batteries

MDPI

Article

Multimodal State of Health Prediction for Lithium-Ion Batteries via Mamba-Based Fusion of Discharge Curves and Impedance Spectra

Yawei Meng [1,2], Qiang Sun [1,2], Jianping Xu [3,*], Antai Bian [1,2], Qizheng Yang [4], Zhi Wang [5,6,*], Zijian Yang [1,2] and Maoyong Zhi [1,2,*]

[1] College of Civil Aviation Safety Engineering, Civil Aviation Flight University of China, Guanghan 618307, China; yaweimeng@cafuc.edu.cn (Y.M.); qiangsun@cafuc.edu.cn (Q.S.); 15650422396@163.com (A.B.); 18048593578@163.com (Z.Y.)
[2] All-Electric General Aviation Aircraft Key Technology Engineering Research Center of Sichuan Province, Guanghan 618307, China
[3] School of Electrical Engineering, Southwest Jiaotong University, Chengdu 611756, China
[4] Mathematical Institute, St Hilda's College, University of Oxford, Oxford OX2 6GG, UK; shil6697@ox.ac.uk
[5] Shenzhen Research Institute, China University of Mining and Technology, Shenzhen 518057, China
[6] School of Safety Engineering, China University of Mining and Technology, Xuzhou 221116, China
* Correspondence: jpxu-swjtu@163.com (J.X.); zhiwang@cumt.edu.cn (Z.W.); zhimaoyong@cafuc.edu.cn (M.Z.)

Abstract

Existing deep learning methods for lithium-ion battery State of Health (SOH) prediction rely almost exclusively on discharge voltage–current curves, ignoring electrochemical impedance spectroscopy (EIS) data that directly reflects internal degradation mechanisms. Fusing these two modalities is non-trivial: discharge curves are high-dimensional temporal sequences residing on a continuous dynamical manifold, while impedance features are low-dimensional static snapshots with fundamentally different statistical distributions. However, naive concatenation introduces modal conflicts rather than complementary gains. We propose the Hybrid Sensing Synergy Architecture (HSSA), which combines a Mamba backbone ($O(L)$ complexity) for discharge curve modeling with a Q-former module that aligns impedance features into the temporal representation space via learnable query tokens and cross-attention. A prepend fusion strategy injects the aligned queries as prefix tokens, enabling the backbone to condition on internal electrochemical context from the first time step. On the NASA battery dataset, HSSA achieves MAE of 0.887 (large-scale, 11 batteries, a 9.8% improvement over unimodal Mamba), 1.457 (medium-scale, five batteries, a 28.0% improvement), and 2.705 (small-scale, four batteries, an 8.7% improvement), demonstrating consistent improvements across all data regimes. On out-of-sample battery B28, HSSA achieves 65.3% improvement. Ablation studies confirm that Q-former alignment is essential and prepend fusion significantly outperforms concatenation-based alternatives.

Keywords: lithium-ion battery; State of Health (SOH); Mamba; selective state space model; multimodal fusion; electrochemical impedance spectroscopy (EIS); Q-former

Academic Editor: Yong-Joon Park

Received: 14 April 2026
Revised: 20 May 2026
Accepted: 26 May 2026
Published: 29 May 2026

1. Introduction

Lithium-ion batteries have evolved from simple energy storage units into critical components that determine the safety, economics, and sustainability of modern energy systems [1,2]. Meanwhile, to pursue higher energy density and better safety, advanced materials such as solid polymer and composite solid-state electrolytes for lithium metal batteries

are also being rapidly developed [3–5]. In electric vehicles, grid-scale storage, aerospace power systems, and portable electronics, batteries must operate reliably under complex and time-varying conditions. However, inevitable degradation processes, including capacity fade, internal resistance growth, thermal characteristic deterioration, and power output decline, directly impact system performance, maintenance strategies, and life cycle costs [6,7]. Accurate characterization of battery degradation state and early prediction of remaining useful life have become core scientific challenges in battery management systems (BMSs), intelligent operation and maintenance, and energy storage optimization [8,9].

From an engineering perspective, battery life prediction is not an isolated curve-fitting problem, but a comprehensive state-awareness task oriented toward real-world operational scenarios [10]. On one hand, BMS must continuously monitor current health status during operation to adjust charging/discharging strategies, limit extreme conditions, and prevent risk accumulation. On the other hand, systems need to infer future degradation trends based on current health levels to support decision-making for maintenance, replacement, retirement, and second-life utilization [11]. Especially in automotive and grid storage scenarios where numerous battery cells and modules operate simultaneously, lack of reliable life prediction capability makes fine-grained management infeasible and prevents balancing safety, economics, and sustainable utilization [12].

Despite years of research, lithium-ion battery life prediction remains fundamentally complex [1,2]. First, battery degradation is not a linear, stationary, or monotonic process, but results from coupled physicochemical mechanisms [6]. Continuous growth of the solid electrolyte interphase (SEI) on the anode, active lithium loss, electrode material structural damage, electrolyte side reactions, and interfacial impedance growth all affect battery performance at different stages with varying intensities. Second, the relationship between external observation signals and internal degradation mechanisms is not one-to-one. Many degradation changes first accumulate at the microscopic level, then gradually manifest through capacity fade, polarization enhancement, and thermal response changes. Third, battery operation is influenced by external factors such as temperature, C-rate, rest duration, sampling noise, and operating condition transitions, causing batteries of the same type to exhibit significantly different degradation trajectories under different conditions. This means life prediction models must not only identify long-term degradation trends but also distinguish between genuine degradation and short-term fluctuations. Existing research has transitioned from mechanism-based models to data-driven deep learning methods [8,12,13], but still faces challenges. Recurrent neural networks (RNN/LSTM) suffer from memory bottlenecks: conventional recurrent structures are theoretically limited by fixed-dimensional hidden states, leading to inevitable information compression loss and gradient vanishing when processing discharge sequences spanning thousands of time steps [14,15]. This “recurrence bottleneck” causes models to struggle with capturing deep causal relationships between early degradation signals and current states in later sequence stages.Transformers solve long-range dependency problems through global receptive fields, but their original self-attention mechanism suffers from $O(L^2)$ time and space complexity, introducing severe memory pressure for long sequences [16,17]. Modern optimizations have attempted to alleviate these issues: hardware-aware exact attention like FlashAttention [18] optimizes memory access to accelerate training, yet its theoretical computational bottleneck inherently remains quadratic. Conversely, approximate linear attention [19] achieves $O(L)$ complexity by replacing the softmax operation with kernel functions, but this introduces severe representational trade-offs. Specifically, linear attention suffers from “attention dilution”: without softmax, it fails to form sharp, “peaky” distributions, causing it to blur localized transient anomalies (e.g., sudden voltage drops) critical for battery degradation analysis. Furthermore, compressing historical context into a fixed-capacity state

