Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Remaining Useful Life Estimation of Lithium-Ion Batteries: A Controlled Benchmark of Physics-Informed Features and Mamba-Based Architectures on CALCE CS2

Dikshant Dikshant

dikshant1103@gmail.com

Motilal Nehru National Institute of Technology https://orcid.org/0009-0006-9434-610X

Dr Praveen Kumar Agarwal

Motilal Nehru National Institute of Technology

Research Article

Keywords: Battery RUL, Feature Engineering, Mamba SSM, Physics-Informed Learning, CALCE Dataset, Degradation Anchor Cross-Attention

Posted Date: August 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10638876/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Remaining Useful Life Estimation of Lithium-Ion Batteries: A Controlled Benchmark of Physics-Informed Features and Mamba-Based Architectures on CALCE CS2

Dikshant, Dr Praveen Kumar Agarwal

M.Tech (CAD/CAM), Department of Mechanical Engineering, MNNIT Allahabad

Supervised by Dr Praveen Kumar Agarwal, Associate Professor, Dept. of Mechanical Engineering, MNNIT Allahabad

Abstract—Remaining useful life (RUL) prediction underpins maintenance scheduling, warranty assessment and safe operating limits in battery management systems (BMS) for electric vehicles and stationary storage. A 13-feature physics-informed description of each charge–discharge cycle is defined, built around the capacity normalisation ratio cap_pct = C(t)/C(0), which expresses cell health independently of absolute initial capacity. Two architectures based on the Mamba selective state-space model are evaluated: MambaRUL, in which a Mamba encoder is combined with Degradation Anchor Cross-Attention (DACA) using three learnable query vectors associated with the three degradation phases, and PhysicsMambaRUL, which adds three parallel heads—direct RUL regression, a fade-rate estimate and a double-exponential capacity model—combined through learned softmax weights. All models are trained and evaluated on CALCE CS2 cells under a single protocol, with CS2_37 and CS2_38 permanently held out. Among the models compared, a GRU trained on the same 13 features gives the lowest error, at RMSE = 4.39 and 7.48 cycles (R^2 = +0.996 and +0.993) on the two test cells, ahead of PhysicsMambaRUL (13.60 and 20.49 cycles) and MambaRUL (28.54 and 38.57 cycles). Removing cap_pct from the feature set raises MambaRUL RMSE by 10.83 and 11.67 cycles on the same cells. In an exploratory test-cell analysis, the physics-consistency term ($\lambda = 0.3$) lowered PhysicsMambaRUL RMSE on CS2_37 from 19.47 to 14.72 cycles. RUL targets are normalised by the complete observed lifetime of each cell, a quantity that is an oracle at inference; the cycle-domain errors reported here therefore describe performance under a controlled CALCE CS2 benchmark rather than deployment-ready online RUL accuracy.

Keywords—Battery RUL, Feature Engineering, Mamba SSM, Physics-Informed Learning, CALCE Dataset, Degradation Anchor Cross-Attention

I. INTRODUCTION

Lithium-ion cells are the dominant energy storage technology in electric vehicles (EVs), grid-scale storage and portable electronics. The pack accounts for a large share of vehicle cost, and the capacity lost over service life limits both usable range and residual value. Remaining useful life (RUL) prediction—the number of cycles left before a cell reaches its end-of-life (EOL) capacity threshold—is therefore a core function of a battery management system (BMS), where it informs maintenance scheduling, warranty assessment and derating decisions. Growing EV deployment in India under the FAME scheme has increased the demand for health monitoring that stays reliable across cells and duty cycles.

Capacity fade does not follow a single mechanism, and three regimes are commonly distinguished. Early cycling is dominated by growth of the solid–electrolyte interphase (SEI) on the anode, which consumes cyclable lithium and produces slow, approximately linear fade. Loss of active material (LAM) then becomes the leading contribution and the fade rate increases. In the final regime, lithium plating and dendrite formation drive accelerated loss; the transition into it is usually called the knee point. Each regime leaves a different signature in the discharge voltage profile and in the shape of the capacity trajectory, so a model that treats the whole sequence uniformly has to infer the operative regime implicitly. This observation motivates the phase-aware attention component described in Section II-B.

Recurrent networks remain a common choice for this task, but gradient flow degrades over the long cycle sequences typical of full-lifetime records, which limits the history a single model can use in practice. Transformers [11] replace recurrence with self-attention and avoid that difficulty, at a cost that grows quadratically with sequence length. Selective state-space models offer a third option: Mamba [10] keeps a recurrent hidden state whose transition parameters depend on the input, which gives linear-time inference together with content-dependent gating. Battery-specific variants have appeared recently, among them RUL-Mamba [5] and SambaMixer [6], which apply selective state-space models to prognostics using conventional cycle-level inputs. Their reported errors are not restated here, since they are obtained on different datasets and under different evaluation protocols. What remains less clear is how much of the attainable accuracy comes from the sequence model and how much from the representation supplied to it. The role of explicit capacity normalisation is investigated through the ablation study in Section IV-C.

Three aspects are evaluated in this study. The first is a 13-feature physics-informed representation of each cycle, organised around the normalised capacity ratio cap_pct = C(t)/C(0). The second is a pair of architectures: MambaRUL, which couples a Mamba encoder with Degradation Anchor Cross-Attention, and PhysicsMambaRUL, which adds fade-rate and curve-fitting heads under a physics-consistency penalty. The third is a controlled comparison of these architectures against LSTM, GRU and Transformer baselines trained on the same features, the same split and the same seed. Under this protocol the GRU gives the lowest test-cell RMSE, and the ablations in Section IV isolate the effect of individual design choices.

II. METHODOLOGY

A. Feature Engineering

Thirteen features are computed from each charge–discharge cycle to describe the electrochemical state of the cell. The resulting cycle-level vectors are assembled into sliding windows of width W = 30 cycles with stride 1, giving input tensors of shape (B, 30, 13). The window length was chosen so that a window spans a segment over which fade is approximately smooth while remaining short relative to the cell lifetimes in this dataset (55–337 cycles): a shorter window carries too little trend information to support an estimate of the local fade rate, and a longer one mixes distinct degradation regimes within a single input.

The central quantity in the feature set is cap_pct(t) = C(t)/C(0), the discharge capacity at cycle t divided by the initial capacity of the same cell. Two properties follow. The indicator is bounded in [0, 1] and decreases monotonically apart from measurement noise and capacity regeneration, and it removes cell-to-cell differences in absolute initial capacity, so that trajectories from cells of unequal capacity become directly comparable. The same reasoning is used in fatigue analysis, where damage is expressed as a fraction of the failure limit rather than as an absolute count of load cycles.

TABLE I. PHYSICS-INFORMED FEATURE SET (13 FEATURES; CAP_PCT ROW HIGHLIGHTED)

No.	Feature	Formula	Physical Meaning	Phase
1	**cap_pct**	C(t)/C(0)	Normalised health indicator [0–1]	All
2	discharge_cap	C(t) [Ah]	Absolute discharge capacity/cycle	All
3	voltage_mean	mean(V_d)	Mean discharge voltage drops with SEI	1–2
4	voltage_std	std(V_d)	Voltage variability increases at the knee	2–3
5	voltage_min	min(V_d)	Min voltage; sensitive to Li plating	3
6	voltage_max	max(V_d)	Max voltage; CC-CV reference	1
7	current_mean	mean(|I|)	Mean discharge current magnitude	All
8	temperature_mean	mean(T)	Cell temperature; proxy for R_int rise	2–3
9	energy_wh	Σ(VIΔt)	Total energy delivered per cycle [Wh]	All
10	cum_energy	cumsum(E)	Cumulative energy — SoH proxy	All
11	ic_area	∫|dQ/dV|dV	IC area — SEI film thickness indicator	1–2
12	soc_at_knee	SOC@inflect.	Voltage plateau shift; electrode stoich.	2–3
13	rolling_slope	slope(C, W=10)	Local capacity fade rate (10-cycle win.)	2–3

Among the feature configurations evaluated in this study, inclusion of cap_pct produced the largest observed reduction in RMSE (Section IV-C).

