

Article

Retrieval-Augmented Degradation Priors for Data-Efficient Early Lithium-Ion Battery Lifetime Prediction

Yuelin Zou ^{1,2,*}, Yajun Zhang ³ and Ke Lv ⁴¹ School of Computer Science and Technology, Xinjiang University, Urumqi 830046, China² State Grid Xinjiang Electric Power Co., Ltd., Information and Communication Company, Urumqi 830046, China³ School of Software Engineering, Xinjiang University, Urumqi 830091, China⁴ School of Engineering Science, University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: zouyuelin@stu.xju.edu.cn

Abstract

Early lithium-ion battery lifetime prediction is difficult because only a small fraction of a cell's lifetime is observed when a prediction is needed. We evaluate a transparent, validation-tuned combination of a supervised early-cycle predictor and a nearest-neighbor retrieval estimate built from similar historical cells. The method is assessed against supervised-only and retrieval-only baselines over 20 repeated cell-wise splits. On the Severson/Toyota Research Institute/Massachusetts Institute of Technology (Severson/TRI/MIT) dataset (124 lithium iron phosphate/graphite cells), fusion reduces the mean absolute error (MAE) of end-of-life (EOL) prediction by 10.08 cycles at 10 observed cycles and 11.90 cycles at 20 observed cycles relative to the validation-selected supervised baseline; the confidence intervals for these gains exclude zero, whereas gains at 50 and 100 cycles are unsupported. Retrieval-only prediction is competitive in the shortest windows. In a targeted 40-split external confirmation on BatteryLife-XJTU (Xi'an Jiaotong University) at 50 observed cycles, fusion reduces mean MAE from 34.11 to 28.02 cycles (paired bootstrap 95% confidence interval for the gain: 0.47–11.41 cycles; one-sided Wilcoxon $p = 0.007$). The results support retrieval as an interpretable complement to supervised prediction in compatible weak-signal regimes, not as a universal replacement for supervised models.

Keywords: lithium-ion batteries; battery lifetime prediction; state of health; remaining useful life; retrieval-augmented forecasting; in-context learning; data-efficient learning; time-series forecasting

1. Introduction

Battery lifetime prediction has greatest practical value before a cell has visibly aged. Fast-charging protocol screening, manufacturing triage, warranty planning, and second-life selection all require early rankings, yet the end-of-life (EOL) labels used to train and validate such rankings are available only after long cycling experiments [1]. In the Severson/Toyota Research Institute/Massachusetts Institute of Technology (TRI/MIT) dataset, commercial lithium iron phosphate/graphite cells reach EOL after roughly hundreds to more than two thousand cycles [2]. A prediction after 10, 20, or 50 cycles is therefore made from only a small fraction of the cell's observable lifetime.

Early cycle-life prediction has often relied on a single-cell mapping from early features to lifetime, without reference to any other battery's degradation history. Severson et al. showed that early-life information can predict lifetime before obvious capacity fade occurs,



Received: 3 June 2026

Revised: 16 July 2026

Accepted: 26 July 2026

Published: 4 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](#)[Attribution \(CC BY\)](#) license.

particularly when voltage-curve and discharge-capacity features are constructed carefully [2]. Gaussian process regression has been used for uncertainty-aware SOH forecasting, where calibrated predictive uncertainty can support maintenance and decision-making [3], and tree ensembles including random forests, XGBoost, and LightGBM capture nonlinear interactions among early-cycle features without requiring large neural architectures [4–6]. These methods are transparent and compatible with modest data, but they carry no prior beyond what is encoded in the target cell’s own early trajectory. When only a handful of fully aged support cells is available, or when the target cell has changed little over the first tens of cycles, the fitted mapping from early features to EOL can be unstable.

Recurrent neural networks and related architectures have been widely applied to battery SOH and RUL estimation [7–9], while particle filters, Bayesian updating, and hybrid empirical formulations infer RUL by propagating uncertainty through a single cell’s partial degradation history [10–12]. Physics-informed and model-driven approaches increasingly combine electrochemical or degradation constraints with data-driven SOH estimation, including across temperatures and operating conditions [13–16]. Attention-based RUL prediction, sequence-to-sequence forecasting, random-forest variants, autoregressive models with exogenous variables, hybrid decomposition networks, and patch-based Transformer architectures provide further model classes [17–28]. These approaches can introduce useful physical structure or model capacity, but early-window prediction remains limited by the amount and type of cell-specific evidence available for the target task.

A third strand does draw on other cells, but folds that information into learned parameters rather than exposing it as retrievable evidence. BatLiNet performs inter-cell learning across aging conditions, and BatteryLife aggregates multiple battery-life datasets to benchmark modern time-series methods across formats, chemistries, temperatures, and protocols [29,30]; battery-specific time-series foundation models and language-model-based health forecasting extend the same movement from handcrafted features toward general, pretrained sequence representations [31,32]. In these methods, cross-cell knowledge is compressed into weights during training, so a prediction cannot be traced back to the specific historical cells that support it. Review articles now document a broad SOH/RUL literature spanning all three strands, which makes it increasingly important for new studies to state the model family, data regime, forecast horizon, and prediction target under evaluation [8,33].

Pretrained time-series models add a useful but incomplete capability. Chronos tokenizes time series and learns a general forecasting model from large collections of temporal data [34]; TimesFM uses a decoder-only foundation-model design for zero-shot time-series forecasting [35]. Both can forecast local capacity evolution without retraining on each battery dataset, but their prior comes from time series in general rather than from battery degradation in particular, and EOL prediction is a threshold-crossing problem over a horizon that may extend hundreds of cycles beyond the observed prefix. For long-life cells, the 80% SOH threshold often lies outside the fixed forecast horizon, and the same limitation applies to the battery-specific Transformer and foundation-model studies above: improved trajectory or SOH forecasting still requires additional assumptions before it becomes long-horizon EOL inference [24–26,31,32].

This study is not presented as a new deep-learning or Bayesian algorithm. It evaluates whether a conventional nearest-neighbor analogue estimate, made explicit at prediction time, can add useful historical-cell information to a validation-selected supervised predictor in the first tens of cycles. The terminology “retrieval-based degradation prior” refers to this operational role: retrieved cells provide an empirical distribution of later trajectories and EOL values conditional on early-shape similarity. It does not imply a formal Bayesian prior or an in-context foundation model.

The central hypothesis in this paper is that similar historical cells can serve as an explicit degradation prior when the query cell's early trajectory is only weakly informative. We refer to this component as a retrieval-augmented degradation prior. A query cell is represented by early capacity shape and simple degradation statistics; support cells with similar representations form the local context, and their downstream lifetimes and future trajectories define a non-parametric prior over plausible degradation paths. The context is therefore not an abstract latent state but a retrieved set of real cells with observed aging histories.

HybridSupervisedRetrieval combines this prior with a supervised early-life predictor, and the two components earn their place for different reasons: the supervised model captures dataset-specific feature-to-lifetime relationships that retrieval alone cannot learn, while retrieval contributes full-horizon degradation experience from historical analogs that no amount of early-cycle feature engineering can supply. Validation-tuned fusion determines the relative weight of the two sources. For a target cell i , the final prediction is

$$\hat{y}_i = \alpha \hat{y}_i^{\text{sup}} + (1 - \alpha) \hat{y}_i^{\text{ret}}, \quad (1)$$

where $\hat{y}_i^{\text{sup}}$ is produced by a candidate supervised model, $\hat{y}_i^{\text{ret}}$ is produced by a candidate retrieval configuration over similar support cells, and α is selected on a validation set. The hybrid search evaluates the supervised learner, retrieval configuration, and fusion weight jointly by validation mean absolute error (MAE). Table 1 reports the independently selected supervised baseline, whereas the later hybrid comparison reports the selected fused predictor; α is interpreted as the selected mixture weight rather than as evidence that either information source dominates across all splits.

The analysis workflow is summarized in Figure 1.

