XJTU. XJTU provides commercial 18650 NMC532/graphite cells with a nominal capacity of 2 Ah. The cells were operated at 20°C under two cycling conditions, offering degradation records under a controlled thermal environment (Wang et al., 2024).

ISU_ILCC. ISU_ILCC contains small Li-polymer cells with NMC cathodes and graphite anodes. A large number of charge/discharge usage conditions are included, resulting in one of the most diverse protocol collections among the Li-ion datasets (Li et al., 2024).

CALB. CALB contains large-format commercial prismatic Li-ion cells tested in an industrial environment. The cells employ NMC cathodes and graphite anodes with a nominal capacity of 58 Ah, and their cycling experiments span four operating temperatures and two industrial protocols (Tan et al., 2025).

Na-ion. The Na-ion dataset contains commercial 18650 sodium-ion cells with a nominal capacity of 1 Ah. Cycling is performed at 25°C under multiple charge/discharge rates ranging from 2C to 6C, resulting in twelve distinct operating protocols (Tan et al., 2025).

Zn-ion. The Zn-ion dataset consists of laboratory-assembled coin cells using MnO_2 cathodes and metallic-zinc anodes. Variations in electrolyte additives, coin-cell formats, and operating temperatures provide a diverse collection of zinc-ion degradation conditions (Tan et al., 2025).

Data Processing. We follow BatteryLife preprocessing, removing formation cycles, reference performance tests, and abnormal cycles under its dataset-specific cleaning rules. Each valid charge and discharge segment is independently resampled by linear interpolation to 150 positions, giving 300 positions with current, voltage, and capacity channels. Current and capacity are divided by nominal

capacity, and voltage by its within-cycle maximum. Each battery yields eligible prefix samples from its first 100 valid cycles, with one lifetime prediction per prefix. Unobserved cycles are zero-padded and masked, so each sample uses only its available early-cycle prefix. All prefixes from the same battery share its lifetime label and remain in the same data partition.

Life Label Construction. Battery lifetime is determined from the capacity degradation trajectory. For Li-ion, Na-ion, and Zn-ion batteries, the lifetime label is defined as the first cycle at which the state of health reaches 80%. For CALB, a 90% threshold is adopted because most industrially tested batteries do not degrade to 80% within the available testing period. Following BatteryLife, batteries whose final SOH exceeds the corresponding threshold by more than 2.5 percentage points are excluded. If the final SOH remains above the threshold by at most 2.5 percentage points, linear extrapolation of the terminal trend estimates the cycle at which the threshold is reached. Thus, the admissible final-SOH intervals for extrapolation are $(80\%, 82.5\%]$ for Li-ion, Na-ion, and Zn-ion, and $(90\%, 92.5\%]$ for CALB. For CALB, the first-cycle discharge capacity is used as the reference capacity when computing SOH. Batteries with lifetime labels no larger than 100 cycles are excluded, following the benchmark's early-prediction setting. Labels describe the lifetime endpoint, whereas model inputs contain only the selected early-cycle prefix.

Data Partition. Each domain follows BatteryLife's approximately 6:2:2 battery-level training, validation, and test split. The Li-ion partition uses random seed 2021. For each of CALB, Na-ion, and Zn-ion, three different partitions use split seeds 42, 2021, and 2024. All methods use the same partition for each corresponding run. Prefix samples are generated within these battery-level partitions, and all prefixes from one battery remain in the same subset. Battery membership is specified by the released split files; validation data guide model selection, and test labels do not enter this process.

C.3 Evaluation Metrics

For a test set containing N prefix samples, let y_i and $\hat{y}_i$ denote the ground-truth and predicted lifetimes of the i-th sample, respectively. The relative prediction error is defined as:

$$e_i = \frac{|y_i - \hat{y}_i|}{y_i}. \tag{35}$$

Based on this quantity, mean absolute percentage error is defined as:

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N} e_i. \tag{36}$$

MAPE measures the average relative deviation between predicted and ground-truth lifetimes, where a lower value indicates more accurate lifetime prediction. We further use 15%-Acc to quantify the proportion of test prefixes whose predictions fall within a prescribed relative-error range:

$$15\%\text{-Acc} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\left(e_i \le 0.15\right). \tag{37}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function, which equals one when the relative prediction error does not exceed 15% and zero otherwise. A higher 15%-Acc indicates that more test prefixes are predicted within the accepted error range. Each run pools predictions over all evaluated battery-prefix pairs, with N denoting their total number. Metrics are first computed separately for each run, then reported as means and standard deviations across three runs. Multiple prefixes share a battery label but represent different observation lengths.

C.4 Baseline Details

iTransformer. iTransformer encodes the time series of each variable as a variate token and employs an inverted Transformer to model inter-variate dependencies for battery lifetime prediction (Liu et al., 2024).

DiffBatt. DiffBatt introduces diffusion modeling into battery degradation prediction and learns generative degradation patterns from early cycling observations for lifetime estimation (Eivazi et al., 2024).

BatLiNet. BatLiNet exploits relative degradation information across batteries and learns lifetime-related differences through inter-cell comparison under diverse aging conditions (Zhang et al., 2025).