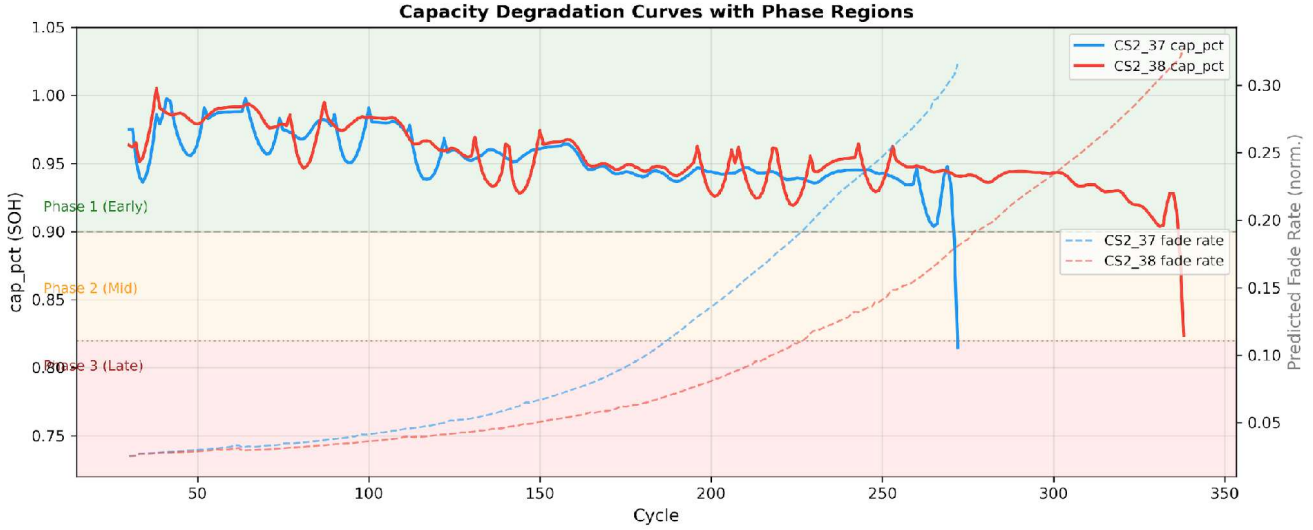


Fig. 1. Capacity fade of the two CALCE CS2 test cells, plotted as cap_pct against cycle number with the estimated local fade rate on the right-hand axis. The shaded bands mark the three degradation regions (SEI growth → LAM → lithium plating) that motivate the DACA anchor design.

B. MambaRUL Architecture

The MambaRUL architecture is built on the Mamba selective state-space model (SSM) [10]. The selective SSM recurrence is defined as:

$$h_t = \bar{A}_t \odot h_{t-1} + \bar{B}_t \odot x_t, \quad y_t = C_t \cdot h_t$$

where $\bar{A}_t$, $\bar{B}_t$ and C_t are input-dependent (selective) parameters obtained from lightweight linear projections of x_t. Gating is therefore content-dependent: information relevant to the current degradation state is written into the hidden state while less informative variation is attenuated, and the recurrence stays O(N) in sequence length rather than the $O(N^2)$ cost of full self-attention.

Degradation Anchor Cross-Attention (DACA) is applied to the output sequence of the Mamba encoder. Three learnable query vectors Q_early, Q_mid and Q_late $\in \mathbb{R}^d$ act as anchors, one for each of the three degradation phases described in Section I. For each window, the DACA module computes:

$$\text{Attn}(Q_{\text{phase}}, K, V) = \text{softmax}(Q_{\text{phase}} K^T / \sqrt{d})\, V, \quad \text{phase} \in \{\text{early}, \text{mid}, \text{late}\}$$

The three anchor outputs are concatenated and projected to a fixed-dimensional representation. Q_early is intended to attend to early-window behaviour associated with SEI growth, Q_mid to the approximately linear fade regime, and Q_late to the acceleration near the knee point. No phase labels are supplied during training, so this correspondence is a design intention rather than an enforced constraint; the anchors are free to specialise as the loss dictates, and all three are evaluated at every window, which allows information from different degradation timescales to be combined without an explicit phase classifier.

TABLE II. ARCHITECTURE COMPARISON

Model	Processing	Complexity	Params	Attention
LSTM (2-layer)	Recurrent	O(N)	205,441	None
GRU (2-layer)	Recurrent	O(N)	154,113	None
Transformer	Parallel	$O(N^2)$	270,721	Full self-attn.
MambaRUL (prop.)	**Sel. SSM**	**O(N)**	**153,025**	**DACA (3 anchors)**
PhysMambaRUL (p.)	**Sel. SSM**	**O(N)**	**155,217**	**DACA+phys. heads**

C. PhysicsMambaRUL

PhysicsMambaRUL extends MambaRUL with three parallel prediction heads attached to the shared Mamba encoder output h_shared, each expressing a different aspect of the degradation process.

Head 1 — Direct RUL: $\hat{y}_1 = \text{sigmoid}(\text{MLP}(h_\text{shared})) \in [0, 1]$. The sigmoid bounds the prediction to the normalised RUL range and prevents unphysical negative values.

Head 2 — Fade Rate: The capacity fade rate is estimated for each window and converted into a physics-based RUL estimate, $RUL_2 = (\text{cap_pct} - 0.80) / \text{fade_rate}$. A value of 0.80 normalised capacity is adopted in this study as the benchmark EOL criterion. This head expresses the relation between the current rate of capacity decline and the number of cycles remaining.

Head 3 — Double-Exponential Curve: Parameters [a, b, d] are predicted subject to $a \in [0.5, 1.0]$, $b \in [-0.01, 0]$, $c = 1-a$ and $d \in [-0.05, 0]$, giving $\text{cap}(t) = a \cdot \exp(b \cdot t) + c \cdot \exp(d \cdot t)$. RUL_3 is evaluated by differentiable soft counting of the cycles until cap(t) crosses 0.80, which keeps the head trainable end to end.

Outputs are combined via learned softmax weights:

$$\hat{y}_{\text{final}} = w \cdot [\hat{y}_1, RUL_{2,\text{norm}}, RUL_{3,\text{norm}}], \quad w = \text{softmax}(\text{Linear}_{3\to3}([\hat{y}_1, RUL_{2,\text{norm}}, RUL_{3,\text{norm}}]))$$

The physics consistency loss is defined as:

$$L = L_{\text{sse}}(\hat{y}_{\text{final}}, y) + \lambda \times (L_{\text{sse}}(RUL_{2,\text{norm}}, y) + L_{\text{sse}}(RUL_{3,\text{norm}}, y))$$

where $\lambda = 0.3$ is a fixed design choice; a sensitivity analysis for this coefficient is reported in Section IV-E. The additional terms penalise disagreement between the direct head and the two physics-constrained heads, which constrains the shared encoder towards representations from which both estimates can be recovered.

D. Training Protocol

RUL targets are normalised per cell as y_norm = RUL / L_cell, where L_cell is the total observed lifetime of that cell, and predictions are returned to cycles as pred_cycles = $\hat{y}$_norm × L_cell. Every cell is thereby mapped onto [0, 1] irrespective of its absolute lifetime. For the held-out test cells, L_cell is the complete observed lifetime and is therefore an oracle quantity that would not be available to an operating system before end of life; the cycle-domain RMSE values reported in Section IV should accordingly be read as controlled benchmark results under full-lifetime normalisation rather than as deployment-ready online RUL accuracy. Feature scaling statistics are estimated on the training cells alone. Training uses the Adam optimiser with weight decay 10^{-4} and gradient clipping at max_norm = 1.0, the latter being needed to keep the Mamba state updates stable. Hyperparameters differ by model family: the LSTM, GRU, Transformer and MambaRUL baselines use learning rate 10^{-4}, at most 150 epochs and early-stopping patience of 25 epochs, while PhysicsMambaRUL uses learning rate 8×10^{-5}, at most 200 epochs and patience of 30 epochs. A cosine annealing schedule with a 15-epoch linear warmup and a minimum learning rate of 10^{-6} is applied in both cases. Early stopping and checkpoint selection are driven by validation RMSE on CS2_34, and the checkpoint with the lowest validation RMSE is retained. At inference, the predicted sequence is made non-increasing with np.minimum.accumulate, consistent with monotonic capacity fade. Each configuration is trained once with seed = 42. Repeated evaluation of that same checkpoint over three passes reproduced identical results, which reflects deterministic inference; seed-to-seed training variance was not assessed in the present study.