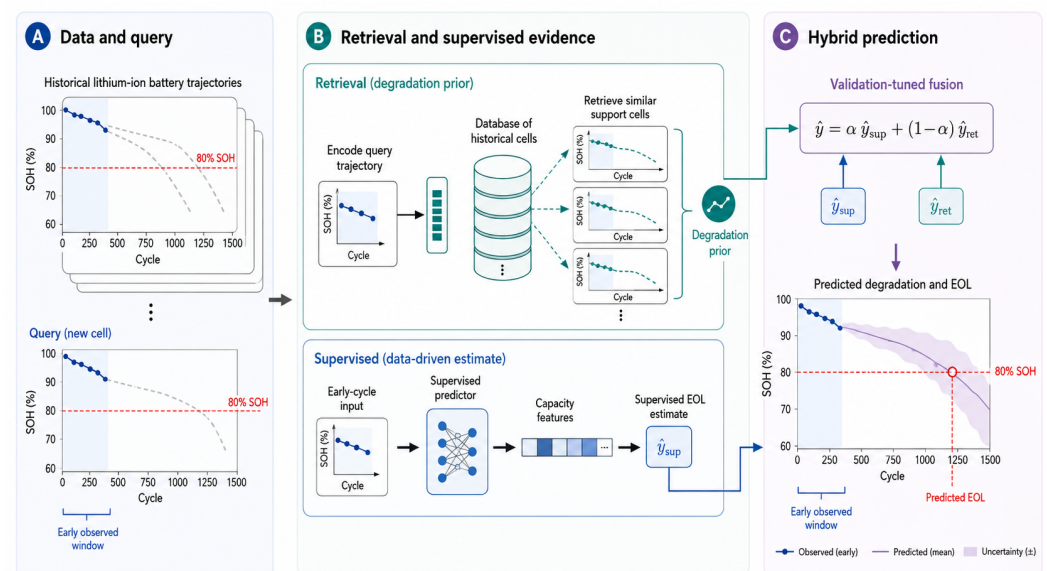


Figure 1. Overview of the proposed retrieval-augmented early battery lifetime prediction framework. The target cell's early capacity window is used both by a supervised predictor and by a retrieval module that identifies similar support cells. Retrieved support cells provide a degradation prior, while validation-tuned fusion combines supervised and retrieval-based evidence to produce the final end-of-life (EOL) prediction.

The contributions of this work are as follows:

- We formulate early lithium-ion battery lifetime prediction as a data-efficient prior-integration problem in which the first 10–100 cycles are used to predict full EOL.

- We evaluate retrieval-augmented degradation priors through retrieval-only prediction, supervised prediction, and validation-tuned supervised–retrieval fusion.
- We benchmark Ridge regression, random forests, ExtraTrees, SVR, XGBoost, LightGBM, naive extrapolation, Chronos-Bolt zero-shot forecasting, repeated-split retrieval-only prediction, and the proposed hybrid method on a processed Severson/TRI/MIT table with 124 cells and 100,460 cycle records.
- We quantify robustness over 20 repeated cell-wise splits, showing that the fused predictor provides its clearest MAE reduction over the supervised baseline at 10 and 20 observed cycles, while longer windows show weaker or unsupported gains.
- We conduct a compatibility-filtered external screen using BatteryLife-SNL, BatteryLife-HUST, BatteryLife-XJTU, and BatteryLife-CALB to identify datasets and windows where explicit retrieval priors or fusion provide additional evidence.

Table 1. Validation-selected supervised end-of-life (EOL) prediction baseline over 20 repeated cell-wise splits. The selected supervised model is chosen separately within each split by validation mean absolute error (MAE).

Observed Cycles	MAE (Cycles)	RMSE (Cycles)	Spearman
10	170.69 ± 51.91	279.62	0.68
20	183.82 ± 48.64	296.65	0.68
50	181.43 ± 56.87	293.61	0.68
100	172.40 ± 41.74	274.11	0.72

2. Materials and Methods

2.1. Task Definition

Let $Q_i(t)$ denote the discharge capacity of cell i at cycle t . Normalized capacity is used as state of health (SOH):

$$\text{SOH}_i(t) = \frac{Q_i(t)}{Q_i(1)}. \quad (2)$$

The end-of-life cycle is the first cycle at which normalized capacity falls below the 80% threshold:

$$\text{EOL}_i = \min\{t : \text{SOH}_i(t) \leq 0.8\}. \quad (3)$$

The main task is early EOL regression. Given only the first N cycles,

$$\mathbf{x}_{i,N} = [\text{SOH}_i(1), \dots, \text{SOH}_i(N)], \quad (4)$$

the model predicts EOL_i . We evaluate $N \in \{10, 20, 50, 100\}$ to cover four early-use regimes: near-immediate screening after formation-like cycling, short early-life assessment, mid-early protocol evaluation, and a longer but still incomplete observation window. Here, “early aging” denotes the observed prefix used for prediction; it does not denote a distinct aging mechanism or a replacement for conventional SOH/RUL prediction. We also evaluate fixed-horizon capacity trajectory forecasting, in which the model predicts $\text{SOH}_i(N+1 : N+H)$. Trajectory forecasting tests near-future capacity prediction, whereas EOL regression tests long-range threshold-crossing inference from a limited prefix.

2.2. Datasets, Cleaning, and Experimental Splits

The primary dataset is the public Severson/TRI/MIT battery cycle-life dataset [2], which contains commercial lithium iron phosphate/graphite cells cycled under fast-charging conditions. We processed the three MATLAB-format source files batches following the cleaning logic of the public Braatz group loader. The procedure removes known abnormal cells, corrects cells continued across batches, filters invalid zero-capacity obser-

variations, and reconstructs valid cycle indices. The processed table contains 124 cells and 100,460 cycle-level records with cell identifier, cycle number, raw cycle, discharge capacity, normalized capacity, charge capacity, internal resistance, temperature, protocol metadata, source batch, and EOL cycle.

All preprocessing for the main Severson experiments is completed before feature construction. Model selection and fusion-weight selection are then performed only within the training portion of each split. Splits are cell-wise, so every cycle from a held-out test cell is excluded from model fitting, validation, retrieval support construction, standardization fitting, hyperparameter selection, and fusion-weight selection. Training cells are divided into a fit subset and a validation subset. The fit subset trains candidate supervised models and constructs candidate retrieval support sets; the validation subset selects the supervised model, retrieval feature set, neighbor count, aggregation rule, target transform, and fusion weight. Held-out test cells are used only for final evaluation. Because all methods use the same processed cell table, performance differences reflect model behavior rather than different cleaning rules.

To estimate split sensitivity, the main supervised, retrieval-only, and hybrid comparisons are repeated over 20 random cell-wise fit/validation/test splits. Each split contains 86 fit cells, 19 validation cells, and 19 test cells. For each early window and split, the supervised baseline is selected by validation MAE from the candidate supervised models. The retrieval-only baseline is selected by validation MAE from candidate retrieval configurations and then evaluated on the held-out test cells using the fit and validation cells as support. The hybrid configuration is selected independently by validation MAE over candidate supervised learners, retrieval configurations, and fusion weights. The selected supervised and hybrid predictors are then evaluated on the same held-out test cells, allowing paired split-level comparison of MAE. We report mean and standard deviation over the 20 splits, median split-level MAE gain, bootstrap 95% confidence intervals for the split-level MAE gain, and a one-sided Wilcoxon signed-rank test on the split-level MAE gains.

An external sanity check uses data from a previous GPDM-related battery degradation project, containing NASA and Oxford cells [36,37]. The external set includes 12 cells and 1700 cycle-level records. Differences in protocol, sampling, chemistry, and lifetime definition make this set a domain-shift sanity check rather than evidence of solved cross-dataset generalization. A small external set can reveal method collapse under protocol shift, but it cannot establish universal cross-dataset robustness or a definitive ranking among methods.

The primary dataset and available early observation windows are summarized in Figure 2.