matrix leads to poor associative recall over long sequences, where early subtle degradation signatures are often washed out by accumulated noise. Therefore, exploring sequence modeling paradigms that natively combine linear complexity with expressive, input-dependent context selection remains highly necessary for complex electrochemical time series.

In recent years, selective state space models (SSMs) represented by Mamba have opened a third path for sequence modeling [20,21]. By introducing hardware-aware scanning mechanisms and dynamic gating strategies, Mamba achieves Transformer-like global dependency modeling capability while maintaining linear computational complexity $O(L)$. From a signal processing perspective, Mamba essentially constructs a learnable continuous dynamical system that can adaptively filter noise and focus on key features reflecting battery degradation [22]. In the battery domain specifically, Kirch et al. [22] first applied Mamba (SambaMixer) to SOH prediction using discharge curves alone, demonstrating superior accuracy over LSTM and Transformer baselines. However, all existing Mamba-based battery research remains confined to the unimodal setting, processing only external macroscopic response signals (voltage, current, temperature) without incorporating internal electrochemical measurements. To our knowledge, no prior work has extended Mamba to multimodal fusion of discharge dynamics with impedance spectroscopy. This paper fills this gap by proposing a principled cross-modal alignment mechanism that enables Mamba to jointly reason over both modalities.

In reality, lithium-ion battery degradation is a multi-scale coupled process involving electrochemical dynamics and macroscopic phenomenological features. From a multi-physics evolution perspective, battery failure manifests not only as nonlinear capacity fade in external charge–discharge responses, but more fundamentally originates from deep evolution of internal electrochemical impedance [23]. Impedance features, as fingerprints of battery internal dynamic characteristics, directly map subtle changes in electrode interface polarization, charge transfer resistance, and ion diffusion limitations [24]. Their sensitivity far exceeds macroscopic capacity, exhibiting significant early warning properties. From a representation learning logic, discharge time series curves capture the system's external dynamic trajectory, while impedance features carry static snapshots of underlying mechanisms. Combining both essentially constructs a holistic health representation where external dynamic patterns and internal physical constraints mutually support each other.

However, this multimodal fusion faces a severe semantic gap. Discharge curves reside on a high-dimensional temporal manifold characterized by continuous dynamics, long-range autocorrelation, and smooth local geometry, whereas impedance features occupy a low-dimensional static manifold with sparse, frequency-domain structure, and fundamentally different covariance statistics. Concretely, the per-token feature variance of discharge sequences is typically one to two orders of magnitude smaller than that of impedance vectors, and their principal component directions are nearly orthogonal. Naive concatenation implicitly assumes that these two manifolds can be linearly aligned in a shared feature space. Unfortunately, this assumption breaks down for heterogeneous modalities with mismatched dimensionality, information density, and gradient scales. In practice, concatenation forces the backbone to process tokens with vastly different statistical properties using a single set of parameters, leading to gradient conflicts where one modality dominates optimization at the expense of the other [8]. This explains why simple fusion strategies often degrade rather than improve performance (as we confirm empirically in Section 4).

The core proposition of this work is therefore: how to bridge the distributional gap between macroscopic temporal responses and microscopic impedance snapshots through nonlinear latent space alignment. Rather than assuming linear compatibility, we employ a Q-former [25] that learns a set of query tokens to project impedance features onto the

temporal representation manifold via cross-attention, effectively performing a learned nonlinear mapping between the two feature spaces. By introducing impedance information as aligned contextual tokens rather than raw concatenated features, we enable the model to leverage complementary degradation signals without suffering from modal conflicts, thereby extending unimodal correlation modeling toward multimodal complementary fusion for battery health prediction.

Based on this, we propose a multimodal architecture for lithium-ion battery life prediction that integrates Mamba-based temporal modeling with cross-modal impedance fusion, which investigates multimodal SOH prediction from discharge dynamics and electrochemical impedance spectroscopy, with particular focus on the cross-modal alignment mechanism. Performance gains require explicitly resolving the distributional mismatch between modalities; naive feature concatenation is insufficient. Experiments on the NASA and CALCE datasets across varying data scales, with systematic ablations, validate this conclusion.

Our contributions are:

- **Multimodal SOH Prediction Framework**: We design a three-stage multimodal framework combining Mamba backbone [21] for discharge curve modeling, impedance MLP encoder for internal degradation feature extraction, and Q-former [25] for cross-modal alignment. To our knowledge, this is among the first works to systematically integrate electrochemical impedance spectroscopy [23] with discharge dynamics for battery SOH prediction using state space models.
- **Q-former-based semantic alignment**: We introduce a query-based cross-attention mechanism [25] that maps sparse impedance features into a unified semantic space compatible with temporal discharge representations. Unlike naive concatenation, Q-former learns to extract degradation-relevant information from impedance spectra and align it with temporal dynamics through learnable query tokens.
- **Prepend fusion strategy**: We propose injecting aligned impedance query tokens as prefix tokens into the Mamba backbone, allowing the model to dynamically integrate internal mechanism signals with external response patterns. This strategy significantly outperforms concatenation-based and pooling-based fusion approaches.
- **Comprehensive experimental validation**: On the NASA battery dataset [26], our method achieves MAE 0.887 (9.8% improvement) on the large-scale setting and MAE 1.457 (28.0% improvement) on the medium-scale setting. On out-of-sample battery B28, our method achieves 65.3% improvement over the unimodal baseline, demonstrating strong generalization to unseen batteries. Systematic ablation studies demonstrate that Q-former alignment is essential and prepend fusion with 16 queries and three-layer depth provides optimal performance.