III. EXPERIMENTAL SETUP

A. Dataset

The CALCE CS2 dataset [12] is used. It comprises LCO-chemistry lithium-ion cells with a nominal capacity of 1.1 Ah, cycled at 1C constant-current discharge with CC-CV charging at 25°C in a temperature-controlled environment. The dataset was selected because it is widely used in battery RUL benchmarking and provides cycle-level capacity, voltage and temperature records of the kind on which data-driven cycle-life prediction depends [13]. The split applied here is summarised in Table III.

TABLE III. CALCE CS2 DATASET SPLIT

Split	Cell IDs	Lifetime	Windows	Purpose
Train	CS2_35, CX2_16, CX2_33, CX2_36, CX2_38, CX2_37	309, 186, 55, 246, 280, 301	1,204	Model fitting
Val.	CS2_34	131	103	Checkpoint sel.
Test	**CS2_37, CS2_38**	**271, 337**	**243, 309**	**Final eval. ONLY**

The test cells CS2_37 and CS2_38 were held out from model fitting, feature-scaler estimation, checkpoint selection and early stopping. CX2_37 and CX2_38 in the training split belong to the CX2 series and are distinct cells from the CS2_37 and CS2_38 test cells; the similarity of the identifiers is a naming coincidence in the CALCE archive. Feature normalisation statistics (mean and standard deviation) were computed from the training cells only and applied unchanged to the validation and test cells. Sliding windows were generated within each cell after the cell-level split, so no window spans two cells and no window crosses a split boundary. The window counts in Table III are n_cycles − W + 1 per cell with W = 30 and stride 1, evaluated after removal of cycles whose recorded capacity exceeds that of the preceding cycle by more than 5%; this filter removes five cycles in CX2_16 and none in any other cell, which is why the training total is 1,204 rather than 1,209 windows. Cell-ID disjointness, scaler provenance and the minimum L2 distance between training and test windows were checked to confirm that the split was applied as described. Final evaluation on CS2_37 and CS2_38 was performed once per configuration. The lifetime of CS2_38 (337 cycles) exceeds that of the longest training cell (309 cycles) by 9.1%, so the test set includes a cell whose lifetime lies outside the range represented in training.

B. Baselines

Three baseline architectures are evaluated: an LSTM (2 layers, 128 hidden units), a GRU (2 layers, 128 hidden units) and a Transformer (2 layers, 4-head self-attention, 128-dimensional). All baselines are trained on the same 13-feature set, the same split and the same seed = 42, so that differences between them reflect architecture rather than data handling.

C. Evaluation Metrics

Five metrics are reported: RMSE in cycles for absolute error; R^2 for the proportion of variance in RUL explained by the prediction; RMSE% = (RMSE / cell_lifetime) × 100, which places the error on a per-cell scale; and W10 and W20, the percentages of prediction windows falling within 10 and 20 cycles of the true RUL. W20 = 100% is used as the primary benchmark criterion; at the 1C cycling rate of this dataset, 20 cycles corresponds to roughly 4–7 days of operation. This is a benchmark criterion and is not, on its own, evidence of deployment readiness.

IV. RESULTS AND DISCUSSION

A. Main Results

Results on the permanently held-out test cells are summarised in Table IV. Repeated evaluation of the same checkpoint reproduced identical results; seed-to-seed training variance was not assessed.

TABLE IV. MAIN RESULTS — ALL MODELS ON CS2_37 (271 CYC.) AND CS2_38 (337 CYC.)

Model	**CS2_37**					**CS2_38**				
	RMSE	**R^2**	**RMSE%**	**W10**	**W20**	**RMSE**	**R^2**	**RMSE%**	**W10**	**W20**
GRU	**4.39**	**+0.996**	**1.6%**	**96.3%**	**100.0%**	**7.48**	**+0.993**	**2.2%**	**83.5%**	**100.0%**
LSTM	6.57	+0.991	2.4%	90.5%	96.3%	11.71	+0.983	3.5%	70.9%	92.2%
Transformer	12.40	+0.969	4.6%	48.1%	99.2%	18.40	+0.957	5.5%	36.2%	57.3%
PhysicsMambaRUL	13.60	+0.962	5.0%	43.2%	84.8%	20.49	+0.947	6.1%	29.4%	64.1%
MambaRUL	28.54	+0.834	10.5%	32.1%	45.3%	38.57	+0.813	11.4%	22.7%	37.9%

With the 13-feature input, the GRU reaches RMSE = 4.39 cycles and R^2 = +0.996 on CS2_37 and RMSE = 7.48 cycles and R^2 = +0.993 on CS2_38, and all of its predictions on both cells fall within 20 cycles of the true RUL. The LSTM follows closely, at 6.57 and 11.71 cycles with W20 = 96.3% on CS2_37 and 92.2% on CS2_38, so the feature set is usable by both recurrent baselines. PhysicsMambaRUL gives 13.60 and 20.49 cycles (R^2 = +0.962 and +0.947), close to the Transformer baseline, which indicates that the physics-constrained multi-head output is workable but confers no advantage here. Stated plainly, the feature-enhanced GRU is more accurate than either proposed Mamba variant. Under this protocol, the choice of input representation appears to have a larger effect on test-cell RMSE than the choice among the evaluated sequence models; that reading applies to the present experiment and is not extended to other datasets or chemistries. MambaRUL with the default n_anchors = 3 setting has the highest error of the five models, at 28.54 and 38.57 cycles. The exploratory ablation in Section IV-D found lower RMSE with n_anchors = 2 on the held-out cells; because that ablation was run on the test cells, it was not used to select the reported configuration.