LiPM. LiPM is a pretrained model for lithium-ion battery analysis that learns transferable degradation knowledge from large-scale battery data and adapts it to battery lifetime prediction (Li et al., 2025).

CPMLP. CPMLP employs CyclePatch to encode each cycle into a cycle-level feature and uses multilayer perceptrons to model cross-cycle information for battery lifetime prediction (Tan et al., 2025).

CPTransformer. CPTransformer applies Transformer layers to CyclePatch features to capture dependencies among early cycles for battery lifetime prediction (Tan et al., 2025).

CPGRU. CPGRU combines CyclePatch with gated recurrent units to model the ordered evolution of early-cycle information for battery lifetime prediction (Tan et al., 2025).

DITING. DITING constructs an early-cycle health reference and extracts weak degradation information from deviations relative to this reference for battery lifetime prediction (Miao et al., 2026).

C.5 Training Details

Architecture. For K patches, the 300 positions are extended by repeating the final position to length $K\lceil 300/K \rceil$. Each patch has $3\lceil 300/K \rceil$ input features. With hidden width d, the shared encoder, association scores, and marginal weights are

$$h_{c,i} = W_e x_{c,i} + b_e, \qquad A_c(i,j) = \frac{(W_q h_{c,i})^\top (W_k h_{c+1,j})}{\sqrt{d}}, \tag{38}$$

$$P_c(i) = (1 - K\epsilon)\,\mathrm{softmax}_i(w_p^\top h_{c,i}) + \epsilon, \qquad \epsilon = 10^{-4}. \tag{39}$$

Here, $W_q, W_k \in \mathbb{R}^{d\times d}$. Flattened projected conditionals form K^2 channels for a Conv1D layer with d outputs, kernel size 5, stride 1, and padding 2, followed by GELU and masked average pooling. The prediction head is Linear(d, d), GELU, and Linear$(d, 1)$, predicting standardized lifetime. The label scaler restores predictions to cycle units. Width is selected from Table 5. Counts refer to the selected configuration.

Variable-length inputs. For each prefix, a transition is valid only when both adjacent cycles are observed. Invalid positions are zeroed before convolution and masked again before pooling. The pooling denominator counts valid adjacent pairs. All methods receive the same prefix samples and observation masks.

Table 5: Hyperparameter search space and fixed settings used in the experiments.

Category	Hyperparameter	Candidate or fixed value
Optimization	Optimizer	Adam
Optimization	Training objective	MSE on standardized lifetimes
Optimization	Learning rate	$(10^{-4},\ 5\times10^{-4},\ 10^{-3},\ 5\times10^{-3})$
Optimization	Batch size	$\{16, 32, 64, 128\}$
Representation	Hidden dimension	$\{16, 32, 64, 128, 256\}$
Regularization	Dropout rate, where used	$\{0, 0.05, 0.1\}$
Input	Maximum prefix length	100 valid cycles
Input	Sampling positions per cycle	150 charge + 150 discharge
Input	Observed variables	Current, voltage, and capacity
LEAF	Number of within-cycle patches K	16
LEAF	Transition-encoder kernel size	5
Model selection	Selection metric	Validation MAPE
Reporting	Evaluation runs	3

Optimization. Training follows the BatteryLife pipeline with Adam and MSE loss. Lifetime labels are standardized using the mean and standard deviation of training-cell lifetimes. Validation and test labels use the same fitted scaler, and predictions are transformed back to cycle units before

computing MAPE and 15%-Acc. Epoch limits, early stopping, and learning-rate scheduling follow the benchmark training configuration. The checkpoint with the lowest validation MAPE is used for testing. More than 20 hyperparameter configurations are evaluated for each method, and validation MAPE determines the selected configuration. LEAF searches the four widths, four learning rates, and four batch sizes listed in Table 5. Dropout is tuned only for models that contain dropout layers. Baseline-specific components follow their released implementations, with common data partitions, input samples, and evaluation metrics. Figure 10 further illustrates the sensitivity of LEAF to key hyperparameters on five representative datasets, including batch size, learning rate, number of patches, and convolution kernel size.