The rest of this paper is organized as follows. Section 2 reviews related work on battery SOH prediction and multimodal fusion. Section 3 presents our multimodal architecture design. Section 4 reports experimental results and ablation studies. Section 5 concludes with limitations and future directions.

2. Related Work

2.1. Battery State of Health Prediction

Early battery SOH prediction methods relied on equivalent circuit models (ECMs) and electrochemical models [27] that describe internal battery dynamics through differential equations. While these physics-based approaches provide interpretability, they require extensive parameter calibration and struggle to generalize across different battery chemistries and operating conditions [24]. The rise in data-driven methods has shifted the paradigm toward learning degradation patterns directly from operational data [28]. Support vector

regression (SVR) and Gaussian process regression (GPR) were among the first machine learning techniques applied to battery prognostics [29], offering uncertainty quantification but limited capacity for modeling complex temporal dependencies. Recent deep learning approaches have achieved significant improvements [30,31]. Convolutional neural networks (CNNs) extract local patterns from charge–discharge curves, while recurrent architectures such as LSTM [14] and GRU model temporal evolution of battery states [15]. However, these methods face fundamental limitations: LSTMs suffer from vanishing gradients and fixed-capacity hidden states when processing long sequences spanning thousands of cycles, while CNNs lack explicit temporal modeling. Transformer-based models [16] address long-range dependencies through self-attention but incur quadratic complexity $\mathcal{O}(L^2)$, making them computationally prohibitive for high-frequency battery monitoring data [17]. Our work leverages Mamba's linear complexity selective state space mechanism to efficiently capture long-term degradation dynamics while maintaining global context [22]. We note that efficient Transformer variants [18,19] also reduce the quadratic bottleneck; however, their advantages are most pronounced at large-scale pretraining—a regime unavailable in battery monitoring. More importantly, Mamba's selective gating provides a distinct inductive bias (dynamic noise filtering via input-dependent state transitions) that attention-based approximations do not replicate.

2.2. State Space Models and Mamba

State space models (SSMs) provide a principled framework for sequence modeling by representing dynamics as continuous-time linear systems. Structured SSMs such as S4 [20] achieve competitive performance on long-range sequence tasks by parameterizing state matrices with diagonal plus low-rank structures, enabling efficient computation via convolution and recurrence views. However, S4 and its variants (S5 [32], H3 [33]) use time-invariant parameters, limiting their ability to selectively focus on relevant input features. Mamba [21] introduces a selective mechanism where SSM parameters (B, C, and time scale Δ) are input-dependent, computed via learned projections from input tokens. This selectivity enables the model to filter irrelevant information and propagate task-relevant signals through the state space, analogous to gating mechanisms in LSTMs [14] but with hardware-efficient parallel scanning. Mamba has demonstrated strong performance on language modeling, genomics, and audio processing tasks. Our work is among the first to apply Mamba to battery health prediction and extend it to multimodal fusion with impedance data [22].

Table 1 summarizes the theoretical complexity of the three main sequence modeling paradigms. The key distinction is that Transformer's $\mathcal{O}(NL^2d)$ time complexity and $\mathcal{O}(NL^2)$ memory cost grow quadratically with sequence length, making it impractical for long battery monitoring sequences. Both LSTM and Mamba achieve linear time complexity, but LSTM cannot be parallelized during training and loses global context through its fixed-capacity hidden state. Mamba uniquely combines linear complexity, full global context via selective state propagation, and training parallelism via the associative scan algorithm.

Table 1. Theoretical complexity comparison of sequence models (N layers, sequence length L, model dim d, state dim d_{state}, hidden size h). FLOPs estimated at $L = 128$, $d = 512$, $N = 4$, $d_{\text{state}} = 16$, $h = 256$.

Model	Time Complexity	Memory	Global Context	Parallelizable	FLOPs
LSTM	$\mathcal{O}(NL(hd + h^2))$	$\mathcal{O}(NLh)$	No (fades)	No	≈100 M
Transformer	$\mathcal{O}(NL^2d)$	$\mathcal{O}(NL^2)$	Yes (full)	Yes	≈168 M
Mamba	$\mathcal{O}(NLd)$	$\mathcal{O}(NLd_{\text{state}})$	Yes (state)	Yes	≈4.2 M

2.3. Multimodal Learning for Time Series

Multimodal learning aims to leverage complementary information from heterogeneous data sources to improve prediction performance and robustness. In time series domains, multimodal fusion has been explored for applications such as human activity recognition (combining accelerometer and gyroscope data), healthcare monitoring (fusing ECG, EEG, and physiological signals), and autonomous driving (integrating LiDAR, camera, and radar). Common fusion strategies include early fusion (concatenating raw features), late fusion (combining predictions from modality-specific models), and intermediate fusion (aligning representations in a shared latent space). Cross-modal attention mechanisms, inspired by vision–language models such as CLIP [34] and Flamingo, have emerged as effective tools for aligning heterogeneous modalities. The Q-former architecture, originally proposed for vision–language pretraining in BLIP-2 [25], uses learnable query tokens and cross-attention layers to extract task-relevant information from one modality and bridge it to another. However, existing multimodal time series work primarily focuses on homogeneous temporal modalities (e.g., multiple sensor streams with similar sampling rates and dimensionality). Battery health prediction presents a unique challenge: discharge curves are high-dimensional temporal sequences ($\sim$1000 time steps), while impedance spectra are low-dimensional static features ($\sim$10–50 dimensions) with fundamentally different physical meanings. From a distributional perspective, these two modalities occupy distinct statistical manifolds: discharge features exhibit smooth temporal autocorrelation and low per-token variance, whereas impedance features are sparse, frequency-structured, and have substantially higher variance per dimension. Simple concatenation or fully connected projection assumes linear compatibility between these manifolds, which fails when the modalities differ in dimensionality, information density, and gradient magnitude. Our Q-former-based alignment addresses this gap by learning a nonlinear mapping from impedance features into the temporal representation space through cross-attention with learnable query tokens, enabling each query to selectively extract degradation-relevant information from specific frequency bands without imposing a linear alignment assumption.