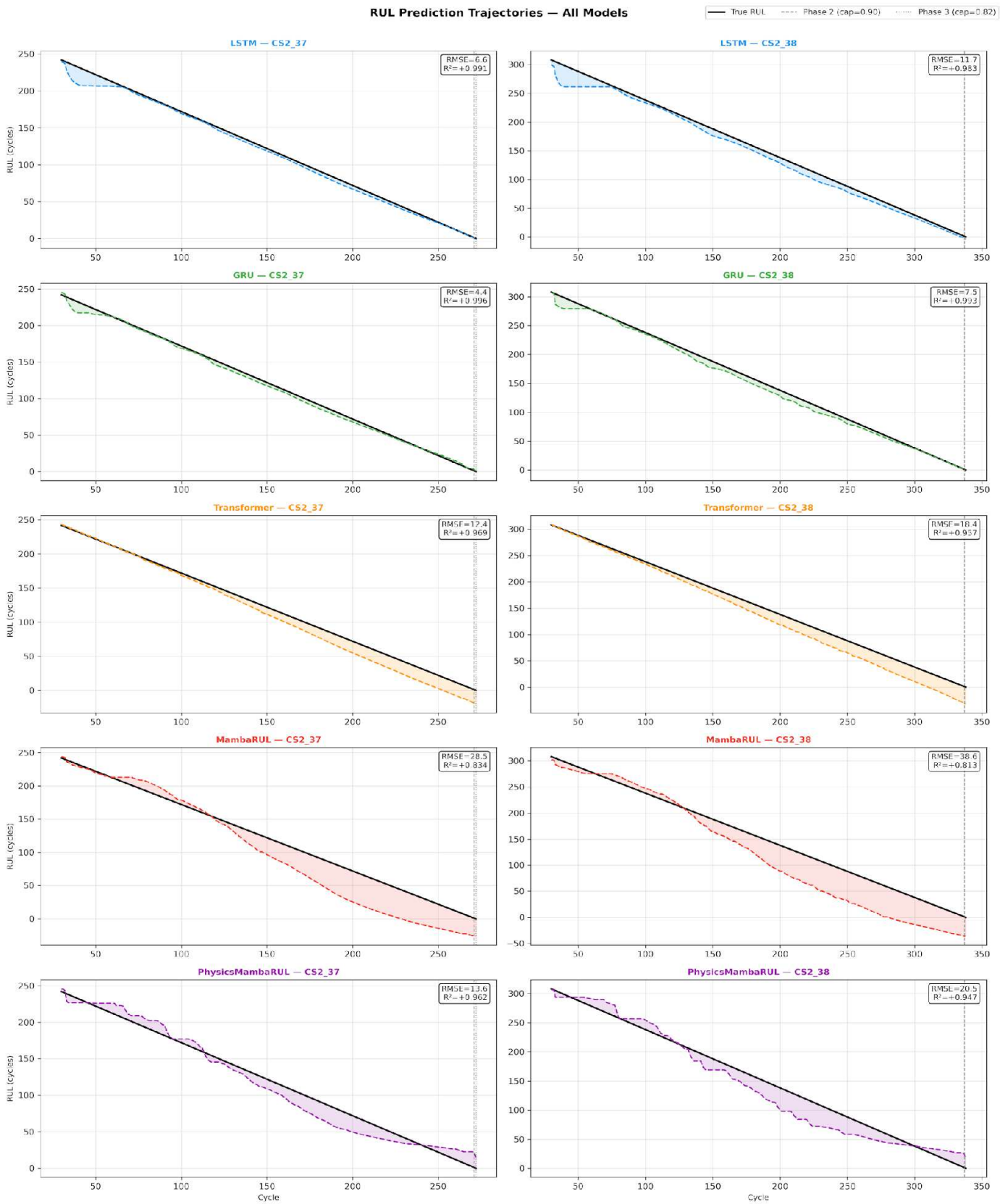


Fig. 2. Predicted and true RUL trajectories for the five models on CS2_37 (left) and CS2_38 (right). The GRU and LSTM traces follow the true trajectory across the full degradation span, while MambaRUL over-predicts in early life.

B. Literature Comparison

Published battery RUL methods are set alongside this work in Table V. The comparison is qualitative: reported error values are not restated, because datasets, cells, EOL definitions, prediction horizons, train/test partitions, target variables and error metrics differ across studies. Within the CALCE CS2_37/CS2_38 protocol used here, the GRU trained on the 13-feature set gives lower RMSE (4.39 cycles on CS2_37) than the two Mamba architectures evaluated under the same conditions. U-H-Mamba reports results on a hybrid laboratory and real-world dataset and is therefore not evaluated under the protocol used here.

TABLE V. QUALITATIVE COMPARISON WITH REPRESENTATIVE BATTERY RUL METHODS

Method	**Year**	**Dataset**	**Main methodology**	**Directly comparable to this work?**
XGBoost-HT [1]	2024	synthetic + LFP fast-charging cells	gradient boosting + feature engineering	No
CNN-LSTM [7]	2023	NASA	CNN + LSTM + grey relational analysis	No
MLP-Mixer/Expert [8]	2024	NASA/CALCE	MLP-Mixer + mixture of experts	Partially
TCN + data extension [2]	2024	NASA/CALCE	CEEMDAN-TCN with data extension	Partially
VMD-SSA-LSTM [4]	2025	NASA/CALCE	VMD + SSA + LSTM	Partially
Hybrid ensembles [3]	2025	CALCE CS2	GB + RF + kNN voting ensemble	Partially
RUL-Mamba [5]	2025	NASA/Oxford/Tongji	Mamba + FAN + GRN	No
SambaMixer [6]	2025	NASA	Mamba-based SOH estimation	No
U-H-Mamba [9]	2026	hybrid laboratory/real-world datasets	uncertainty-aware hierarchical Mamba	No
This work	2026	CALCE CS2_37/CS2_38	13 physics-informed features + GRU/Mamba/PhysicsMamba RUL	Yes

Numerical results reported in the literature are not necessarily directly comparable because datasets, cells, EOL definitions, prediction horizons, train/test partitions, target variables, and error metrics differ across studies.

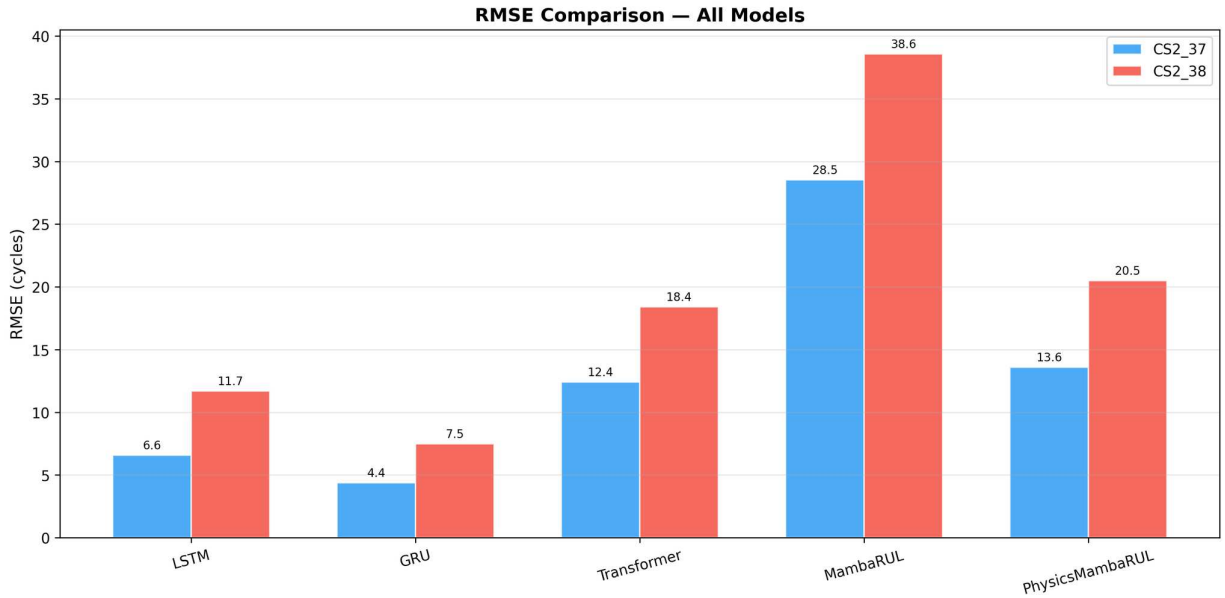


Fig. 3. RMSE of the models evaluated in this study on CS2_37 and CS2_38. Published results are not plotted, because reported RMSE values are not directly comparable to this work: datasets, cells, EOL definitions, prediction horizons, train/test partitions, target variables and error metrics differ across studies (see Table V).

C. Ablation Study: cap_pct Feature

The contribution of cap_pct is isolated by training MambaRUL twice, once with all 13 features and once with the same set minus cap_pct, holding every other hyperparameter fixed. Results are given in Table VI.

TABLE VI. CAP_PCT FEATURE ABLATION

Configuration	37 RMSE	37 R^2	38 RMSE	38 R^2
12 features (excl. cap_pct)	36.81	+0.725	42.69	+0.771
13 features (incl. cap_pct)	**25.98**	**+0.863**	**31.02**	**+0.879**

Removing cap_pct raises RMSE from 25.98 to 36.81 cycles on CS2_37 and from 31.02 to 42.69 cycles on CS2_38, so its inclusion accounts for reductions of 10.83 cycles (29.4%) and 11.67 cycles (27.3%) on the two held-out cells under the present protocol. A plausible explanation is that, without cap_pct, the current health state has to be inferred from the relative levels and trends of the voltage, current and temperature features, which is a noisy high-dimensional inference; with cap_pct present, health is available directly as a scalar, leaving the network to learn the mapping from fade rate to remaining cycles rather than to reconstruct health from indirect signals.