We additionally perform a compatibility-filtered external screen using processed BatteryLife datasets [30]. BatteryLife-SNL, BatteryLife-HUST, BatteryLife-XJTU, and BatteryLife-CALB are converted to the same long-format capacity table used by the primary experiments. The screen includes datasets with at least 20 cells and enough complete early-cycle observations for the selected window. For each dataset and window, we evaluate four candidate families under repeated cell-wise splits: validation-selected supervised prediction, validation-selected retrieval-only prediction, validation-tuned supervised–retrieval fusion, and a validation-selected selector that chooses among the three families by validation MAE. Initial screens use 20 splits; the SNL 50-cycle and XJTU 20/50-cycle signals are independently extended to 40 non-overlapping splits and assessed with paired bootstrap confidence intervals and one-sided Wilcoxon tests. The external screen is used to identify compatible regimes where the retrieval prior is informative; it is not interpreted as a claim of uniform cross-dataset dominance.

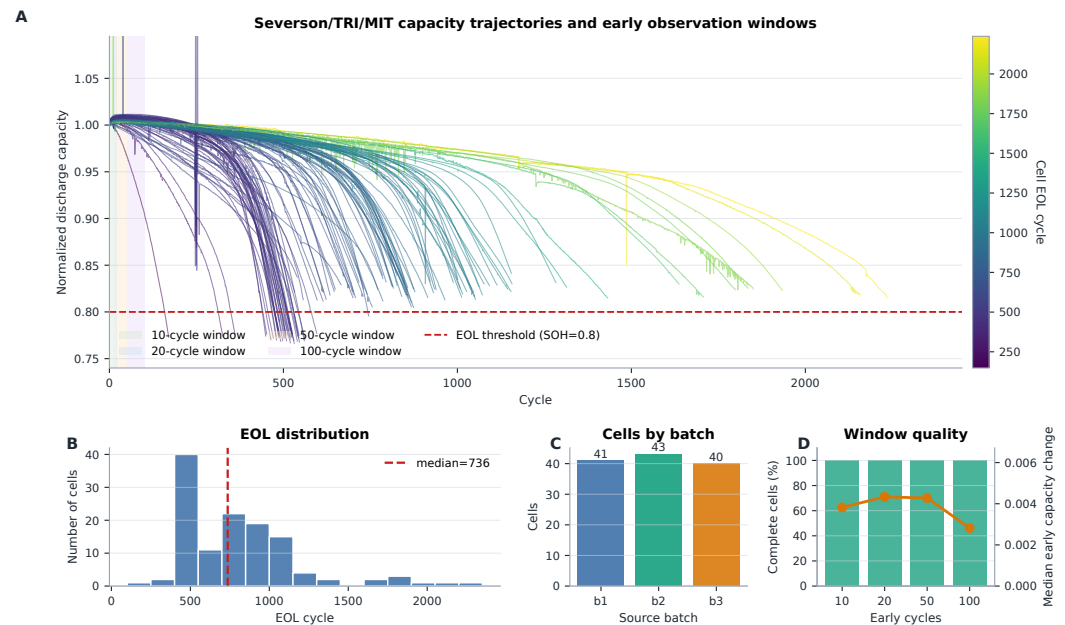


Figure 2. Severson/TRI/MIT capacity trajectories, EOL distribution, source batches, and early observation windows: **(A)** capacity trajectories with early observation windows and the EOL threshold; **(B)** distribution of EOL values; **(C)** numbers of cells by source batch; and **(D)** usable early-window coverage. The main experiments use only the first 10, 20, 50, or 100 cycles of each test cell as input, while EOL is defined by the later 80% SOH threshold crossing.

2.3. Supervised Early-Life Predictors

For each early window, features are constructed from the normalized capacity sequence. The feature set contains the raw early capacity values, mean, standard deviation, minimum, maximum, last value, first-to-last change, linear slope, quadratic trend coefficient, area under the early capacity curve, first-half versus second-half change, and first-difference statistics. We deliberately use capacity-only predictors for the cross-dataset comparison because capacity is available in all harmonized public tables. The processed Severson table also contains internal resistance, temperature, protocol metadata, and source batch, but voltage-derived and differential-capacity features are not harmonized across the external datasets and are therefore outside the present comparison. This design assesses the added value of historical-cell retrieval under a portable input definition; it is not a claim that capacity-only features are the strongest possible battery-lifetime feature set.

We compare Ridge regression, k -nearest neighbors, random forest, ExtraTrees, gradient boosting, support vector regression (SVR), XGBoost, and LightGBM [4–6]. The set covers regularized linear regression, local non-parametric regression, bagged tree ensembles, boosted tree ensembles, and kernel regression, which are common regimes for small and medium battery-health datasets. In the repeated-split comparison, the supervised baseline is selected by validation MAE separately within each split and early window.

2.4. Retrieval-Based Degradation Prior

The context component is a non-parametric retrieval model. The target cell is treated as a query, and fit cells serve as support examples. Validation cells are used only to choose configurations; test cells never enter the support database. For a target cell i , a retrieval feature vector $\mathbf{z}_{i,N}$ is constructed from early capacity shape and degradation statistics. Standardization parameters are fitted on the corresponding fit support set and then applied to validation or test queries. Euclidean distances are computed after standardization:

$$d(i, j) = \|\text{standardize}(\mathbf{z}_{i,N}) - \text{standardize}(\mathbf{z}_{j,N})\|_2. \quad (5)$$

The nearest k support cells define the local context set $\mathcal{N}_k(i)$. Retrieval EOL prediction is a distance-weighted aggregation of support-cell EOL labels:

$$\hat{y}_i^{\text{ret}} = \sum_{j \in \mathcal{N}_k(i)} w_{ij} \text{EOL}_j, \quad \sum_{j \in \mathcal{N}_k(i)} w_{ij} = 1. \quad (6)$$

The retrieval estimate is thus a validation-selected nearest-neighbor analogue model: query encoding, nearest-neighbor context selection, and distance-weighted label aggregation. We evaluate five prespecified retrieval configurations covering feature sets, 3–15 neighbors, mean/median/trimmed aggregation, temperature-scaled distance weighting, and optional log-target prediction. For trajectory forecasting, retrieved support trajectories are aligned by cycle index and aggregated to produce a mean trajectory and empirical 10th–90th percentile bands. These bands describe dispersion among retrieved analogues; they are not calibrated probabilistic prediction intervals. Each prediction remains traceable to a small set of historical cells.

2.5. HybridSupervisedRetrieval

The proposed hybrid method combines supervised and retrieval predictions:

$$\hat{y}_i = \alpha \hat{y}_i^{\text{sup}} + (1 - \alpha) \hat{y}_i^{\text{ret}}, \quad (7)$$

where $\alpha \in [0, 1]$ is selected by validation MAE. In each split, candidate supervised models and retrieval configurations are fitted on the fit subset. Their validation predictions are then used to select the fusion weight. After the supervised model, retrieval configuration, and α are fixed, the selected configuration is evaluated on the held-out test set without test-set feedback. The same equation allows the information source to change with the early window: retrieval can dominate when N is small, while supervised features can receive more weight when more of the target trajectory has been observed.

The fusion weight also provides a direct test of the proposed mechanism. If retrieval behaves only as a weak nearest-neighbor baseline, validation should assign it little weight and the hybrid prediction should approach the supervised estimate. If historical analogs carry useful information when early-cycle supervised features are weak, validation should select a smaller α and reduce test MAE. The weight trend is therefore interpreted together with the error metrics.

2.6. Chronos-Bolt Zero-Shot Forecasting Baseline

Chronos-Bolt tiny is evaluated as a pretrained time-series foundation-model baseline [34]. Given the first N normalized capacity values, the model produces zero-shot probabilistic forecasts for fixed horizons. We test horizons of 64 and 256 cycles and report trajectory RMSE, trajectory MAE, prediction interval coverage probability at 80% (PICP80), and mean prediction interval width (MPIW80). Because fixed-horizon forecasts often do not reach the EOL threshold for long-life cells, Chronos-Bolt is used mainly as a trajectory forecasting baseline rather than as the main full-horizon EOL method. Accurate local capacity forecasts do not by themselves determine a threshold-crossing cycle that lies far beyond the forecast horizon.