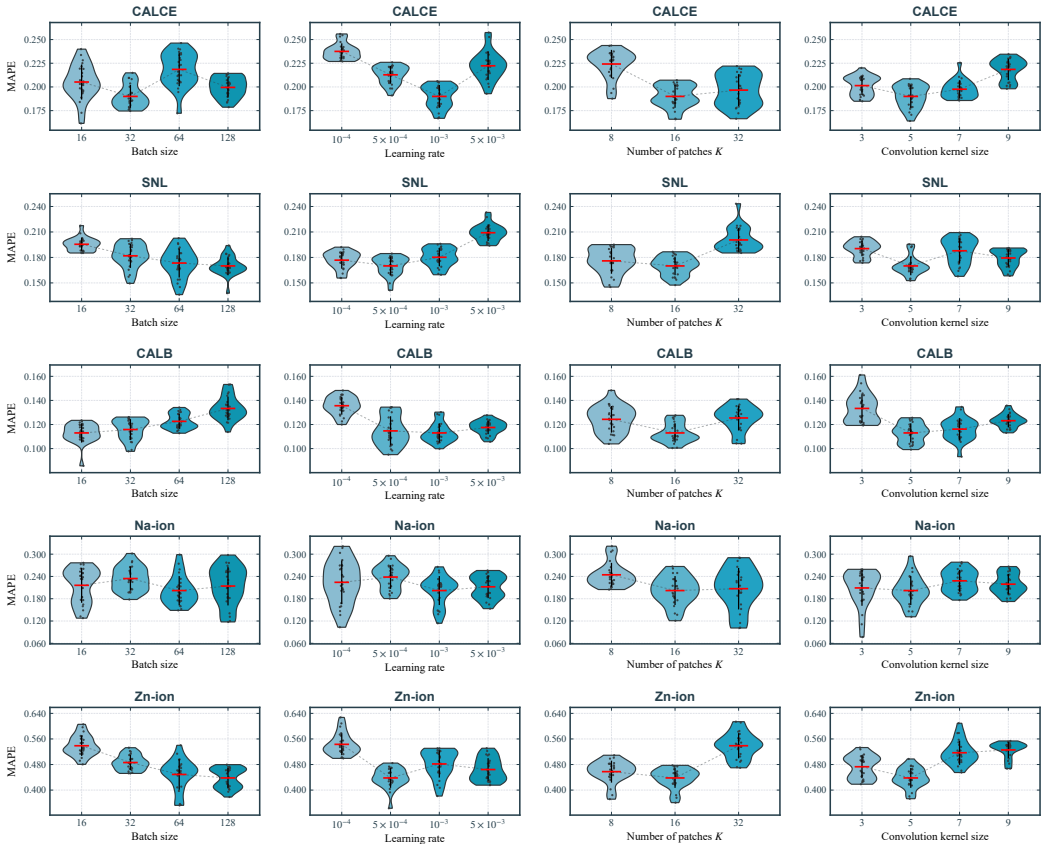


Figure 10: Hyperparameter sensitivity of LEAF on five representative battery datasets. Violin plots show the distribution of validation MAPE under different choices of batch size, learning rate, number of patches K, and convolution kernel size. Red markers indicate the mean performance, with error bars denoting one standard deviation.

Result aggregation. Results are reported as the mean and standard deviation across three runs. Each run computes MAPE and 15%-Acc over all prefix samples, including those pooled within the Li-ion domain. The Li-ion runs share the split defined by seed 2021. For each of CALB, Na-ion, and Zn-ion, the three runs use the splits defined by seeds 42, 2021, and 2024. The reported standard deviations summarize variation across these runs. Prefixes from one battery share its lifetime label and are evaluated within the same split.

Computational environment. All experiments were conducted on a server running Ubuntu 22.04.4 LTS, with 16 allocated Intel Xeon Gold 6430 CPU cores, 120 GB of system memory, and one NVIDIA GeForce RTX 4090 GPU with 24 GB of memory. Each training run used the single GPU. The models were implemented in Python 3.12.3 and PyTorch 2.4.1 with CUDA 12.1.

D Additional Experimental Results

D.1 Conditional-Transition Discrepancies across All Datasets

For cycles a, b, let JS use natural logarithms and define

$$d_m(a,b) = \mathrm{JS}(P_a, P_b), \qquad d_t(a,b) = \sum_i \frac{P_a(i) + P_b(i)}{2} \mathrm{JS}\big(\widehat{K}_a(\cdot \mid i), \widehat{K}_b(\cdot \mid i)\big). \tag{40}$$

Pairs $1 \le a < b \le C-1$ are averaged within each held-out battery and then across batteries with equal weight, while the gap profile groups pairs by $b-a$. These quantities describe learned patch-index distributions. The voltage-distribution visualizations instead use observation-level histograms with common bins determined from the training data.

Figure 11 extends the analysis in Figure 3 to all 16 battery datasets. Each point compares the marginal discrepancy between two cycle positions with the corresponding discrepancy between their adjacent-cycle conditional transitions. Across datasets, many cycle pairs lie above the diagonal, showing that conditional transitions often differ more strongly than the marginal cycle states. The magnitude of this separation varies with battery chemistry and operating condition, but the same contrast appears across the evaluated datasets.

Figure 12 examines the same quantities as the gap between cycle positions increases. Marginal discrepancies remain relatively small over much of the early-cycle window, whereas conditional-transition discrepancies are generally larger and tend to increase with cycle separation. Local fluctuations differ across datasets because of their cycling protocols and degradation characteristics. The gap profiles therefore provide a second view of the separation between cycle-wise state variation and cross-cycle conditional variation.

D.2 Fixed-Marginal Transition Structures across All Datasets

Figure 13 extends the fixed-marginal transition analysis in Figure 4 to all datasets. Each circular diagram connects source states in the current cycle with target states in the next cycle. Purple and green connections indicate associations enhanced and suppressed relative to the independence baseline, respectively, while link width represents the magnitude of the deviation. The outer bars show the marginal probabilities preserved by projection.