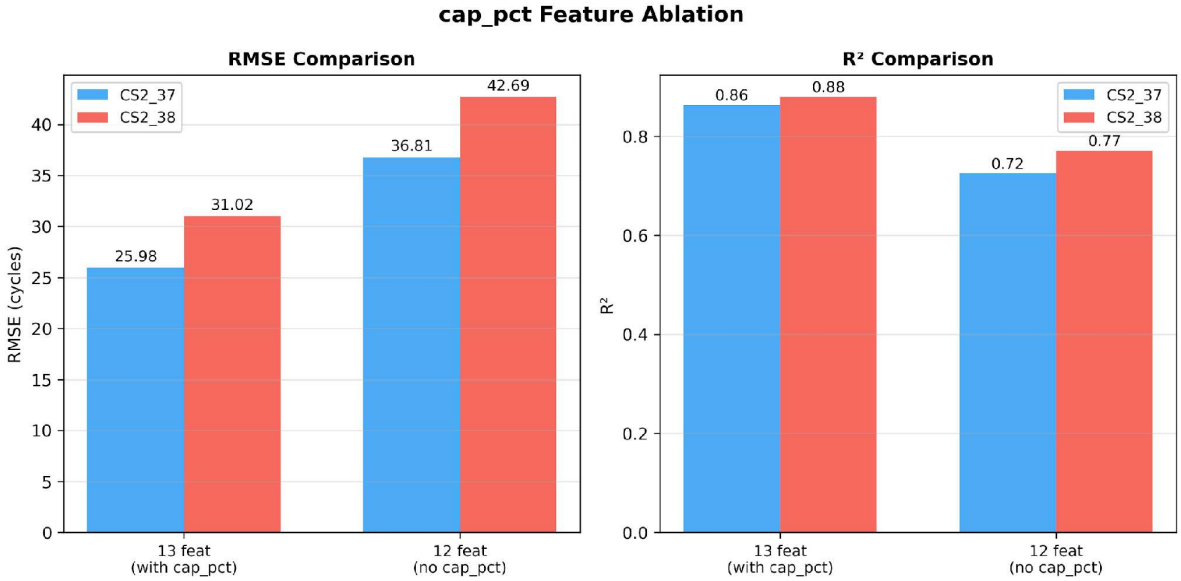


Fig. 4. Effect of cap_pct on MambaRUL test RMSE. Including the feature lowers RMSE by 10.83 cycles on CS2_37 and 11.67 cycles on CS2_38; among the feature configurations evaluated in this study, this was the largest observed reduction.

D. Ablation Study: Degradation Anchor Count

The effect of the number of degradation anchors in DACA is examined by training MambaRUL with n_anchors $\in \{0, 1, 2, 3\}$, where n_anchors = 0 corresponds to the Mamba encoder without cross-attention. Results are given in Table VII.

TABLE VII. DEGRADATION ANCHOR COUNT ABLATION (N = 2 HIGHLIGHTED; EXPLORATORY TEST-SET ANALYSIS)

n	Configuration	37 RMSE	37 R^2	38 RMSE	38 R^2
0	No DACA (vanilla Mamba)	21.59	+0.905	27.94	+0.902
1	Single-phase anchor	32.06	+0.791	44.29	+0.753
2	**Early + late anchors**	**15.15**	**+0.953**	**14.91**	**+0.972**
3	Default (3-phase)	25.98	+0.863	31.02	+0.879

Within this exploratory ablation, the two-anchor configuration produced the lowest RMSE on the two held-out cells (15.15 and 14.91 cycles), below the three-anchor configuration in this ablation by 10.83 and 16.11 cycles. This result requires confirmation on validation data before being considered for model selection, and the reported configuration therefore remains n_anchors = 3. One interpretation, offered as a hypothesis rather than an established finding, is that two anchors—one associated with early-life SEI growth and one with late-life lithium plating—cover the phases that carry most of the fade signal in these LCO cells, and that the single-anchor case, which is the weakest of the four, collapses two dissimilar regimes onto one query.

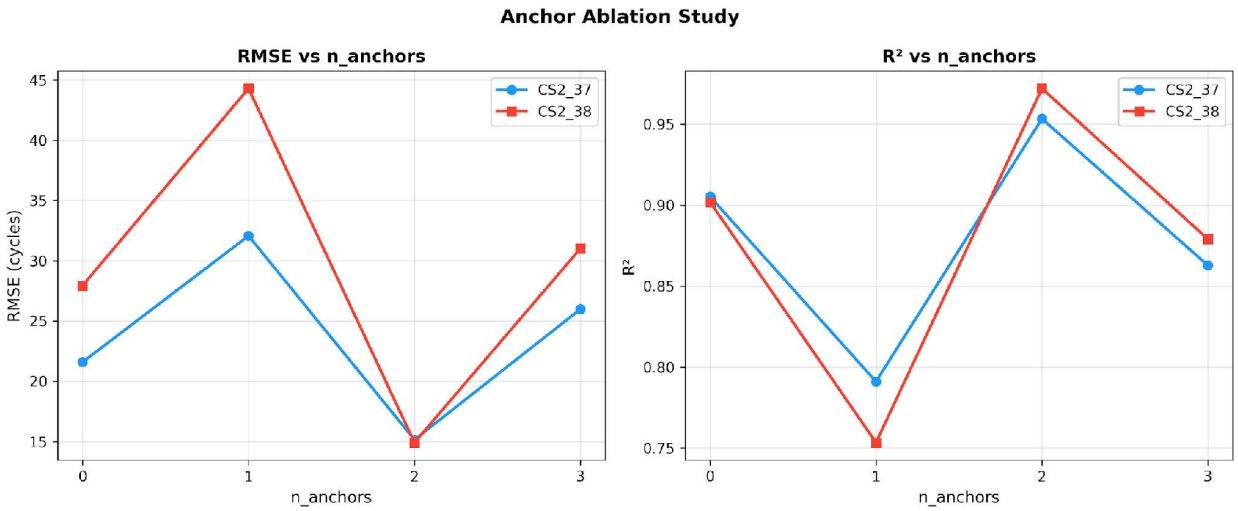


Fig. 5. Degradation Anchor Cross-Attention ablation. Among the configurations evaluated on the held-out test cells, n = 2 anchors produced the lowest RMSE. The analysis was exploratory and was not used for model selection.

E. Physics Analysis

The physics-consistency coefficient λ was varied on the held-out test cells, with results in Table VIII; the learned head weights across degradation phases are reported in Table IX. Because this sensitivity analysis was carried out on the final test cells, it was exploratory and was not used for model selection or hyperparameter tuning. The value λ = 0.3 was fixed a priori.

TABLE VIII. PHYSICS CONSISTENCY LOSS ABLATION

λ	37 RMSE	37 R^2	38 RMSE	38 R^2
0.0 (no phys. loss)	19.47	+0.923	23.11	+0.933
0.3 (phys. loss active)	**14.72**	**+0.956**	**22.12**	**+0.938**

TABLE IX. PHYSICSMAMBARUL HEAD WEIGHT EVOLUTION BY DEGRADATION PHASE (CS2_37 / CS2_38)

Phase	cap_pct	Windows	w_1	w_2	w_3	Interpretation
Early	>0.90	237/295	0.397	0.411	0.192	Rate and direct heads similar
Mid	0.85–0.90	5/13	0.338	0.474	0.188	Transition: fade rate gains weight
Late	≤0.85	1/1	0.327	**0.485**	0.188	Largest weight on fade-rate head

With λ = 0.3, PhysicsMambaRUL RMSE falls from 19.47 to 14.72 cycles on CS2_37 (24.4%) and from 23.11 to 22.12 cycles on CS2_38 (4.3%). The two cells respond differently, and no mechanism is established here; one possibility is that the shorter CS2_37 trajectory has fewer close analogues among the training cells, so the additional constraint carries more weight. Both figures come from the exploratory test-cell analysis described above.