2.4. Critical Comparison with Battery-Specific Multimodal Methods

Existing battery SOH methods that incorporate multiple data sources fall short in three ways. First, feature-level combination approaches [30] concatenate EIS-derived statistics (e.g., charge transfer resistance extracted from equivalent circuit fitting) with capacity measurements and pass them to a shallow regressor, entirely discarding the temporal structure of discharge dynamics. Second, late fusion approaches append static aggregate features (mean voltage, temperature) to recurrent hidden states [31], which cannot model the causal influence of evolving internal impedance on discharge trajectories. Third, simple projection approaches project impedance vectors into the model's embedding space via a linear layer before concatenation, which—as we demonstrate empirically in Section 4—degrades performance on large datasets because the linear projection cannot resolve the distributional mismatch between the two modalities. Our Q-former alignment addresses all three limitations by learning a nonlinear, attention-based mapping from impedance frequency tokens to the temporal discharge representation space.

2.5. Why Q-Former for Heterogeneous Modality Alignment

We formalize the semantic gap to motivate the Q-former design. Let $\mathbf{H}_{\text{dis}} \in \mathbb{R}^{T_{\text{dis}} \times d_{\text{model}}}$ and $\mathbf{Z}_{\text{imp}} \in \mathbb{R}^{(T_{\text{imp}}/2) \times d_{\text{model}}}$ denote the projected discharge and impedance token sequences,

respectively ($T_{\text{dis}} \gg T_{\text{imp}}/2$). A linear fusion approach (concatenation followed by a shared linear layer) implicitly assumes:

$$\mathbf{h}_{\text{fused}} = \mathbf{W}[\mathbf{H}_{\text{dis}}; \mathbf{Z}_{\text{imp}}] + \mathbf{b} \tag{1}$$

This is equivalent to assuming that the joint distribution $p(\mathbf{H}_{\text{dis}}, \mathbf{Z}_{\text{imp}})$ lies on a linear subspace. However, the two modalities exhibit fundamentally different statistical structures. Discharge tokens have low per-token variance with strong temporal autocorrelation ($\text{Cov}(\mathbf{h}_t, \mathbf{h}_{t+\tau}) \gg 0$ for small τ), while impedance tokens have high per-token variance with weak inter-token correlation. The Maximum Mean Discrepancy (MMD) between the two feature distributions quantifies this gap:

$$\text{MMD}^2(\mathcal{H}_{\text{dis}}, \mathcal{Z}_{\text{imp}}) = \|\mathbb{E}[\phi(\mathbf{h})] - \mathbb{E}[\phi(\mathbf{z})]\|^2_{\mathcal{H}_k} \tag{2}$$

where $\phi(\cdot)$ is a kernel feature map. When MMD^2 is large, linear alignment in Equation (1) introduces systematic bias, as the shared weight matrix $\mathbf{W}$ cannot simultaneously minimize reconstruction error for both modalities. We use an RBF (Gaussian) kernel $k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2/2\sigma^2)$ with bandwidth σ set by the median heuristic ($\sigma^2 = \text{median}\{\|\mathbf{x}_i - \mathbf{x}_j\|^2\}$) [35]. We emphasize that the MMD statistic here serves as a qualitative diagnostic to motivate the alignment mechanism, rather than a formal hypothesis test; the exact value depends on kernel and feature normalization choices. In contrast, the Q-former performs a nonlinear, attention-based alignment:

$$\mathbf{Q}_{\text{out}} = \text{CrossAttn}(\mathbf{Q}_{\text{learn}}, \mathbf{Z}_{\text{imp}}, \mathbf{Z}_{\text{imp}}) \in \mathbb{R}^{N_{\text{query}} \times d_{\text{model}}} \tag{3}$$

where $\mathbf{Q}_{\text{learn}} \in \mathbb{R}^{N_{\text{query}} \times d_{\text{model}}}$ are learnable query tokens. This mechanism has three key advantages over linear fusion: (1) the cross-attention computes a data-dependent weighted combination of impedance tokens, enabling nonlinear feature selection; (2) multiple queries ($N_{\text{query}} > 1$) preserve multi-aspect information without collapsing to a single vector; and (3) the output queries $\mathbf{Q}_{\text{out}}$ are explicitly trained to be compatible with the discharge token distribution, minimizing the effective MMD in the fused representation space.

The Q-former was originally designed for vision–language alignment in BLIP-2 [25], where image patch sequences and text token sequences have analogous heterogeneity. In our battery setting, impedance frequency–pair tokens and discharge time step tokens exhibit a structurally similar asymmetry: a small set of static, information-dense tokens (impedance) must be aligned with a long sequence of temporally correlated tokens (discharge). The Q-former's learnable query mechanism is naturally suited to this scenario, as it can selectively attend to different frequency bands (low-frequency diffusion vs. high-frequency charge transfer) and produce aligned tokens that the Mamba backbone processes seamlessly alongside discharge tokens.

2.6. Electrochemical Impedance Spectroscopy in Battery Diagnosis

Electrochemical impedance spectroscopy (EIS) measures battery impedance across a range of frequencies, providing insights into internal processes such as charge transfer resistance, double-layer capacitance, and solid electrolyte interphase (SEI) growth [23]. EIS has been widely used in battery diagnostics for state estimation, fault detection, and degradation mechanism analysis [24]. Traditional approaches fit EIS data to equivalent circuit models to extract physical parameters [27], but this requires domain expertise and model selection. Recent machine learning methods treat EIS as feature vectors for SOH estimation, using techniques such as principal component analysis (PCA) for dimensionality reduction and neural networks for regression [30,31]. However, most existing work uses impedance

features in isolation or combines them with simple statistical features (e.g., mean capacity, cycle count) rather than rich temporal discharge dynamics [36]. A few recent studies have explored multimodal fusion of EIS and voltage curves, but rely on simple concatenation or separate processing followed by late fusion, which fails to capture the deep coupling between external response and internal mechanisms. Our work is among the first to use a learnable cross-modal alignment mechanism (Q-former [25]) to bridge impedance and temporal discharge representations in a unified state space model framework.