The projected transitions exhibit non-uniform source–target structures across all datasets. Because both adjacent-cycle marginals remain fixed, the enhanced and suppressed links correspond to entries of the dependence residual at the prescribed state compositions. Their row and column sums are zero under the marginal constraints. The detailed patterns differ across datasets, indicating that the learned cross-cycle dependence varies with battery system and aging condition.

D.3 Local Transition Contributions across All Datasets

Contributions are computed using absolute integrated gradients from the independence conditional $K_c^0(j \mid i) = P_{c+1}(j)$ to $\widehat{K}_c$, with 64 midpoint steps and fixed fitted parameters and marginals. The straight path preserves the prescribed joint marginals. Scores are normalized by their battery-level total plus 10^{-12}, averaged over valid adjacent-cycle pairs, and grouped into four consecutive source and target regions.

Figure 14 extends the local contribution analysis in Figure 6 to all battery datasets. The 16 local states are grouped into four source and target regions. Rows represent source groups in the current cycle and columns represent target groups in the subsequent cycle. Filled markers report the median contribution, while hollow markers report the 90th percentile.

Contribution scores are unevenly distributed across source–target pairs. Large values are concentrated in a subset of transition regions, and their locations differ across datasets rather than following a common within-cycle interval. The fitted predictors therefore emphasize different local associations under different battery systems and aging conditions.

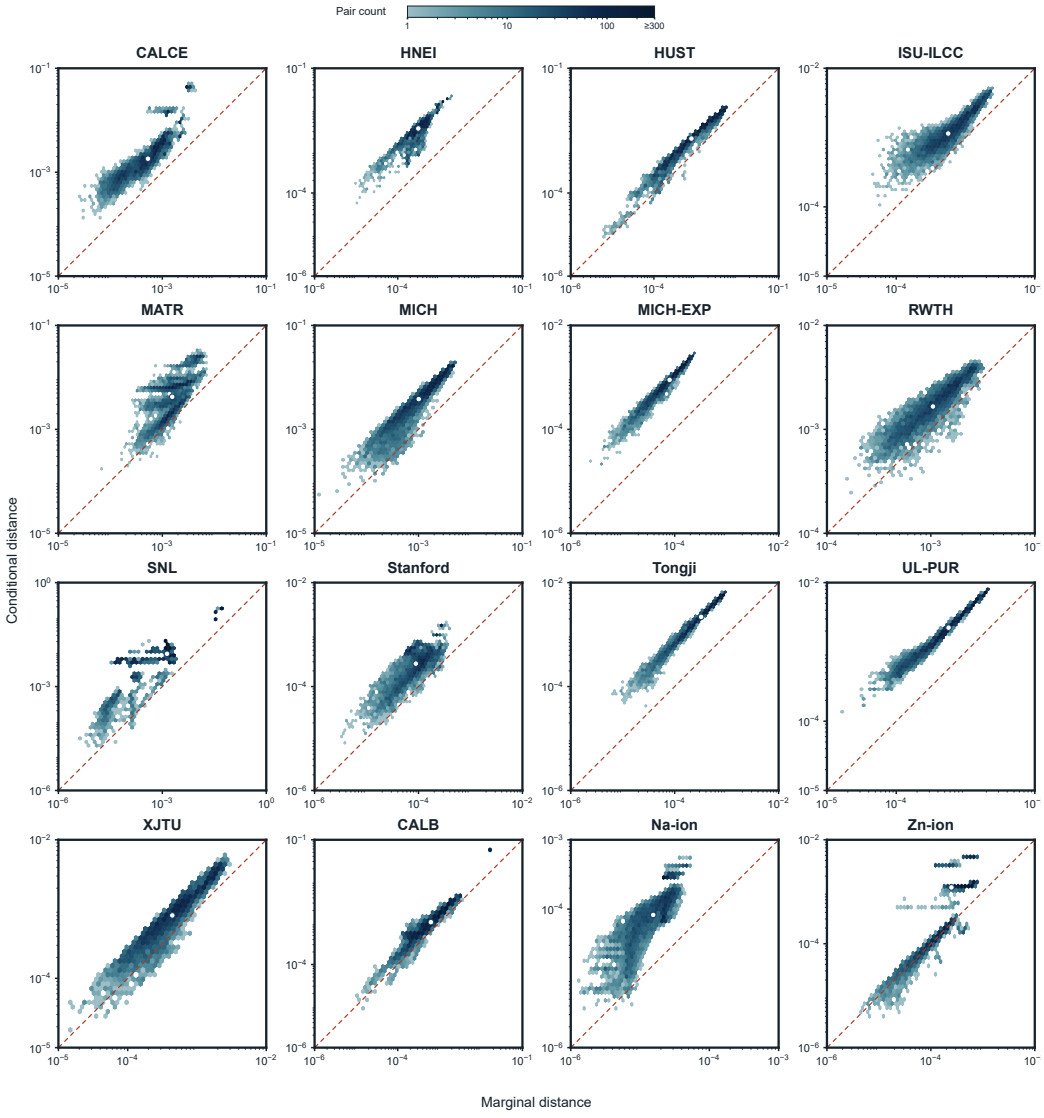


Figure 11: Marginal and conditional-transition discrepancies across all battery datasets. The horizontal and vertical axes denote marginal and conditional-transition distances, respectively; the dashed diagonal indicates equal discrepancies, and color represents pair density.