The learned combination weights change with degradation state (Table IX). In early life (cap_pct > 0.90) the fade-rate head and the direct head carry comparable weight (w_2 = 0.411, w_1 = 0.397); in late life (cap_pct ≤ 0.85) the fade-rate head carries the largest share (w_2 = 0.485) and the direct head falls to w_1 = 0.327, while the curve-fitting head stays near 0.19 throughout. The observed change in head weights is consistent with the increasing contribution of the fade-rate estimate at lower cap_pct values. No phase labels were supplied during training, so the trend was not imposed directly; it is reported as an observation about the learned weighting and not as evidence that the electrochemical mechanisms have been recovered. The late-life row rests on a single window per cell and should be treated as indicative only.

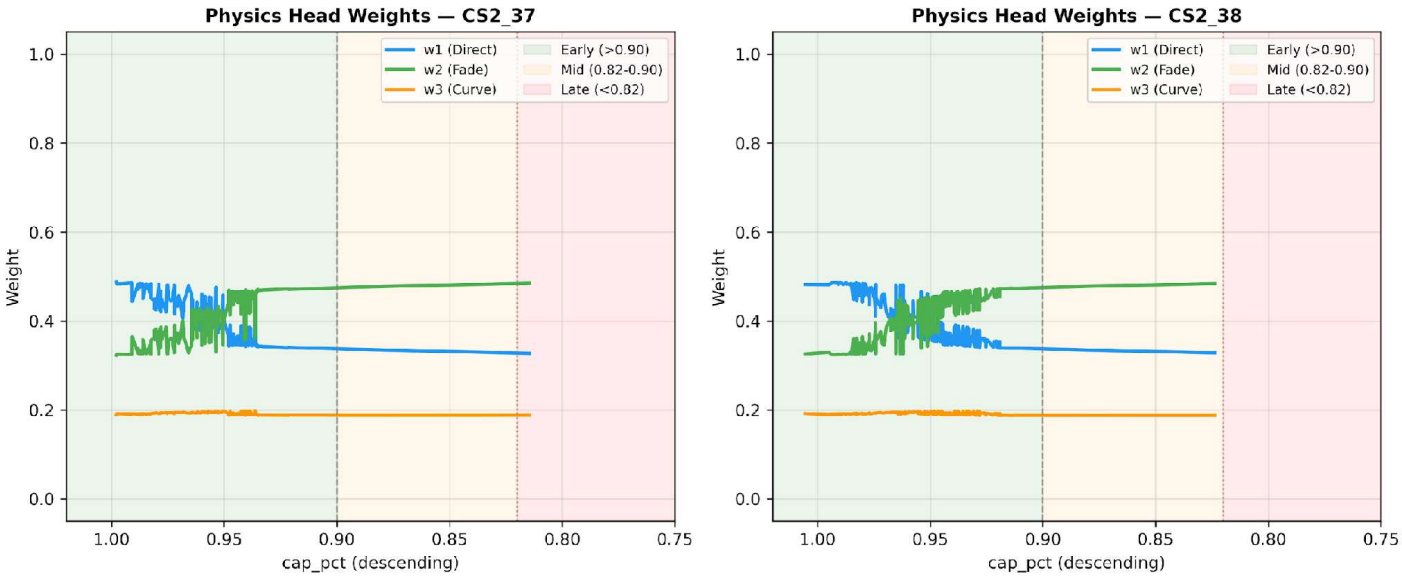


Fig. 6. Learned head weights against degradation state. The fade-rate head weight w_2 rises from 0.411 in early life to 0.485 in late life, w_1 falls over the same range, and w_3 stays approximately constant. The late-life bin contains few windows.

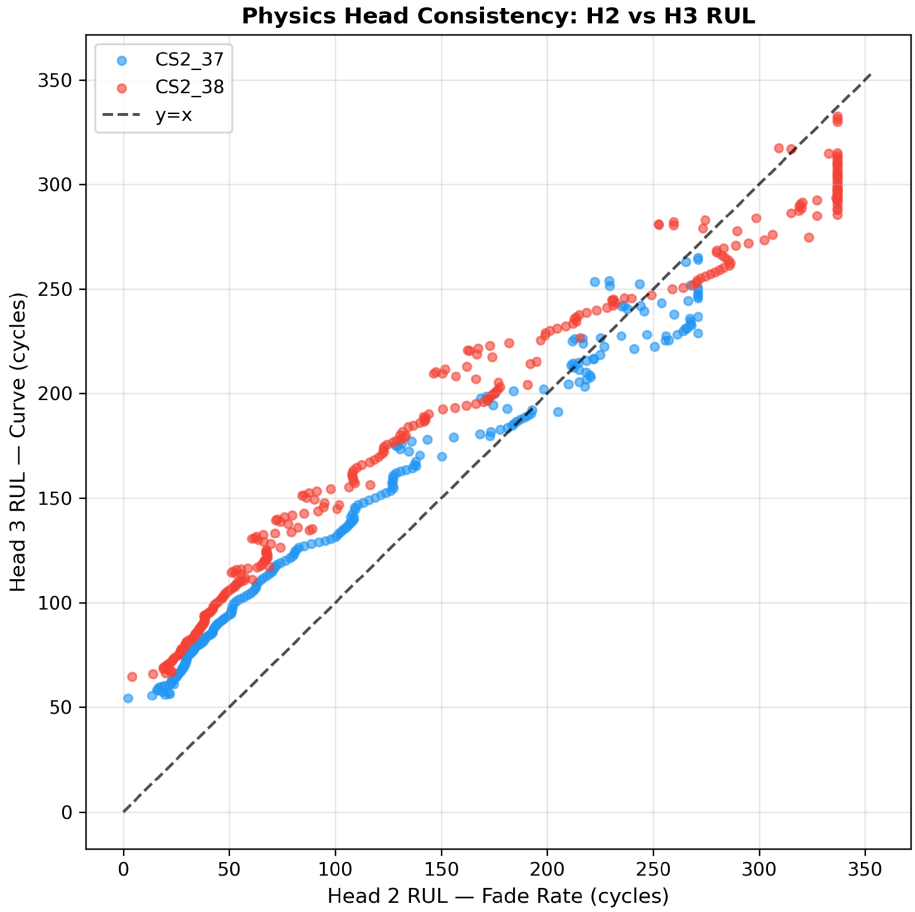


Fig. 7. Consistency between the two physics heads of PhysicsMambaRUL: the Head 3 (double-exponential curve) RUL estimate plotted against the Head 2 (fade-rate) estimate for CS2_37 and CS2_38, with the y = x line for reference. Most points lie above the diagonal, so the curve head returns the larger estimate over much of the range, with closest agreement at low RUL.

F. Per-Cell Analysis

Per-cell results are collected in Table X. CS2_38 gives higher error than CS2_37 for every model. Its lifetime of 337 cycles exceeds that of the longest training cell (309 cycles) by 9.1%, so part of its trajectory lies outside the lifetime range represented

in training. Under the present protocol, the model maintained reasonable performance on CS2_38 despite its longer lifetime than the training cells: the GRU reaches RMSE = 7.48 cycles (R^2 = +0.993, W20 = 100%) and the LSTM RMSE = 11.71 cycles (R^2 = +0.983). Since L_cell is supplied at denormalisation, this observation concerns behaviour under the benchmark protocol and does not establish generalisation to unseen lifetimes in an online setting.