2.7. Positioning of This Work

Our work sits at the intersection of three research threads: (1) efficient sequence modeling via Mamba's selective SSM, (2) multimodal fusion through cross-modal attention, and (3) battery health prediction using both external dynamics and internal impedance. Unlike prior battery SOH methods that rely solely on discharge curves or use impedance in isolation, we systematically integrate both modalities through a principled alignment mechanism. Unlike general multimodal time series methods that assume homogeneous modalities, we address the unique semantic gap between high-dimensional temporal sequences and low-dimensional static features. Our prepend fusion strategy allows the Mamba backbone to dynamically integrate impedance information throughout the temporal processing, rather than treating it as a separate branch or post hoc combination. This design is motivated by the physical intuition that internal impedance evolution constrains and explains external discharge behavior, and should therefore inform the model's temporal reasoning at every step.

3. Method

We propose a multimodal architecture for battery SOH prediction that integrates Mamba-based temporal modeling of discharge curves with cross-modal fusion of impedance features. Figure 1 illustrates the overall framework. The Multi-Modal Encoder processes diverse inputs including complex-valued impedance data $a + bj$ alongside dynamic signals such as temperature, current, and voltage. A Cross-Modal Alignment module facilitates feature synchronization by employing self-attention and cross-attention layers to align impedance features with temporal representations through learnable query tokens. These aligned representations are then integrated into a Samba module, where a Mamba backbone [21] processes discharge time series with linear complexity while an impedance encoder extracts low-dimensional degradation features from EIS data [23]. The system utilizes a Q-former module [25] and an MLP to combine concatenated features and classification tokens for final SOH estimation through the predict head. We describe each component in detail below.

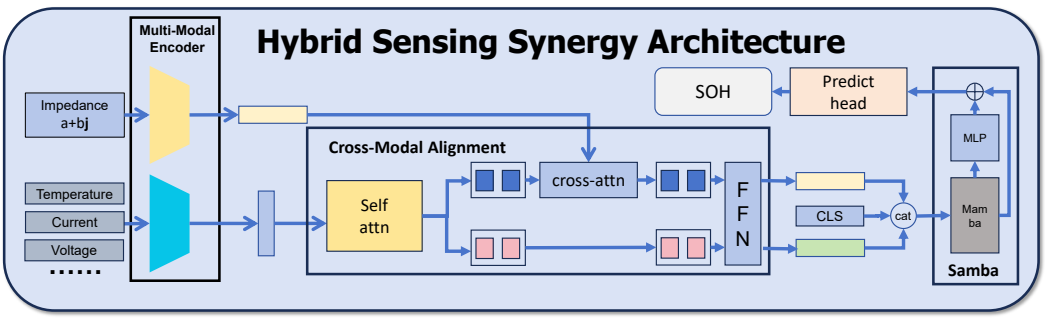

Figure 1. An overview of our system design.

3.1. Problem Formulation

Let $\mathbf{X}_{\text{dis}} \in \mathbb{R}^{T_{\text{dis}} \times 4}$ denote a discharge curve sequence, where T_{dis} is the sequence length and the 4 channels correspond to voltage, current, temperature, and timestamp. Let $\mathbf{f}_{\text{imp}} \in \mathbb{R}^{T_{\text{imp}}}$ denote the raw impedance feature vector extracted from EIS measurements, where T_{imp} is the impedance feature dimension (typically 10–50). For each discharge cycle k, we denote the corresponding discharge sequence as $\mathbf{X}_{\text{dis}}^{(k)}$ recorded at time $t_{\text{dis}}^{(k)}$, and the associated impedance feature as $\mathbf{f}_{\text{imp}}^{(k)}$, which is taken from the most recent preceding EIS measurement at time $t_{\text{EIS}}^{(k)}$. We enforce a strict causal constraint:

$$t_{\text{EIS}}^{(k)} < t_{\text{dis}}^{(k)}, \quad \forall k \tag{4}$$

ensuring that no future impedance information leaks into the prediction for cycle k. When EIS measurements are not available at every cycle (typical in practice, with EIS acquired every 5 to 10 cycles), we use the most recent preceding measurement: $\mathbf{f}_{\text{imp}}^{(k)} = \mathbf{f}_{\text{imp}}^{(k')}$ where $k' = \max\{j < k : \text{EIS available at cycle } j\}$.

Our goal is to predict the State of Health SOH $\in [0, 100]$, defined as the ratio of current capacity to nominal capacity:

$$\text{SOH} = \frac{C_{\text{current}}}{C_{\text{nominal}}} \times 100\% \tag{5}$$

The key challenge is to learn a mapping $f : (\mathbf{X}_{\text{dis}}^{(k)}, \mathbf{f}_{\text{imp}}^{(k)}) \to \text{SOH}^{(k)}$ that effectively fuses heterogeneous modalities: $\mathbf{X}_{\text{dis}}^{(k)}$ is a high-dimensional temporal sequence capturing external battery response, while $\mathbf{f}_{\text{imp}}^{(k)}$ is a low-dimensional static feature vector reflecting internal electrochemical processes. The temporal alignment constraint above ensures that the fusion strategy must handle asynchronous, causally ordered modalities rather than perfectly synchronized data streams.

3.2. Mamba Backbone for Discharge Curve Modeling

The Mamba backbone processes discharge sequences through a stack of selective state space blocks. Each Mamba block operates as follows.

3.2.1. Selective State Space Model

Given an input sequence $\mathbf{x} \in \mathbb{R}^{L \times d_{\text{model}}}$, the selective SSM [21] computes output $\mathbf{y} \in \mathbb{R}^{L \times d_{\text{model}}}$ by first projecting $\mathbf{x}$ to obtain input-dependent parameters:

$$\mathbf{B} = \text{Linear}_B(\mathbf{x}), \quad \mathbf{C} = \text{Linear}_C(\mathbf{x}), \quad \Delta = \text{Softplus}(\text{Linear}_\Delta(\mathbf{x})) \tag{6}$$

where $\mathbf{B} \in \mathbb{R}^{L \times d_{\text{state}}}$, $\mathbf{C} \in \mathbb{R}^{L \times d_{\text{state}}}$, and $\Delta \in \mathbb{R}^{L}$ control the state transition, output projection, and time scale respectively. The SSM then evolves a hidden state $\mathbf{h}_t \in \mathbb{R}^{d_{\text{state}}}$ via discretized dynamics:

$$\mathbf{h}_t = \bar{\mathbf{A}}\mathbf{h}_{t-1} + \bar{\mathbf{B}}_t \mathbf{x}_t \tag{7}$$

$$\mathbf{y}_t = \mathbf{C}_t \mathbf{h}_t \tag{8}$$

where $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}_t$ are discretized versions of continuous-time parameters using zero-order hold. The key innovation is that $\mathbf{B}$, $\mathbf{C}$, and Δ are input-dependent, enabling the model to selectively propagate relevant information through the state space.