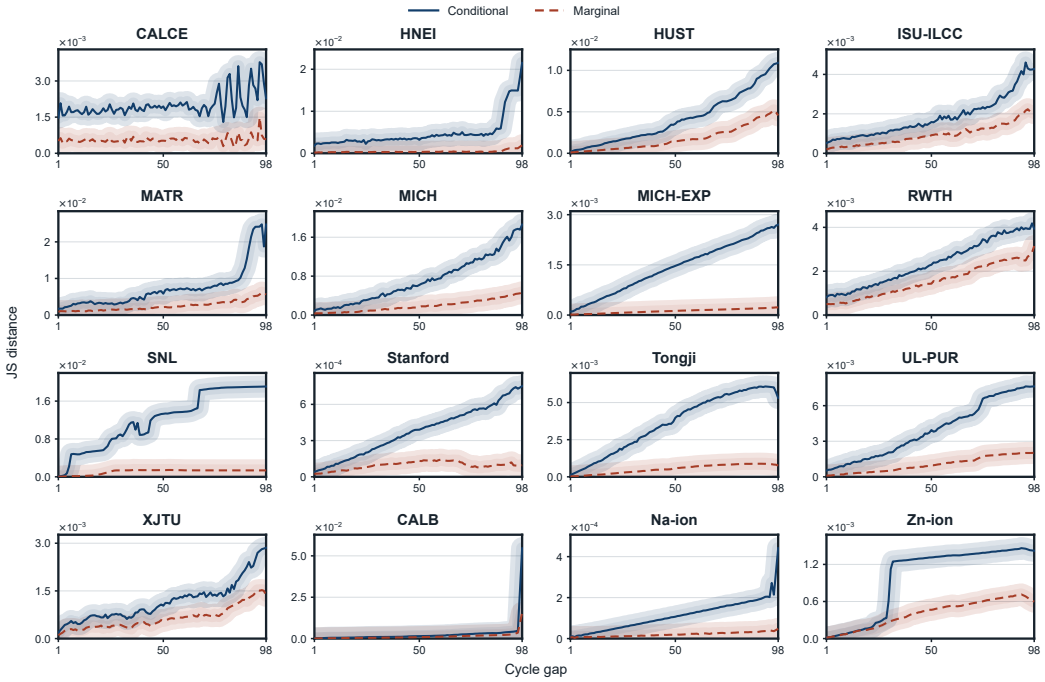


Figure 12: Marginal and conditional-transition JS distances across increasing cycle gaps for all battery datasets.

D.4 Cross-Domain Transferability

Li-ion is used as the source domain. Baseline trains the same architecture directly on the target-domain training set used in the main evaluation. Frozen evaluates the Li-ion pretrained model without target-domain updates. FT and DA use the first $\lceil n_{\text{train}}/2 \rceil$ target-domain batteries for adaptation while retaining the original validation and test sets. FT updates all model parameters, whereas DA additionally aligns source and target representations. Equal-sized source and target minibatches define

$$\mathcal{L}_{\text{DA}} = \tfrac{1}{2}(\mathcal{L}_s + \mathcal{L}_t) + \lambda \|\bar{z}_s - \bar{z}_t\|_2^2 / d, \tag{41}$$

where z denotes the pooled representation and $\lambda \in \{0.01, 0.1, 1\}$ is selected according to target validation MAPE. Target test inputs and labels are not used during adaptation.

Table 6 compares LEAF, CPMLP, and DITING under the four transfer settings. Direct transfer without adaptation produces substantial degradation, while FT and DA recover much of the target-domain performance. The effect of adaptation differs across battery domains and evaluation metrics, with DA providing the strongest improvement in several cross-domain settings.

Table 6: Cross-domain transfer comparison across three methods.

		Baseline		Frozen		FT		DA	
Target	Method	MAPE	15%-ACC	MAPE	15%-ACC	MAPE	15%-ACC	MAPE	15%-ACC
Zn-ion	**LEAF**	0.438	0.351	3.012	0.061	0.421	0.369	**0.374**	**0.423**
	CPMLP	0.523	0.311	2.684	0.083	0.497	0.327	0.454	0.358
	DITING	0.595	0.274	3.327	0.052	0.561	0.298	0.492	0.336
Na-ion	**LEAF**	0.202	**0.444**	3.864	0.018	0.217	0.417	**0.181**	0.432
	CPMLP	0.261	0.355	3.176	0.037	0.276	0.342	0.238	0.386
	DITING	0.289	0.364	4.263	0.011	0.301	0.353	0.254	0.397
CALB	**LEAF**	**0.113**	0.778	1.428	0.236	0.128	0.751	0.119	**0.802**
	CPMLP	0.142	0.687	1.736	0.191	0.151	0.672	0.136	0.701
	DITING	0.148	0.652	1.294	0.268	0.166	0.633	0.143	0.680