TABLE X. PER-CELL RESULTS — ALL MODELS (LIFETIMES: CS2_37 = 271, CS2_38 = 337 CYCLES)

Model	Cell	RMSE (cyc.)	R^2	RMSE%
GRU	**CS2_37**	**4.39**	**+0.996**	**1.6%**
GRU	**CS2_38**	**7.48**	**+0.993**	**2.2%**
LSTM	CS2_37	6.57	+0.991	2.4%
LSTM	CS2_38	11.71	+0.983	3.5%
Transformer	CS2_37	12.40	+0.969	4.6%
Transformer	CS2_38	18.40	+0.957	5.5%
PhysicsMambaRUL	CS2_37	13.60	+0.962	5.0%
PhysicsMambaRUL	CS2_38	20.49	+0.947	6.1%
MambaRUL	CS2_37	28.54	+0.834	10.5%
MambaRUL	CS2_38	38.57	+0.813	11.4%

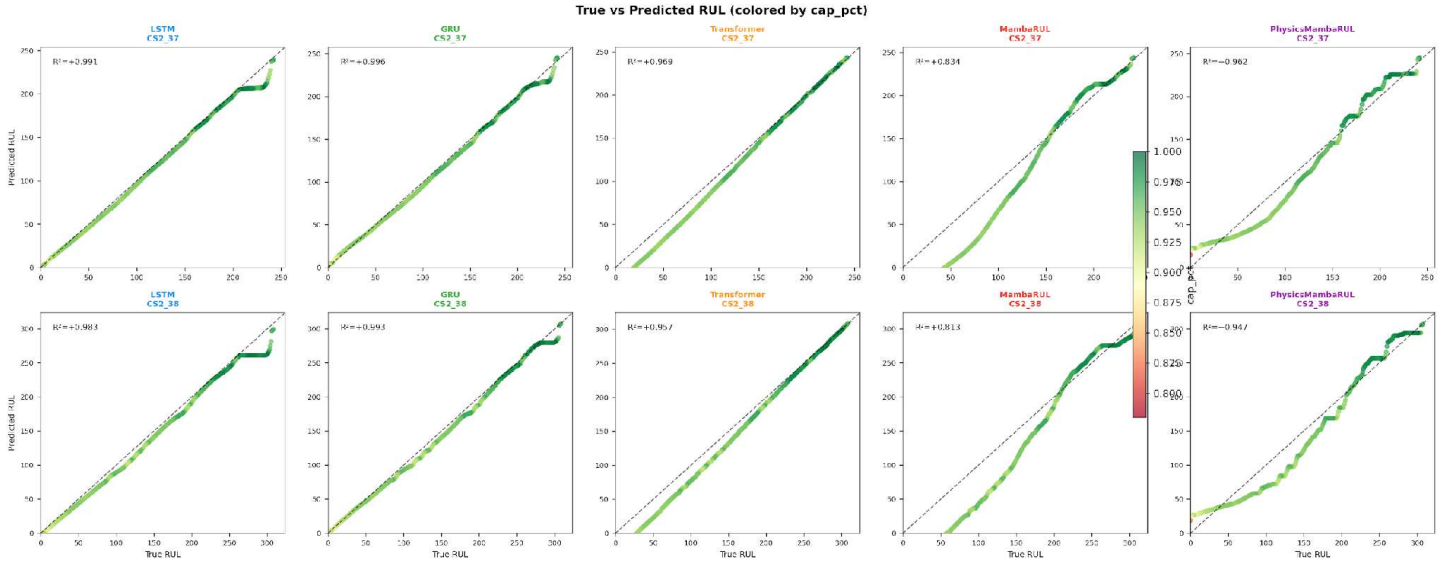


Fig. 8. True against predicted RUL for all models on CS2_37 and CS2_38. Points near the diagonal y = x indicate accurate predictions; the GRU shows the tightest clustering around the diagonal on both cells.

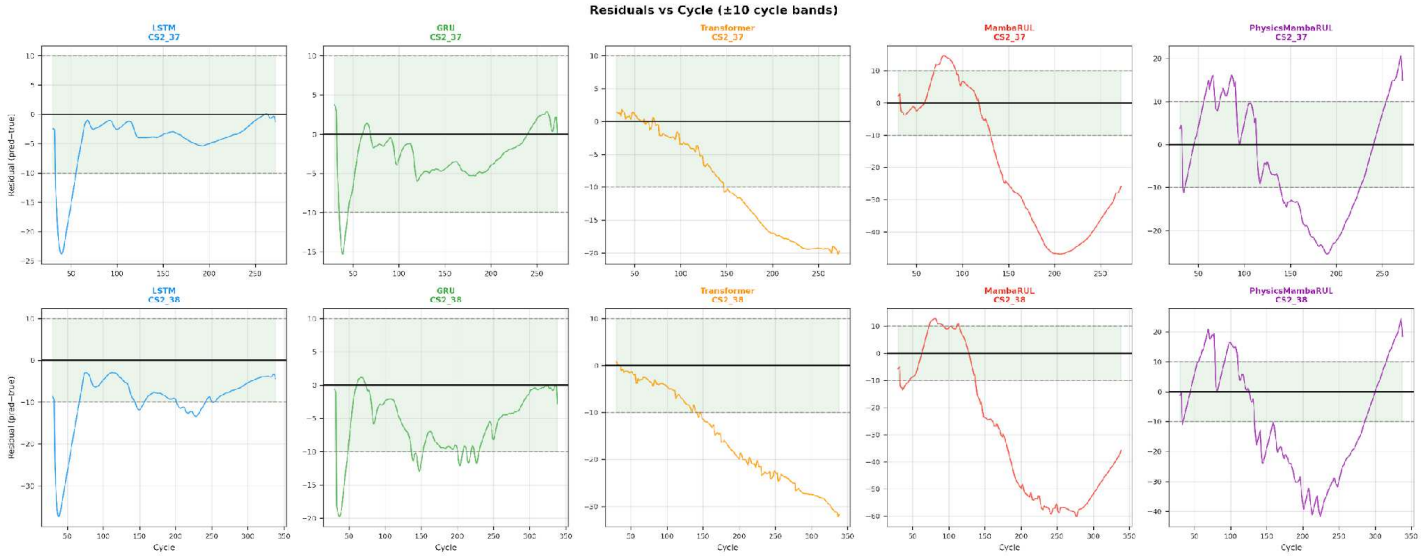


Fig. 9. Residuals (predicted − true RUL) against cycle number. GRU residuals remain comparatively small across the test trajectory, while MambaRUL shows a positive early-life bias.