2.7. Evaluation Metrics

For EOL regression, we report mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), coefficient of determination (R^2), and Spearman rank correlation. MAE is emphasized because it measures cycle-level EOL error directly. Spearman correlation is included because many screening applications depend on ranking cells rather than predicting each EOL cycle exactly.

For capacity forecasting, we report trajectory MAE, trajectory RMSE, PICP80, and MPIW80. For retrieval trajectories, PICP80 is a diagnostic of the empirical 10th–90th percentile bands, not evidence of calibrated predictive uncertainty.

3. Results

3.1. Early-Window Baselines and Forecasting Gap

The supervised baselines provide the reference point for the retrieval experiments. Early normalized-capacity measurements already carry lifetime information, but the model selected by validation changes across splits and observed windows. As shown in Table 1, the validation-selected supervised baseline gives mean MAE values of 170.69, 183.82, 181.43, and 172.40 cycles at 10, 20, 50, and 100 observed cycles. The large split-to-split standard deviations confirm that cell-wise partitioning materially affects the measured error, which motivates the repeated-split hybrid comparison below.

The same dependence on labeled lifetime data appears in the auxiliary few-shot supervised experiments, which are reported as diagnostic single-split sweeps rather than as repeated-split significance tests. At a 50-cycle input window, the strongest ExtraTrees few-shot baseline gives MAE values of approximately 262.38, 235.90, 215.39, and 193.65 cycles when trained with 5, 10, 20, and 50 labeled cells. Expanding the labeled set from 5 to 50 cells reduces MAE by 68.73 cycles, so the supervised mapping remains sensitive to the number of fully aged cells available for fitting.

The same diagnostic status applies to long-horizon extrapolation from early capacity values. A simple linear-persistence extrapolation from 50 and 100 observed cycles gives EOL MAE values of 1636.56 and 1562.07 cycles. The error arises because many early normalized-capacity trajectories are nearly flat; extending a small local slope to the 80% SOH threshold can place the predicted crossing far beyond the observed lifetime range.

Chronos-Bolt is also reported as an auxiliary trajectory baseline rather than as part of the repeated-split EOL significance analysis. For a 64-cycle forecast horizon on the Severson dataset, trajectory RMSE values are 0.00455, 0.00355, 0.00183, and 0.00199 at 10, 20, 50, and 100 observed cycles, with PICP80 values of 0.952, 0.900, 0.895, and 0.746. These forecasts indicate that the pretrained model can represent smooth local degradation without battery-specific training. At a 256-cycle horizon, however, both accuracy and coverage degrade, and many forecasts still do not reach the 80% SOH threshold. Accurate fixed-horizon capacity prediction therefore does not by itself produce full-horizon EOL inference.

3.2. Retrieval as a Historical-Cell Prior

Retrieval uses the future trajectories and EOL labels of similar support cells to supply information that is unavailable from the target prefix alone. Table 2 reports the repeated-split retrieval-only baseline, with the retrieval configuration selected by validation MAE within each split. Retrieval EOL MAE is 164.26 ± 54.07 , 168.54 ± 62.76 , 182.76 ± 55.87 , and 181.02 ± 53.45 cycles at 10, 20, 50, and 100 observed cycles, respectively. The retrieval-only baseline is therefore competitive with the supervised baseline in the shortest windows, supporting the premise that historical-cell analogs carry useful early-life information. It also sets a stricter reference point for the hybrid method: the hybrid comparison in Table 3 should be read as improvement over validation-selected supervised prediction, not as a claim that fusion always dominates retrieval-only prediction.

The repeated-split diagnostics in Figure 3 show that the selected feature set and neighbor count vary across early windows. Retrieval is therefore used as a validation-tuned empirical prior rather than as a fixed nearest-neighbor comparator. Its prediction can also be inspected through the retrieved cells and the empirical uncertainty bands formed from their trajectories.

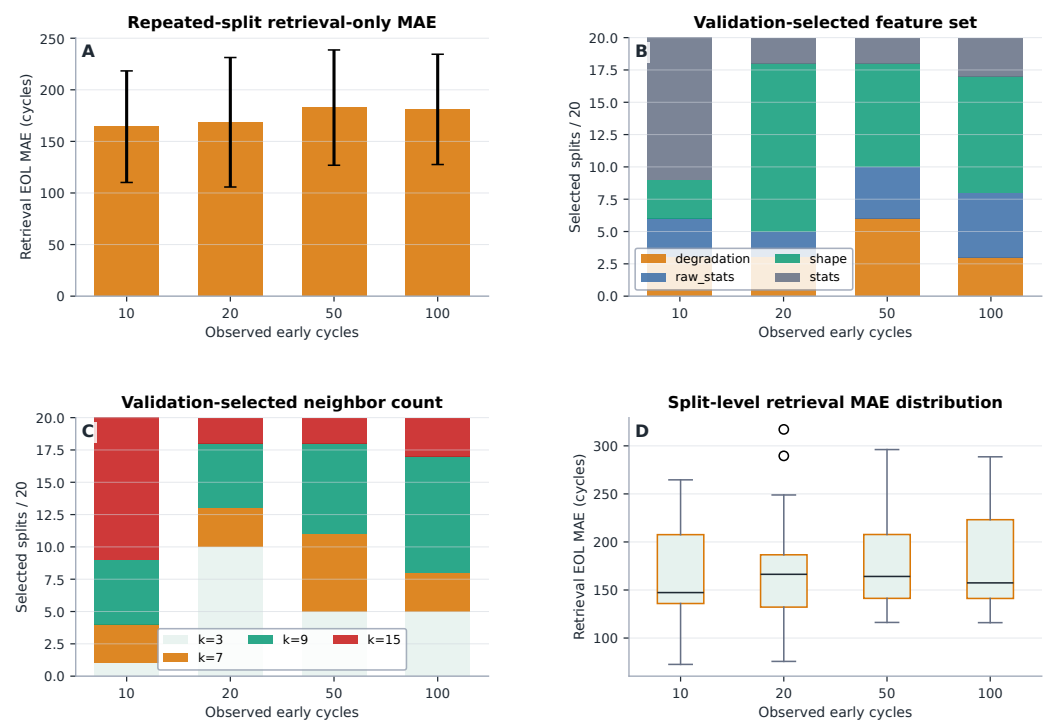
Table 2. Validation-selected retrieval-only EOL prediction over 20 repeated cell-wise splits.

Observed Cycles	Retrieval MAE	Median MAE	RMSE	Spearman	Modal Selected Configuration
10	164.26 ± 54.07	147.32	283.88	0.68	stats, $k = 15$, median (11/20)
20	168.54 ± 62.76	166.33	297.17	0.63	shape, $k = 3$, mean, log (10/20)
50	182.76 ± 55.87	164.10	306.39	0.59	degradation, $k = 7$, trimmed, log (6/20)
100	181.02 ± 53.45	157.41	298.06	0.64	raw_stats, $k = 9$, median (5/20)

Table 3. Repeated-split HybridSupervisedRetrieval comparison against the validation-selected supervised baseline. Positive gain means lower MAE for the hybrid method.

Observed Cycles	Supervised MAE	Hybrid MAE	Mean Gain	Median Gain	95% CI for Gain	Wilcoxon p	Positive Splits
10	170.69 ± 51.91	160.61 ± 54.71	10.08	2.66	[1.19, 20.39]	0.058	13/20
20	183.82 ± 48.64	171.92 ± 47.77	11.90	7.28	[1.69, 23.06]	0.028	11/20
50	181.43 ± 56.87	176.35 ± 56.91	5.08	1.21	[−2.00, 12.81]	0.067	12/20
100	172.40 ± 41.74	172.75 ± 51.11	−0.35	0.00	[−9.91, 7.79]	0.340	9/20

Repeated-split retrieval-only diagnostics

**Figure 3.** Repeated-split retrieval-only diagnostics: (A) validation-selected retrieval MAE; (B) selected retrieval feature sets; (C) selected neighbor counts; and (D) the split-level MAE distribution over 20 cell-wise splits.