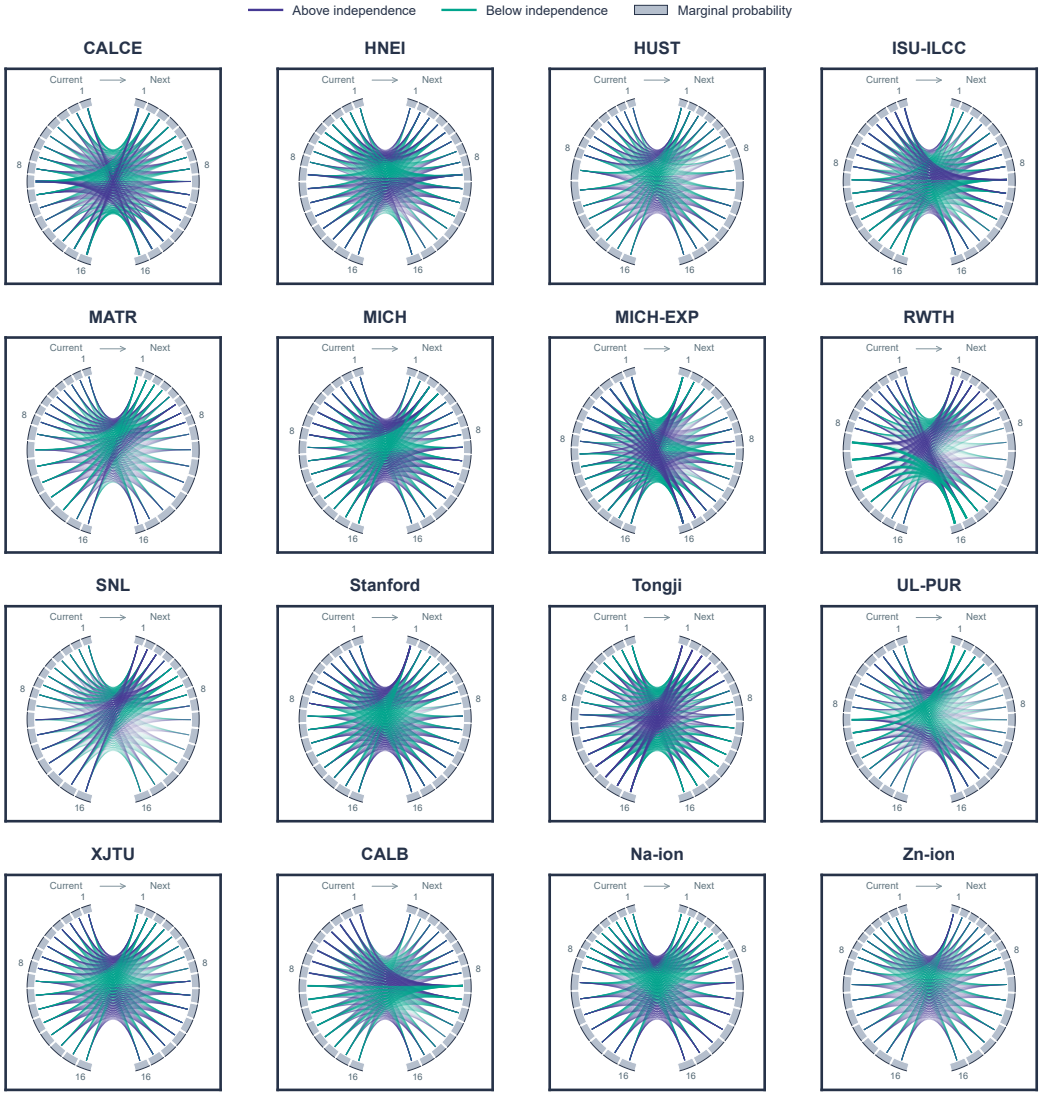


Figure 13: Fixed-marginal conditional-transition structures across all battery datasets. Purple and green links denote transitions enhanced and suppressed relative to independence, respectively, while outer bars indicate preserved marginal probabilities.