3.2.2. Bidirectional Processing

To capture both forward and backward temporal context, we use a bidirectional Mamba architecture that processes the sequence in both directions and concatenates the outputs:

$$\mathbf{y} = \text{Concat}(\text{Mamba}_{\text{fwd}}(\mathbf{x}), \text{Mamba}_{\text{bwd}}(\text{Reverse}(\mathbf{x}))) \tag{9}$$

3.2.3. Stacking and Residual Connections

The full backbone consists of N stacked Mamba blocks with residual connections and layer normalization:

$$\mathbf{h}^{(l+1)} = \text{LayerNorm}(\mathbf{h}^{(l)} + \text{Mamba}(\mathbf{h}^{(l)})) \tag{10}$$

This design achieves $O(L \cdot d_{\text{model}} \cdot d_{\text{state}})$ complexity, linear in sequence length L, compared to $O(L^2 \cdot d_{\text{model}})$ for Transformers [16].

3.2.4. Noise Handling via Selective Gating

The input-dependent parameters $\mathbf{B}$, $\mathbf{C}$, and Δ provide inherent robustness to measurement noise and non-stationary battery conditions. When the model encounters an anomalous token (e.g., a voltage transient caused by measurement noise or a recuperation-induced plateau), the learned time scale Δ can effectively gate down the state update for that position, preventing transient disruptions from corrupting the accumulated degradation state. This behavior is qualitatively different from LSTM's additive gating—which cannot selectively suppress a single time step's influence on the state—and from Transformer's position-unaware attention, which distributes anomalous signal globally.

3.2.5. Implementation Notes

The state matrix $\bar{\mathbf{A}}$ is initialized using the HiPPO-LegS construction, which provides theoretically optimal initial memory for polynomial history approximation. The parameters $\mathbf{B}$ and $\mathbf{C}$ are initialized from $\mathcal{N}(0, 0.01)$; Δ is initialized such that $\text{softplus}(\Delta_0) \approx 0.001$, following the original Mamba implementation. Dropout ($p = 0.25$) is applied after each Mamba block's output projection; drop-path ($p = 0.25$) is applied on residual connections. AdamW weight decay ($\lambda = 0.05$) applies to linear projection weights only, not to SSM parameters. The SSM state dimension is fixed at $d_{\text{state}} = 16$ across all model scales.

3.3. Impedance Encoder

The raw impedance feature vector $\mathbf{f}_{\text{imp}} \in \mathbb{R}^{T_{\text{imp}}}$ contains T_{imp} scalar measurements (alternating real and imaginary parts at $T_{\text{imp}}/2$ frequencies). Rather than treating this as a flat vector, we exploit the physical structure of EIS: each frequency probes a distinct electrochemical process. High frequencies (>1 kHz) reflect ohmic resistance and inductive effects; mid-frequencies (ranging from 1 Hz to 1 kHz) capture charge transfer resistance and double-layer capacitance; low frequencies (<1 Hz) encode solid-state diffusion and SEI dynamics [23]. We therefore split $\mathbf{f}_{\text{imp}}$ into $T_{\text{imp}}/2$ frequency–pair tokens and project each to d_{model} dimensions:

$$\mathbf{Z}_{\text{imp}} = \text{Linear}\Big(\text{reshape}(\mathbf{f}_{\text{imp}}, \tfrac{T_{\text{imp}}}{2}, 2)\Big) \in \mathbb{R}^{\frac{T_{\text{imp}}}{2} \times d_{\text{model}}} \tag{11}$$

This yields a sequence of $T_{\text{imp}}/2$ impedance tokens, one per frequency band, which the Q-former can attend to selectively. A global summary vector is also computed via a 3-layer MLP for auxiliary use:

$$\mathbf{Z}_{\text{imp}} = \text{MLP}(\mathbf{f}_{\text{imp}}) \in \mathbb{R}^{d_{\text{model}}} \tag{12}$$

As shown in Figure 2, the Nyquist plots of batteries B6 and B7 exhibit a systematic outward shift of the impedance arc as SOH decreases, confirming that EIS captures internal degradation mechanisms complementary to external discharge dynamics.

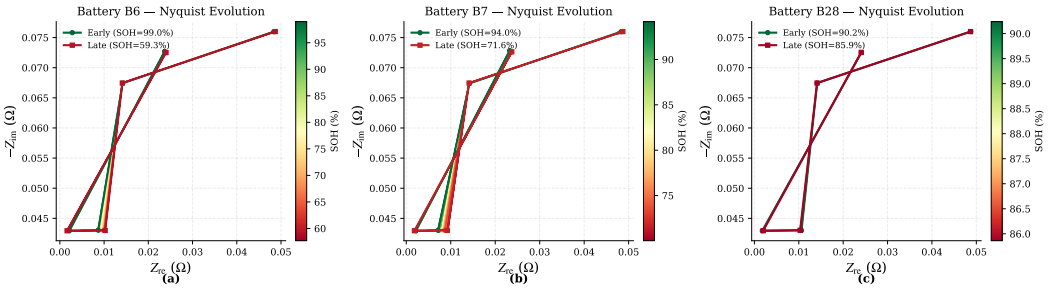

Figure 2. Nyquist plot evolution colored by SOH (%). (**a**) Battery B6. (**b**) Battery B7. (**c**) Battery B28. Each curve represents one impedance measurement at a given cycle. Green curves correspond to healthy states (high SOH) and red curves to degraded states (low SOH). The systematic outward shift of the impedance arc with degradation reflects growing charge transfer resistance and SEI layer thickness, motivating the use of EIS as a complementary modality for SOH prediction.