3.3. Hybrid Prediction and Fusion Behavior

Figure 4 and Table 3 summarize the repeated-split EOL results. HybridSupervisedRetrieval gives its clearest improvement in the shortest windows. Across 20 cell-wise splits, the mean MAE reduction is 10.08 cycles at 10 observed cycles and 11.90 cycles at 20 observed cycles. The corresponding bootstrap 95% confidence intervals for the split-level gain are 1.19–20.39 cycles and 1.69–23.06 cycles. The one-sided Wilcoxon signed-rank tests give $p = 0.058$ and $p = 0.028$, respectively. At 50 cycles, the mean gain is smaller and the confidence interval crosses zero. At 100 cycles, the hybrid method does not improve the validation-selected supervised baseline on average.

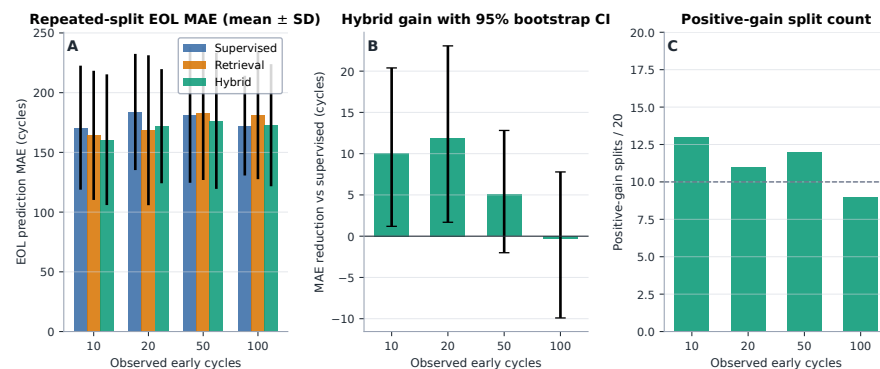


Figure 4. Repeated-split EOL prediction results on held-out Severson/TRI/MIT test cells: (A) mean EOL MAE over 20 cell-wise splits for supervised, retrieval-only, and hybrid predictors; (B) MAE reduction of the hybrid method relative to the validation-selected supervised baseline with bootstrap 95% confidence intervals; and (C) the number of splits in which the hybrid method reduces test MAE.

The supervised model inside the selected hybrid configuration is chosen as part of the fused predictor, rather than copied from the independently selected supervised baseline. Joint validation reflects the objective of the hybrid experiment: the supervised learner, retrieval configuration, and fusion weight are selected for their combined EOL prediction error. The median gains are smaller than the mean gains, especially at 10 cycles, where the median gain is 2.66 cycles while the mean gain is 10.08 cycles. This pattern indicates that the early-window benefit is concentrated in some splits rather than distributed uniformly across all partitions. The repeated-split evaluation shows that small apparent gains at longer windows should be interpreted cautiously.

Across the same splits, hybrid RMSE is 274.53, 291.77, 292.40, and 279.95 cycles at 10, 20, 50, and 100 observed cycles, respectively, compared with 279.62, 296.65, 293.61, and 274.11 cycles for the supervised baseline. Mean split-level Spearman correlations are 0.70, 0.67, 0.65, and 0.73 for the hybrid predictor and 0.68, 0.68, 0.68, and 0.72 for the supervised baseline. Thus, the MAE gains at 10 and 20 cycles do not correspond to a uniform improvement in every metric or every split.

The validation-selected fusion weights in Figure 5 show how the hybrid model allocates weight between its two prediction sources across repeated splits. The mean supervised weight α is 0.43, 0.52, 0.59, and 0.60 at 10, 20, 50, and 100 observed cycles. The selected weights therefore do not indicate a purely retrieval-driven predictor; rather, retrieval contributes as an additional historical-cell prior inside a validation-selected mixture. Together with the retrieval-only results in Table 2, the error gains indicate that the early-window benefit is driven mainly by access to historical-cell analogs, with fusion providing a validation-selected way to combine that prior with supervised features.

3.4. Compatibility-Filtered External Evidence

The BatteryLife external screen clarifies the scope of the retrieval-prior claim. Table 4 reports the most informative compatible windows, with targeted 40-split confirmation for SNL 50 cycles and XJTU 20/50 cycles. SNL illustrates why mean improvements alone are insufficient: retrieval-only lowers mean MAE from 333.30 to 278.85 cycles at 50 observed cycles, but the paired 95% confidence interval for the gain spans zero (−12.88 to 133.84 cycles; $p = 0.264$). It is therefore treated as heterogeneous evidence rather than confirmation. BatteryLife-XJTU provides the clearer fusion-positive case. At 50 cycles, fusion reduces mean MAE from 34.11 to 28.02 cycles; its paired gain is 6.09 cycles (95% confidence interval 0.47–11.41 cycles; $p = 0.007$). At 20 cycles, the corresponding fusion result remains borderline ($p = 0.052$). BatteryLife-HUST provides weaker, exploratory support: retrieval-only improves the mean MAE at 5 and 20 observed cycles, but the gains are not stable across all windows.

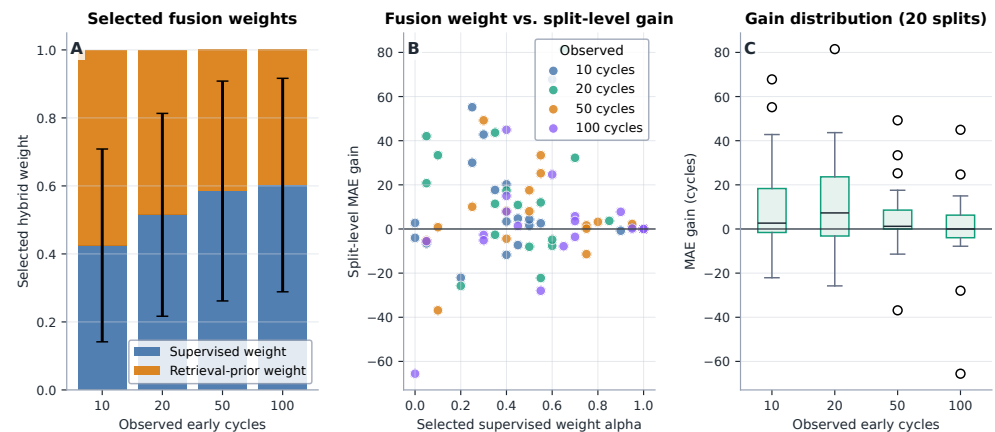


Figure 5. Repeated-split fusion behavior: (A) mean validation-selected supervised and retrieval weights; (B) the relation between selected weight and split-level MAE gain; and (C) the gain distribution over 20 cell-wise splits.

Table 4. Compatibility-filtered external dataset screen using BatteryLife processed datasets. Values are mean MAE over repeated cell-wise splits; SNL 50-cycle and XJTU 20/50-cycle rows use 40 non-overlapping splits, while HUST rows use the initial 20 splits. Lower is better.

Dataset	Cells	Observed Cycles	Supervised	Retrieval-Only	Hybrid	Selector	Main Observation
BatteryLife-SNL	52	50	333.30 ± 241.75	278.85 ± 162.28	319.11 ± 218.07	309.98 ± 217.93	heterogeneous; not confirmed
BatteryLife-XJTU	23	20	33.24 ± 16.36	32.05 ± 20.23	29.12 ± 18.15	29.41 ± 17.06	fusion borderline
BatteryLife-XJTU	23	50	34.11 ± 17.28	28.77 ± 15.03	28.02 ± 14.70	29.42 ± 17.15	fusion confirmed
BatteryLife-HUST	77	5	361.51 ± 67.18	343.49 ± 83.31	352.32 ± 68.94	347.51 ± 58.62	retrieval prior helpful
BatteryLife-HUST	77	20	339.33 ± 69.18	329.40 ± 67.91	337.47 ± 69.72	332.76 ± 65.09	retrieval prior helpful

These external results support a narrower but more defensible claim than a universal model-ranking statement. The SNL confirmation shows that a large mean gain can remain unstable under repeated splits. Fusion is confirmed only for the XJTU 50-cycle setting, while the XJTU 20-cycle result is borderline and other external signals are exploratory. The selector results further show that validation does not always identify the lowest-error test family, so selector performance is treated as an engineering diagnostic rather than as the central contribution.