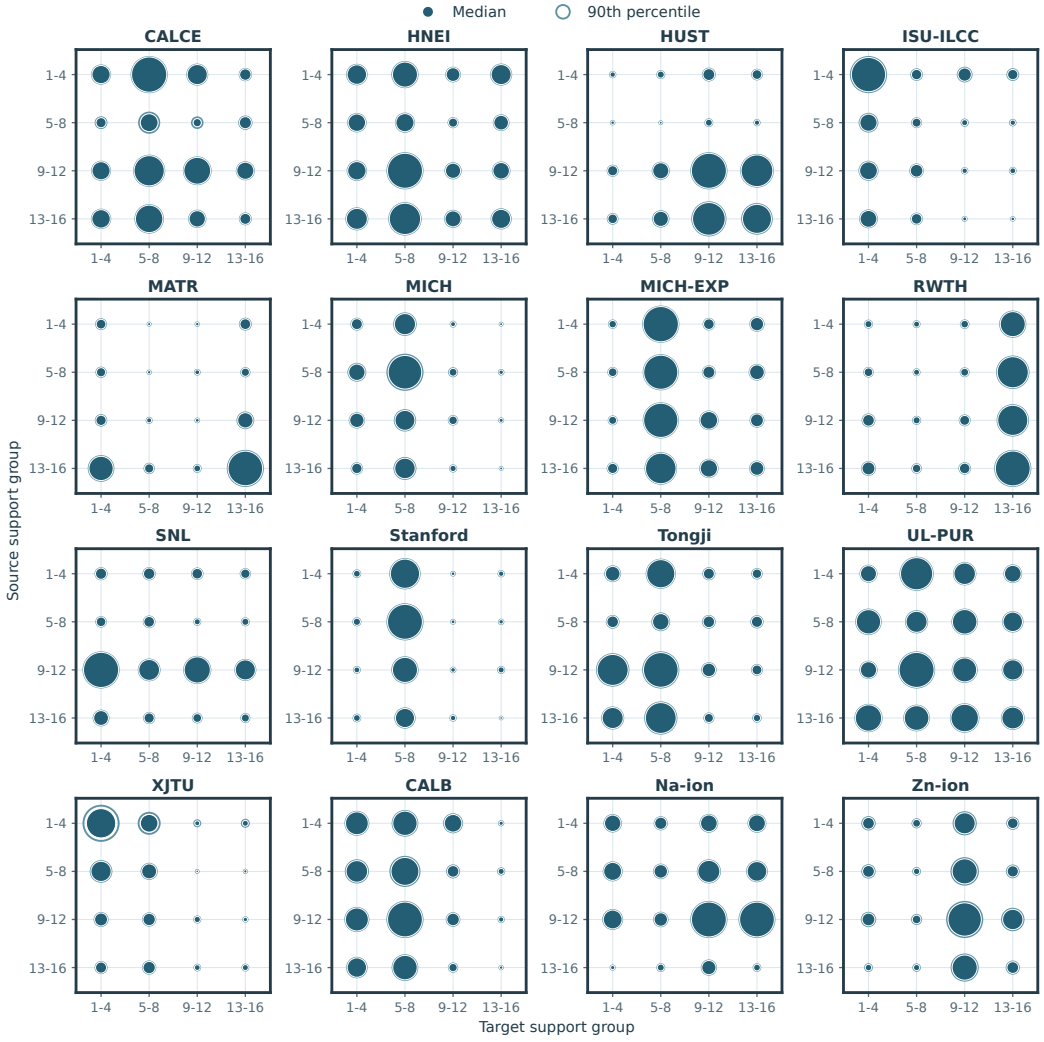


Figure 14: Local conditional-transition contributions across all battery datasets. Rows and columns denote source and target support groups, while marker sizes summarize contribution magnitude using the median and 90th percentile.

D.5 Performance Analysis on Seen and Unseen Aging Conditions

Test batteries are classified as Seen when their condition tuple occurs in the training set and Unseen otherwise. The tuple contains dataset, chemistry, temperature, charge/discharge protocol, cell format, and electrolyte where available; lifetime-derived attributes are excluded. The same fitted models and battery partitions are used for both groups, and each battery contributes equally within its group. Table 7 compares performance under seen and unseen aging conditions. Prediction errors increase on unseen conditions for all three methods, reflecting the additional difficulty introduced by changes in operating conditions and battery characteristics. LEAF maintains lower MAPE and higher 15%-ACC across the four battery domains.

Table 7: Performance under seen and unseen aging conditions.

		Seen		Unseen	
Domain	Method	MAPE	15%-ACC	MAPE	15%-ACC
Li-ion	**LEAF**	0.137	0.719	0.181	0.603
	CPMLP	0.163	0.661	0.217	0.519
	DITING	0.158	0.665	0.207	0.548
CALB	**LEAF**	0.104	0.801	0.146	0.654
	CPMLP	0.129	0.723	0.182	0.579
	DITING	0.136	0.681	0.186	0.529
Na-ion	**LEAF**	0.185	0.475	0.249	0.330
	CPMLP	0.243	0.326	0.336	0.000
	DITING	0.269	0.378	0.348	0.265
Zn-ion	**LEAF**	0.417	0.383	0.506	0.292
	CPMLP	0.498	0.335	0.617	0.224
	DITING	0.557	0.287	0.689	0.111

D.6 Independent Validation and Robustness Analysis

To further characterize stationarity masking and the predictive behavior of LEAF, we conduct complementary analyses from four perspectives, including observation-level distributions, paired performance differences, alternative design choices, and observation budgets, as summarized in Figure 15. We first compare marginal distribution discrepancies with null-adjusted conditional discrepancies directly at the observation level. As shown in the upper-left panel, most of the 16 dataset-level summaries exhibit conditional JS changes that are larger than the corresponding marginal JS distances. The magnitude of this separation varies across datasets. These observation-level results support the stationarity-masking pattern in which relatively small marginal changes coexist with more pronounced changes in adjacent-cycle conditional dependence.