3.4. Q-Former for Cross-Modal Alignment

The core challenge in multimodal fusion is bridging the semantic gap between temporal discharge representations and static impedance features. We adopt a Q-former architecture [25] that uses learnable query tokens to extract and align impedance information with the temporal modality.

3.4.1. Learnable Query Tokens

We initialize N_{query} learnable query embeddings $\mathbf{Q} = \{\mathbf{q}_1, \ldots, \mathbf{q}_{N_{\text{query}}}\} \in \mathbb{R}^{N_{\text{query}} \times d_{\text{model}}}$. These queries serve as "soft prompts" that probe the impedance features for task-relevant information. The number of queries N_{query} governs alignment bandwidth. We use $N_{\text{query}} = 16 > \frac{T_{\text{imp}}}{2} = 5$ deliberately: electrochemical impedance exhibits frequency-dispersion effects where a single physical process (e.g., charge transfer resistance) contributes to multiple adjacent frequency measurements. Multiple queries per physical band allow the Q-former to learn soft, overlapping spectral decompositions rather than a hard 1-to-1 frequency assignment. The ablation study in Section 4 confirms that 8 queries limits capacity (MAE 0.952) while 32 queries introduces redundancy (MAE 0.985), with 16 striking the optimal balance.

3.4.2. Cross-Attention Mechanism

The Q-former consists of $N_{\text{Q-former}}$ transformer layers, each containing self-attention among queries and cross-attention from queries to the impedance token sequence $\mathbf{Z}_{\text{imp}} \in \mathbb{R}^{\frac{T_{\text{imp}}}{2} \times d_{\text{model}}}$:

$$\mathbf{Q}' = \text{SelfAttn}(\mathbf{Q}, \mathbf{Q}, \mathbf{Q}) \tag{13}$$

$$\mathbf{Q}'' = \text{CrossAttn}(\mathbf{Q}', \mathbf{Z}_{\text{imp}}, \mathbf{Z}_{\text{imp}}) \tag{14}$$

Each query attends over all $T_{\text{imp}}/2$ frequency tokens, allowing different queries to specialize in different frequency bands (e.g., low-frequency diffusion vs. high-frequency charge transfer). Self-attention enables queries to interact and form a coherent joint representation. Each Q-former layer also includes a feed-forward network and layer normalization following standard transformer conventions.

3.4.3. Output Aligned Queries

After $N_{\text{Q-former}}$ layers, the output queries $\mathbf{Q}_{\text{out}} \in \mathbb{R}^{N_{\text{query}} \times d_{\text{model}}}$ contain impedance information aligned to the temporal representation space. These queries are designed to be compatible with the Mamba backbone's input format.

3.5. Prepend Fusion Strategy

We propose a prepend fusion strategy that injects aligned impedance queries as prefix tokens to the discharge sequence before feeding it to the Mamba backbone. Let $\mathbf{H}_{\text{dis}} \in \mathbb{R}^{T_{\text{dis}} \times d_{\text{model}}}$ denote the projected discharge sequence (after linear projection and positional encoding). The fused sequence is:

$$\mathbf{H}_{\text{fused}} = \text{Concat}(\mathbf{Q}_{\text{out}}, \mathbf{H}_{\text{dis}}) \in \mathbb{R}^{(N_{\text{query}}+T_{\text{dis}}) \times d_{\text{model}}} \tag{15}$$

The Q-former outputs $\mathbf{Q}_{\text{out}} \in \mathbb{R}^{N_{\text{query}} \times d_{\text{model}}}$ are prepended to the discharge token sequence, forming an augmented input of length $N_{\text{query}} + T_{\text{dis}}$ that is fed into the Mamba backbone. Critically, the prepended impedance-derived tokens are treated identically to discharge tokens within all Mamba blocks: they participate in the same selective state transitions ($\bar{\mathbf{A}}$, $\bar{\mathbf{B}}$, $\mathbf{C}$, Δ) with shared parameters, with no masking or special handling. This design allows impedance context to propagate into the discharge hidden state through the SSM's recurrent dynamics from the very first discharge token onward.

We adopt prepend fusion rather than appending impedance tokens at the sequence end, for two reasons grounded in the SSM's causal structure.

Information propagation. In a unidirectional SSM, the hidden state at position t is a function of inputs at all positions $\leq t$. Prepending impedance tokens at positions $1, \ldots, N_{\text{query}}$ ensures that every discharge token at position $N_{\text{query}} + t$ has access to the full impedance context encoded in the carried-forward hidden state. Appending would place impedance tokens after all discharge positions, making them causally inaccessible to every discharge hidden state; they would only influence the final output if an additional readout mechanism attended to them, which our architecture does not include.

Gradient path length. Prepending reduces the recurrent gradient path from the prediction loss back to the impedance branch by T_{dis} steps compared to appending, providing shorter, more stable gradient flow for training the Q-former and impedance encoder. Impedance measurements provide a compact global descriptor of the battery's internal electrochemical state at a given point in time: charge transfer resistance, SEI thickness, and diffusion characteristics are all encoded in the impedance spectrum before the discharge begins. The Q-former converts this static descriptor into N_{query} task-aligned latent prompts, each specializing in a different aspect of the internal state. By prepending these prompts to the discharge token sequence, the Mamba backbone can condition its selective state propagation on the internal electrochemical context from the very first discharge token, rather than treating impedance as a post hoc correction. This is analogous to prefix-tuning in language models [25], where task-relevant context is injected as prefix tokens to guide the model's attention throughout the sequence.

Unlike mean pooling, which compresses N_{query} queries into a single vector, prepend fusion retains all query tokens, allowing the model to capture multiple aspects of impedance information. The discharge sequence remains intact, preserving its temporal ordering and allowing the Mamba backbone to model temporal dependencies without disruption.