3.5. Trajectory Diagnostics

The trajectory case studies in Figure 6 show short-, medium-, and long-life examples. Retrieval supplies full-horizon degradation trajectories with empirical uncertainty bands, while Chronos-Bolt supplies a short-horizon capacity forecast. The two outputs diagnose different aspects of the prediction: the pretrained forecaster represents local capacity evolution, and retrieval exposes which historical cells are being used as long-range analogs. When a retrieval prediction is poor, the support cells can be examined directly for early-shape similarity and eventual lifetime mismatch.

3.6. Domain-Shift Sanity Check

The GPDM/NASA/Oxford sanity check gives a stricter view of domain shift. In Figure 7, Chronos-Bolt trajectory RMSE is much larger than on Severson, and PICP80 is unstable. Retrieval is competitive in trajectory RMSE for some external settings, but the coverage is still below the desired level. This check limits the scope of the main findings: protocol, chemistry, temperature, sampling, and label-definition shifts remain major barriers to broad deployment, even when retrieval priors are useful on compatible datasets.

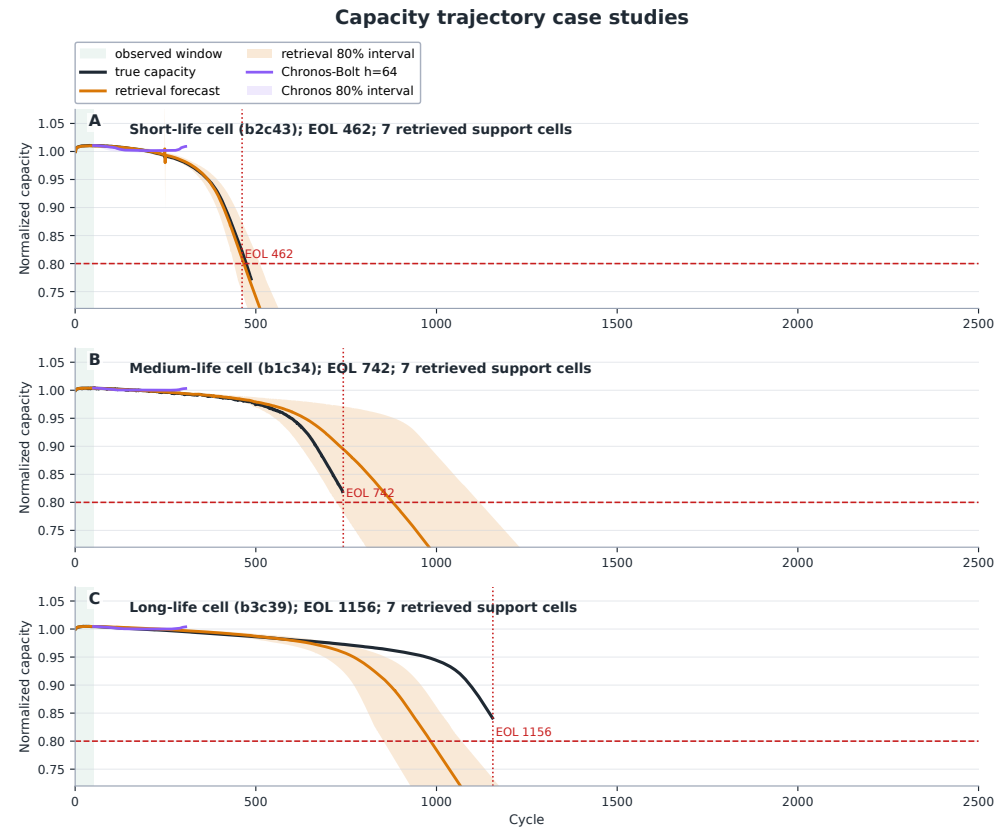


Figure 6. Capacity trajectory case studies: (A) a short-life cell; (B) a medium-life cell; and (C) a long-life cell. Retrieval forecasts provide full-horizon degradation priors and empirical uncertainty bands; Chronos-Bolt provides short-horizon trajectory forecasts and its purple shaded 80% prediction interval.

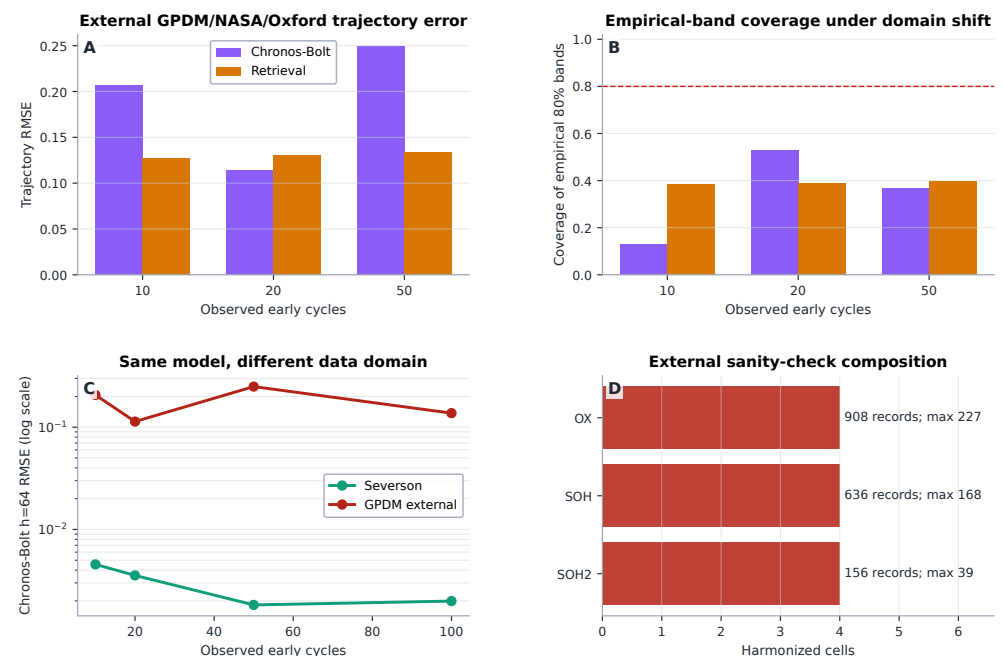


Figure 7. External GPDM/NASA/Oxford sanity check: (A) external trajectory RMSE; (B) empirical 80% prediction-interval coverage; (C) method comparison in the external domain; and (D) external dataset composition. Domain shift makes trajectory forecasting substantially harder than in the primary Severson setting; this experiment is used as a robustness warning rather than as a definitive cross-dataset benchmark.

4. Discussion

4.1. Evidence from Compatible Windows

The repeated-split experiments identify the regimes where retrieval is most useful. In the first 10 or 20 cycles of the Severson/TRI/MIT dataset, normalized-capacity trajectories differ only slightly across many cells, and a supervised model must infer lifetime from small local deviations. Retrieval reframes the inference around aged historical cells: which support cells looked similar at the same early point, and what EOL values did they later reach? The retrieval-only baseline is already competitive in this weak-signal regime, with mean MAE values of 164.26 and 168.54 cycles at 10 and 20 observed cycles. The hybrid method improves over the validation-selected supervised baseline by 10.08 and 11.90 cycles at the same windows, but the 20-cycle retrieval-only result also shows that the value of the proposed framework lies in exposing and using the historical-cell prior, not in proving that fusion always outperforms retrieval alone.

The targeted external confirmations sharpen this interpretation. The apparently large SNL 50-cycle retrieval gain is heterogeneous and does not retain paired statistical support after 40 splits. BatteryLife-XJTU provides the clearer fusion-positive case: at 50 observed cycles, the fused predictor retains a positive paired gain after 40 splits. These results suggest that access to compatible historical-cell evidence can help, but they do not support a universal claim that retrieval-only or fusion will improve every external dataset.