The upper-right panel reports paired MAPE differences between LEAF and CPMLP, CPTransformer, and CPGRU, where positive values indicate lower prediction error for LEAF and the horizontal bars denote 95% paired battery-level bootstrap intervals. The mean differences remain positive across all four battery domains, supporting improvements at the battery level. The lower-left panel compares LEAF with direct adjacent-cycle differences, a KL-based projection variant, a soft marginal constraint, and a variant without projection. Relative to LEAF, direct adjacent differences increase MAPE by 9.8%–16.2% across the four domains, supporting the predictive advantage of the conditional-transition representation over direct numerical differences. Removing fixed-marginal projection increases MAPE by 10.2%, 13.3%, 32.2%, and 7.8% on Li-ion, CALB, Na-ion, and Zn-ion, respectively, as calculated from the mean MAPE values in the ablation table. The soft-constraint variant also yields higher errors across all four domains, while the KL-based projection variant remains closer to LEAF. Together, these comparisons support the contributions of conditional-transition modeling and fixed-marginal projection within the evaluated architecture.

The lower-right panel evaluates test MAPE under observation budgets of 20, 40, 60, 80, and 100 cycles. Prediction errors generally decrease as additional early-cycle observations become available, although some intermediate budgets show fluctuations. LEAF maintains lower test MAPE than the corresponding baseline across the evaluated battery domains and observation lengths. These results indicate that its predictive advantage is present with short observation windows and persists as the observation budget increases.

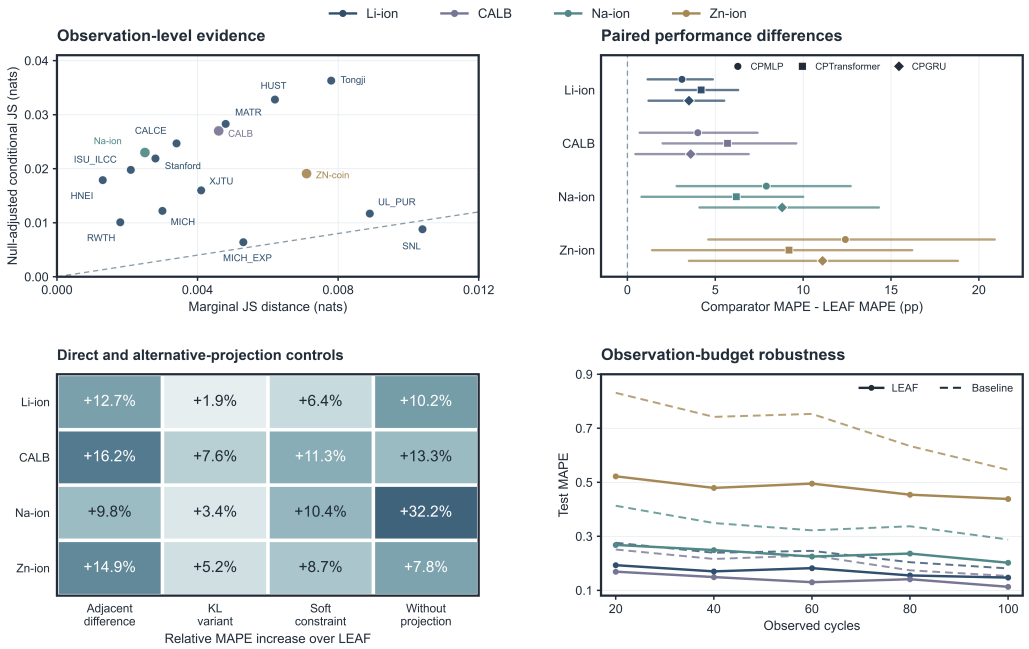


Figure 15: Independent validation and robustness analysis of LEAF. The upper-left panel compares marginal JS distance with null-adjusted conditional JS at the observation level across 16 dataset-level summaries. The upper-right panel reports paired MAPE differences between representative baselines and LEAF, where positive values indicate lower error for LEAF and horizontal bars denote 95% paired battery-level bootstrap intervals. The lower-left panel compares direct adjacent-cycle differences, a KL-based projection variant, a soft marginal constraint, and removal of projection, with values reporting the relative MAPE increase over LEAF. The lower-right panel reports test MAPE under observation budgets from 20 to 100 cycles.

E LIMITATIONS

LEAF models local associations between adjacent cycles and captures their ordered evolution over the early observation window. The current evaluation covers diverse battery systems, aging conditions, and testing environments, while broader industrial applications may involve degradation patterns not represented in the available data. LEAF also focuses on adjacent-cycle conditional transitions, leaving longer-range dependencies across multiple cycles unexplored. Cross-domain results further show that source pretraining alone does not guarantee improved target-domain performance, and transfer effectiveness remains dependent on the target battery system and its data distribution. Future work will extend conditional-transition modeling to longer temporal ranges and evaluate its transferability and generalization under broader operating conditions and industrial settings.