G. Summary

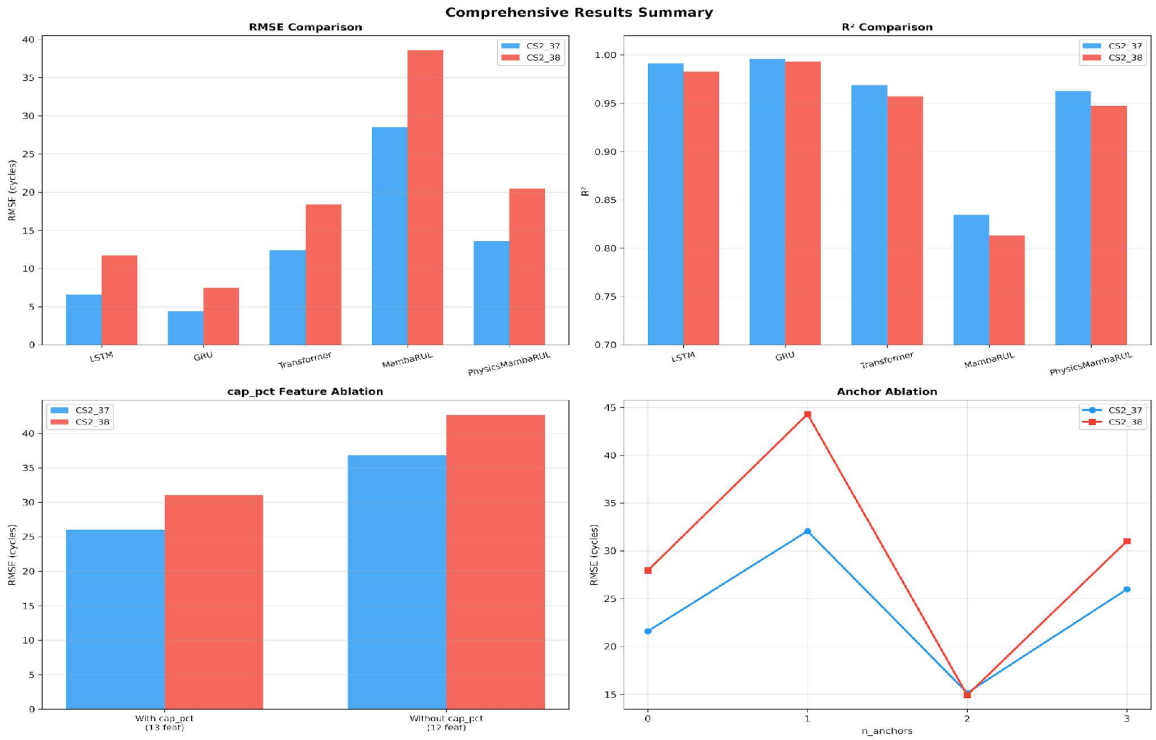


Fig. 10. Summary of RMSE and R² across the models and ablation configurations evaluated in this study. Green bars mark the proposed methods. Published baselines are not plotted, because reported RMSE values are not directly comparable to this work: datasets, EOL thresholds, train/test splits, feature sets and metric definitions differ (see Table V).

V. CONCLUSION

A physics-informed 13-feature representation of each charge–discharge cycle and two Mamba-based architectures, MambaRUL and PhysicsMambaRUL, were evaluated for lithium-ion battery RUL prediction on the CALCE CS2 dataset, against LSTM, GRU and Transformer baselines trained on the same features, the same split and the same seed.

Under the test protocol described in Section III, the GRU trained on the 13-feature set produced the lowest RMSE among the evaluated models, at 4.39 cycles on CS2_37 (R^2 = +0.996) and 7.48 cycles on CS2_38 (R^2 = +0.993), with all predictions falling within 20 cycles on both cells. PhysicsMambaRUL reached 13.60 and 20.49 cycles and MambaRUL 28.54 and 38.57 cycles, so neither proposed architecture improved on the recurrent baselines in this comparison.

Three ablations were carried out. Removing cap_pct increased MambaRUL RMSE by 10.83 cycles on CS2_37 and 11.67 cycles on CS2_38, the largest change observed among the feature configurations evaluated. In an exploratory anchor-count ablation, n_anchors = 2 gave the lowest RMSE of the four settings tested on the held-out cells. In an exploratory test-cell analysis, the physics-consistency term with $\lambda = 0.3$ lowered PhysicsMambaRUL RMSE relative to $\lambda = 0$ on both cells. The latter two comparisons were performed on the final test cells and were not used to select the reported configuration.

Several conditions bound the scope of these results. RUL targets are normalised by the complete observed lifetime of each cell, which is an oracle quantity at inference, so the cycle-domain errors are benchmark values rather than online prediction accuracy. Each configuration was trained once with seed = 42, leaving run-to-run variance unquantified. The anchor-count and λ analyses were exploratory and test-set based. The dataset is a laboratory benchmark cycled under controlled conditions, and no field-aged cells were included. Cross-chemistry behaviour was not evaluated.

The corresponding directions for further work follow from these conditions: replacing the oracle lifetime with an online estimate derived only from cycles observed up to the prediction time, for example from early-cycle capacity-fade slopes; selecting λ and the anchor count on validation data; repeating training across several seeds to quantify variance; and extending the evaluation to further datasets, other cell chemistries and field-recorded data, including duty cycles representative of Indian EV operation. The present results should therefore be interpreted as a controlled CALCE CS2 benchmark rather than a deployment evaluation.

ACKNOWLEDGMENT

The CALCE Battery Research Group at the University of Maryland is acknowledged for publicly providing the CS2 dataset used in this work. This research is conducted as part of the M.Tech thesis work at MNNIT Allahabad under the supervision of Dr Praveen Kumar Agarwal.

REFERENCES

[1]V. Safavi, A. Mohammadi Vaniar, N. Bazmohammadi, J. C. Vasquez, and J. M. Guerrero, "Battery Remaining Useful Life Prediction Using Machine Learning Models: A Comparative Study," Information, vol. 15, no. 3, p. 124, 2024, doi: 10.3390/info15030124.

[2]J. Zhao, D. Liu, and L. Meng, "Remaining Useful Life Prediction of a Lithium-Ion Battery Based on a Temporal Convolutional Network with Data Extension," International Journal of Applied Mathematics and Computer Science, vol. 34, no. 1, pp. 105-117, 2024, doi: 10.61822/amcs-2024-0008.

[3]S. Zhao, D. Sun, Y. Liu, and Y. Liang, "Remaining Useful Life Prediction for Lithium-Ion Batteries Based on Hybrid Ensembles Allied with Data-Driven Approach," Energies, vol. 18, no. 5, p. 1114, 2025, doi: 10.3390/en18051114.

[4]J. Li, Z. Wang, D. Tian, and Z. Sun, "Prediction Method for Remaining Service Life of Lithium Batteries Using SSA-LSTM Combination under Variable Mode Decomposition," Energy Storage Science and Technology, vol. 14, no. 2, pp. 659-670, 2025, doi: 10.19799/j.cnki.2095-4239.2024.0732.

[5]J. Huang, L. Liu, H. Zhao, T. Li, and B. Li, "RUL-Mamba: Mamba-Based Remaining Useful Life Prediction for Lithium-Ion Batteries," J. Energy Storage, vol. 120, p. 116376, 2025, doi: 10.1016/j.est.2025.116376.

[6]J. I. Olalde-Verano, S. Kirch, C. Perez-Molina, and S. Martin, "SambaMixer: State of Health Prediction of Li-Ion Batteries Using Mamba State Space Models," IEEE Access, vol. 13, pp. 2313-2327, 2025, doi: 10.1109/ACCESS.2024.3524321.

[7]D. Chen, X. Zheng, C. Chen, and W. Zhao, "Remaining Useful Life Prediction of the Lithium-Ion Battery Based on CNN-LSTM Fusion Model and Grey Relational Analysis," Electronic Research Archive, vol. 31, no. 2, pp. 633-655, 2023, doi: 10.3934/era.2023031.

[8]L. Zhao, S. Song, P. Wang, C. Wang, J. Wang, and M. Guo, "A MLP-Mixer and Mixture of Expert Model for Remaining Useful Life Prediction of Lithium-Ion Batteries," Front. Comput. Sci., vol. 18, p. 185329, 2024, doi: 10.1007/s11704-023-3277-4.

[9]Z. Wen, X. Liu, W. Niu, H. Zhang, and Y. Cheng, "U-H-Mamba: An Uncertainty-Aware Hierarchical State-Space Model for Lithium-Ion Battery Remaining Useful Life Prediction Using Hybrid Laboratory and Real-World Datasets," Energies, vol. 19, no. 2, p. 414, 2026, doi: 10.3390/en19020414.

[10]A. Gu and T. Dao, "Mamba: Linear-Time Sequence Modeling with Selective State Spaces," arXiv:2312.00752, 2024.

[11]A. Vaswani et al., "Attention Is All You Need," Advances in Neural Information Processing Systems, vol. 30, 2017.

[12]W. He, N. Williard, M. Osterman, and M. Pecht, "Prognostics of Lithium-Ion Batteries Based on Dempster-Shafer Theory and the Bayesian Monte Carlo Method," J. Power Sources, vol. 196, no. 23, pp. 10314-10321, 2011.

[13]K. A. Severson et al., "Data-Driven Prediction of Battery Cycle Life Before Capacity Degradation," Nature Energy, vol. 4, pp. 383-391, 2019, doi: 10.1038/s41560-019-0356-8.