The gain becomes less reliable outside such compatible regimes. In the primary Severson/TRI/MIT experiment, the fused gain at 50 cycles has a confidence interval that includes zero, and at 100 cycles the repeated-split mean gain is slightly negative. Longer prefixes provide more cell-specific information, which can reduce the incremental benefit of retrieved historical analogs. Across external datasets, HUST shows modest early retrieval gains, whereas CALB does not provide a stable positive result after repeated-split confirmation. The evidence therefore supports compatibility-filtered use of retrieval priors rather than a universal model-ranking claim.

4.2. Forecasting Models and Engineering Use

The Chronos-Bolt results separate short-horizon trajectory forecasting from threshold-crossing EOL prediction. The model forecasts 64-cycle capacity trajectories accurately on Severson without battery-specific training, but fixed-horizon forecasts often do not extend to the 80% SOH threshold. A future battery prognosis system could still use pretrained temporal models as trajectory modules, similarity encoders, or regularizers for degradation representations. The present results do not support using zero-shot fixed-horizon forecasting alone as a complete EOL predictor.

For engineering review, the main practical advantage of retrieval is traceability. Each prediction can be traced to specific support cells, allowing an analyst to inspect whether the retrieved cells have plausible early capacity shapes, protocols, and eventual lifetimes. The empirical trajectory bands should be interpreted as analogue dispersion rather than calibrated uncertainty: the retrieval PICP80 values are approximately 0.59–0.68 across the repeated-split windows, below the nominal 0.80 level. Traceability is relevant to fast-charging protocol screening, production batch assessment, early-life cell sorting, and second-life battery selection, where decisions often depend on a small prefix of the lifetime trajectory.

4.3. Limitations and Future Work

The evidence is limited by dataset scope and by the compatibility filter used for the external screen. The primary experiments use the Severson/TRI/MIT dataset, which is public and well suited to early lifetime prediction, but it represents a particular chemistry,

protocol family, and experimental setting. BatteryLife-SNL, BatteryLife-HUST, BatteryLife-XJTU, and BatteryLife-CALB expand the empirical scope, but their results are not uniform: SNL's mean retrieval gain is heterogeneous, XJTU confirms fusion only in a small-cell 50-cycle setting, HUST gives weaker early-window signals, and CALB is not a strong positive case. The GPDM/NASA/Oxford analysis contains only 12 harmonized cells and differs in degradation definitions and sampling; it is best read as a warning about domain shift, not as a definitive cross-dataset ranking.

The signal set is also limited. Capacity is the most portable measurement across the public datasets used here, but voltage, current, temperature, internal resistance, and incremental-capacity features may improve prediction when high-quality cycle-level curves are available. Those signals also require additional preprocessing choices that can affect comparability across datasets. Retrieval coverage is a further constraint: if the support database lacks cells similar to the target cell, the retrieved prior may be misleading. Future work should test protocol-aware retrieval, foundation-model embeddings for degradation similarity, larger cross-dataset benchmarks, and fine-tuned time-series foundation models designed for battery degradation trajectories.

5. Conclusions

We examined data-efficient early lithium-ion battery lifetime prediction using public cycling data. The proposed framework combines supervised early-cycle EOL predictors with retrieval-based degradation priors from similar historical cells, while also evaluating retrieval-only prediction as a direct expression of the same prior. On the Severson/TRI/MIT dataset, repeated cell-wise splits showed that supervised–retrieval fusion reduced mean MAE relative to the validation-selected supervised baseline at 10 and 20 observed cycles, while the gains at 50 and 100 cycles were not robust. External confirmation further showed that a large mean gain on SNL was heterogeneous, whereas XJTU at 50 observed cycles retained a modest fusion-positive gain after 40 non-overlapping repeated splits. Chronos-Bolt results showed that pretrained time-series models can forecast short-horizon capacity trajectories, but fixed-horizon zero-shot forecasting was insufficient for full-horizon EOL inference. Retrieval-augmented degradation priors therefore provide an interpretable complement to supervised battery lifetime models in compatible weak-signal regimes. Broader cross-chemistry and cross-protocol deployment remains a subject for future validation.

Author Contributions: Conceptualization, Y.Z. (Yuelin Zou) and Y.Z. (Yajun Zhang); methodology, Y.Z. (Yuelin Zou) and K.L.; software, Y.Z. (Yuelin Zou); validation, Y.Z. (Yuelin Zou) and Y.Z. (Yajun Zhang); formal analysis, Y.Z. (Yuelin Zou); investigation, Y.Z. (Yuelin Zou); data curation, Y.Z. (Yuelin Zou); writing—original draft preparation, Y.Z. (Yuelin Zou); writing—review and editing, Y.Z. (Yuelin Zou), Y.Z. (Yajun Zhang), and K.L.; visualization, Y.Z. (Yuelin Zou); supervision, Y.Z. (Yajun Zhang) and K.L.; project administration, Y.Z. (Yuelin Zou). All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Key Research and Development Program of Xinjiang Uygur Autonomous Region (2024B03028 and 2022A03016) and the Natural Science Foundation of Xinjiang Uygur Autonomous Region (2022D01C54).

Institutional Review Board Statement: Not applicable. This study used publicly available battery cycling datasets and did not involve humans or animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: The primary Severson/TRI/MIT dataset is publicly available from the original data repository and associated public loaders. The BatteryLife processed datasets used for compatibility-filtered external screening are publicly available from the BatteryLife repository and associated Zenodo/Hugging Face releases. The NASA and Oxford datasets used for domain-shift

sanity checking are publicly available from their respective repositories. For review, the resubmission package includes the repeated-split definitions (seeds 0–19; 86/19/19 fit/validation/test cells), candidate retrieval configurations, metric tables, and the scripts used to generate them. These materials will be released in a public repository upon acceptance, subject to the licenses and redistribution conditions of the original third-party datasets.

Acknowledgments: The authors acknowledge the maintainers of the public battery datasets and open-source software libraries used in this study. OpenAI Codex/ChatGPT (web-based services accessed in 2026) (OpenAI, San Francisco, CA, USA) was used during manuscript preparation to assist with language revision, literature-positioning drafts, consistency checks among the text, tables, and figures, and scripting support for reference metadata validation and figure-format refinement. The tools were not used to generate raw experimental data, alter the reported numerical results, or make final scientific interpretations without author review. The authors reviewed and edited all AI-assisted text and take full responsibility for the content of the manuscript.

Conflicts of Interest: Author Yuelin Zou was employed by the company State Grid Xinjiang Electric Power Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Attia, P.M.; Grover, A.; Jin, N.; Severson, K.A.; Markov, T.M.; Liao, Y.H.; Chen, M.H.; Cheong, B.; Perkins, N.; Yang, Z.; et al. Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature* **2020**, *578*, 397–402. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Severson, K.A.; Attia, P.M.; Jin, N.; Perkins, N.; Jiang, B.; Yang, Z.; Chen, M.H.; Aykol, M.; Herring, P.K.; Fraggedakis, D.; et al. Data-driven prediction of battery cycle life before capacity degradation. *Nat. Energy* **2019**, *4*, 383–391. [\[CrossRef\]](#)
3. Richardson, R.R.; Osborne, M.A.; Howey, D.A. Gaussian process regression for forecasting battery state of health. *J. Power Sources* **2017**, *357*, 209–219. [\[CrossRef\]](#)
4. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
5. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794. [\[CrossRef\]](#)
6. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*; Curran Associates Inc.: Red Hook, NY, USA, 2017; Volume 30.
7. Chemali, E.; Kollmeyer, P.J.; Preindl, M.; Emadi, A. State-of-charge estimation of Li-ion batteries using deep neural networks: A machine learning approach. *J. Power Sources* **2018**, *400*, 242–255. [\[CrossRef\]](#)
8. Roman, D.; Saxena, S.; Robu, V.; Pecht, M.; Flynn, D. Machine learning pipeline for battery state-of-health estimation. *Nat. Mach. Intell.* **2021**, *3*, 447–456. [\[CrossRef\]](#)
9. Zhang, Y.; Xiong, R.; He, H.; Pecht, M.G. Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. *IEEE Trans. Veh. Technol.* **2018**, *67*, 5695–5705. [\[CrossRef\]](#)
10. Liu, D.; Pang, J.; Zhou, J.; Peng, Y.; Pecht, M. Particle learning framework for estimating the remaining useful life of lithium-ion batteries. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 280–293. [\[CrossRef\]](#)
11. Su, X.; Wang, S.; Pecht, M.; Zhao, L.; Ye, Z. Interacting multiple model particle filter for prognostics of lithium-ion batteries. *Microelectron. Reliab.* **2017**, *70*, 59–69. [\[CrossRef\]](#)
12. Zhang, Y.Z.; Xiong, R.; He, H.W.; Pecht, M. Validation and verification of a hybrid method for remaining useful life prediction of lithium-ion batteries. *J. Clean. Prod.* **2019**, *212*, 240–249. [\[CrossRef\]](#)
13. Wang, F.; Zhai, Z.; Zhao, Z.; Di, Y.; Chen, X. Physics-informed neural network for lithium-ion battery degradation stable modeling and prognosis. *Nat. Commun.* **2024**, *15*, 4332. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Wang, S.; Wang, Y.; Xiao, J.; Soo, Y.Y.; Chen, Z. A Physics-Informed Neural Network with Attention-Enhanced Multisource Feature Fusion for Lithium-Ion Battery SOH Estimation. *IEEE Trans. Ind. Electron.* **2026**, 1–13. [\[CrossRef\]](#)
15. Lin, C.; Tuo, X.; Wu, L.; Zhang, G.; Lyu, Z.; Zeng, X. Physics-Informed Machine Learning for Accurate SOH Estimation of Lithium-Ion Batteries Considering Various Temperatures and Operating Conditions. *Energy* **2025**, *318*, 134937. [\[CrossRef\]](#)
16. Wu, C.; Xu, C.; Wang, L.; Fu, J.; Meng, J. Lithium-ion battery remaining useful life prediction based on data-driven and particle filter fusion model. *Green Energy Intell. Transp.* **2025**, *4*, 100267. [\[CrossRef\]](#)

17. Wang, F.K.; Amogne, Z.E.; Chou, J.H.; Tseng, C. Online remaining useful life prediction of lithium-ion batteries using bidirectional long short-term memory with attention mechanism. *Energy* **2022**, *254*, 124344. [[CrossRef](#)]
18. Wu, J.; Cheng, X.; Huang, H.; Fang, C.; Zhang, L.; Zhao, X.; Zhang, L.; Xing, J. Remaining useful life prediction of lithium-ion batteries based on PSO-RF algorithm. *Front. Energy Res.* **2023**, *10*, 937035. [[CrossRef](#)]
19. Zhu, C.; He, Z.; Bao, Z.; Sun, C.; Gao, M. Prognosis of lithium-ion batteries remaining useful life based on a sequence-to-sequence model with variational mode decomposition. *Energies* **2023**, *16*, 803. [[CrossRef](#)]
20. Li, Z.; Bai, F.; Zuo, H.; Zhang, Y. Remaining useful life prediction for lithium-ion batteries based on iterative transfer learning and Mogrifier LSTM. *Batteries* **2023**, *9*, 448. [[CrossRef](#)]
21. Cai, Y.; Li, W.; Zahid, T.; Zheng, C.; Zhang, Q.; Xu, K. Early prediction of remaining useful life for lithium-ion batteries based on CEEMDAN-transformer-DNN hybrid model. *Heliyon* **2023**, *9*, e17754. [[CrossRef](#)] [[PubMed](#)]
22. Huang, Z.; Ma, Z. Remaining useful life prediction of lithium-ion batteries based on autoregression with exogenous variables model. *Reliab. Eng. Syst. Saf.* **2024**, *252*, 110485. [[CrossRef](#)]
23. Nie, Y.; Nguyen, N.H.; Sinthong, P.; Kalagnanam, J. A time series is worth 64 words: Long-term forecasting with transformers. In Proceedings of the International Conference on Learning Representations, Kigali, Rwanda, 1–5 May 2023.
24. Tang, P.; Qiu, Z.; Yao, Z.; Pan, J.; Cheng, D.; Gu, X.; Sun, C. Lithium-ion battery RUL prediction based on optimized VMD-SSA-PatchTST algorithm. *Sci. Rep.* **2025**, *15*, 26824. [[CrossRef](#)] [[PubMed](#)]
25. Liu, L.; Huang, J.; Zhao, H.; Li, T.; Li, B. PatchFormer: A novel patch-based transformer for accurate remaining useful life prediction of lithium-ion batteries. *J. Power Sources* **2025**, *631*, 236187. [[CrossRef](#)]
26. Zhao, J.; Wang, Z.; Wu, Y.; Burke, A.F. Predictive pretrained transformer (PPT) for real-time battery health diagnostics. *Appl. Energy* **2025**, *377*, 124746. [[CrossRef](#)]
27. Zhao, F.M.; Gao, D.X.; Cheng, Y.M.; Yang, Q. Application of state of health estimation and remaining useful life prediction for lithium-ion batteries based on AT-CNN-BiLSTM. *Sci. Rep.* **2024**, *14*, 29026. [[CrossRef](#)] [[PubMed](#)]
28. Chen, S.; Fang, X. Remaining useful life prediction of lithium-ion batteries via spatial attention TLSTM and dilated CNN with evolutionary optimization. *Sci. Rep.* **2025**, *15*, 32340. [[CrossRef](#)] [[PubMed](#)]
29. Zhang, H.; Li, Y.; Zheng, S.; Lu, Z.; Gui, X.; Xu, W.; Bian, J. Battery lifetime prediction across diverse ageing conditions with inter-cell deep learning. *Nat. Mach. Intell.* **2025**, *7*, 270–277. [[CrossRef](#)]
30. Tan, R.; Hong, W.; Tang, J.; Lu, X.; Ma, R.; Zheng, X.; Li, J.; Huang, J.; Zhang, T.Y. BatteryLife: A Comprehensive Dataset and Benchmark for Battery Life Prediction. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2, Toronto, ON, Canada, 3–7 August 2025*; Association for Computing Machinery: New York, NY, USA, 2025; pp. 5789–5800. [[CrossRef](#)]
31. Fan, Z.; Guo, Y.; Song, L.; Du, J.; Wang, H.; Li, X. BatteryTSFM: Generalizable long-horizon degradation prediction across conditions and chemistries via time series foundation models. *Energy AI* **2025**, *22*, 100646. [[CrossRef](#)]
32. Wong, K.L.; Im, S.K.; Fang, X.; Tang, S.K.; Lam, C.T.; Pau, G. Prognostics by generalists: Large language models for lithium-ion batteries health forecasting. *Sci. Rep.* **2026**, *16*, 16723. [[CrossRef](#)] [[PubMed](#)]
33. de la Iglesia, D.H.; Corbacho, C.C.; Dib, J.Z.; Alonso-Secades, V.; Lopez Rivero, A.J. Advanced machine learning and deep learning approaches for estimating the remaining life of EV batteries: A review. *Batteries* **2025**, *11*, 17. [[CrossRef](#)]
34. Ansari, A.F.; Stella, L.; Turkmen, C.; Zhang, X.; Mercado, P.; Shen, H.; Shchur, O.; Rangapuram, S.S.; Pineda Arango, S.; Kapoor, S.; et al. Chronos: Learning the language of time series. *arXiv* **2024**, arXiv:2403.07815. [[CrossRef](#)]
35. Das, A.; Kong, W.; Leach, A.; Mathur, S.; Sen, R.; Yu, R. A decoder-only foundation model for time-series forecasting. *arXiv* **2023**. [[CrossRef](#)]
36. NASA Ames Prognostics Center of Excellence. *Li-ion Battery Aging Datasets*; NASA: Washington, DC, USA, 2007.
37. Birkel, C.R. *Oxford Battery Degradation Dataset 1*; University of Oxford: Oxford, UK, 2017.

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