3.6. Prediction Head and Training Objective

A learnable CLS token $\mathbf{e}_{\text{CLS}} \in \mathbb{R}^{d_{\text{model}}}$ is prepended to the fused sequence before the Mamba backbone, yielding the full input:

$$\mathbf{H}_{\text{input}} = \text{Concat}(\mathbf{e}_{\text{CLS}}, \mathbf{Q}_{\text{out}}, \mathbf{H}_{\text{dis}}) \in \mathbb{R}^{(1+N_{\text{query}}+T_{\text{dis}})\times d_{\text{model}}} \tag{16}$$

After the Mamba backbone processes this sequence, we extract the output hidden state at the CLS position $\mathbf{h}_{\text{CLS}} \in \mathbb{R}^{d_{\text{model}}}$ and pass it through a linear prediction head:

$$\hat{\text{SOH}} = \text{Linear}(\mathbf{h}_{\text{CLS}}) \tag{17}$$

The CLS token aggregates global context from both the impedance queries and the discharge sequence through the Mamba backbone's selective state propagation.

We train the model end-to-end using mean squared error (MSE) loss:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\text{SOH}}_i - \text{SOH}_i)^2 \tag{18}$$

where N is the batch size. We use AdamW optimizer with cosine annealing learning rate schedule and gradient clipping for stable training.

3.7. Architecture Variants and Hyperparameters

We experiment with three model scales: **MultiModal-S** ($d_{\text{model}} = 256$, 4 Mamba layers, 2 Q-former layers, 8 queries), **MultiModal-M** ($d_{\text{model}} = 512$, 8 Mamba layers, 3 Q-former layers, 16 queries), and **MultiModal-L** ($d_{\text{model}} = 1024$, 12 Mamba layers, 4 Q-former layers, 32 queries). Unless otherwise stated, all main results use **MultiModal-M**, which provides the best trade-off between accuracy and computational cost (see ablation in Section 4).

For all experiments, we use $d_{\text{state}} = 16$ for the SSM state dimension, dropout rate 0.25, drop-path rate 0.25, and batch size 32. Training runs for 100 epochs with early stopping based on validation loss.

4. Experiments

4.1. Experimental Setup

4.1.1. Dataset and Splits

We conducted evaluations using the NASA battery aging dataset [26], which contains charge–discharge cycling data from 28 lithium-ion 18,650 cells tested under varied temperature and load conditions. Each cycle provides voltage, current, temperature, and timestamp measurements during discharge, alongside periodic electrochemical impedance spectroscopy (EIS) measurements [23]. We construct three dataset scales of increasing size—NASA-S, NASA-M, and NASA-L—to study how data availability interacts with multimodal fusion. Table 2 summarizes the splits. Batteries B6 and B7 serve as primary test batteries across all scales; B28 (only 28 cycles, operated at 43 °C) is reserved as an additional out-of-sample generalization target evaluated only on NASA-L.

We additionally conducted evaluations using the CALCE Battery Research Group CS2 dataset [17,37] to assess cross-dataset generalizability. The CALCE CS2 series consists of $LiCoO_2$ 18,650 cells (CS2-35/36/37/38) cycled under 1C constant-current protocol at room temperature (∼25 °C). The raw cycling data were collected with Arbin battery test equipment (Arbin Instruments, College Station, TX, USA). We convert the raw Arbin cycler data to match our pipeline exactly: per-cycle discharge .npy files with columns (current, voltage, temperature, sample_time, capacity_t), a metadata CSV, and a 10-dimensional impedance proxy vector derived from the measured DC internal resistance at 5 evenly spaced points

per cycle (filling the real part; imaginary part set to zero as no AC EIS is available). We use CS2-35/36/37 for training (1656 cycles) and CS2-38 for testing (593 cycles), mirroring our NASA train/test protocol.

Table 2. Dataset splits. B6 and B7 are held-out test batteries across all scales. Train/val split is 80/20 by cycle order within each training battery.

Scale	Train Batteries	Test Batteries	Train Cycles	Val Cycles
NASA-S	B5, B25, B29, B48	B6, B7	∼480	∼120
NASA-M	B5, B18, B45, B46, B48	B6, B7	∼600	∼150
NASA-L	B5, B18, B31, B34, B36 · · ·	B6, B7 (+B28 *)	∼1320	∼330

* B28 used for out-of-sample generalization evaluation only (28 cycles).

4.1.2. Data Processing

Each discharge cycle yields a variable-length sequence of four-channel measurements. We apply anchor-based resampling [22] to normalize all sequences to $T_{\text{dis}} = 128$ tokens, selecting anchor points at voltage plateaus and temperature extrema to preserve physically meaningful features. EIS measurements, acquired every 5–10 cycles, are aligned to each discharge cycle using the most recent preceding measurement, enforcing strict causal ordering with no future leakage. The 10-dimensional impedance feature vector consists of alternating real and imaginary parts at five logarithmically spaced frequencies, reshaped into five frequency–pair tokens for Q-former input. All channels are normalized using training set statistics only. Figure 3 illustrates the co-evolution of impedance features and SOH, and Figure 4 shows representative discharge voltage profiles.

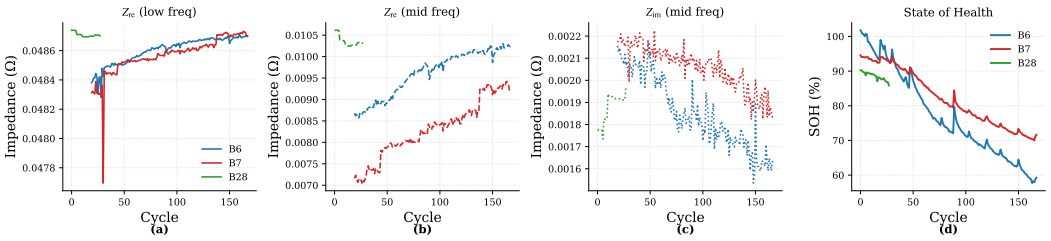

Figure 3. Co-evolution of impedance features and SOH for batteries B6, B7, and B28. (**a**) Low-frequency Z_{re} versus cycle number. (**b**) Mid-frequency Z_{re} versus cycle number. (**c**) Mid-frequency Z_{im} versus cycle number. (**d**) Corresponding SOH curves. The systematic correlation between impedance drift and capacity fade confirms that impedance carries complementary degradation information not captured by discharge curves alone.

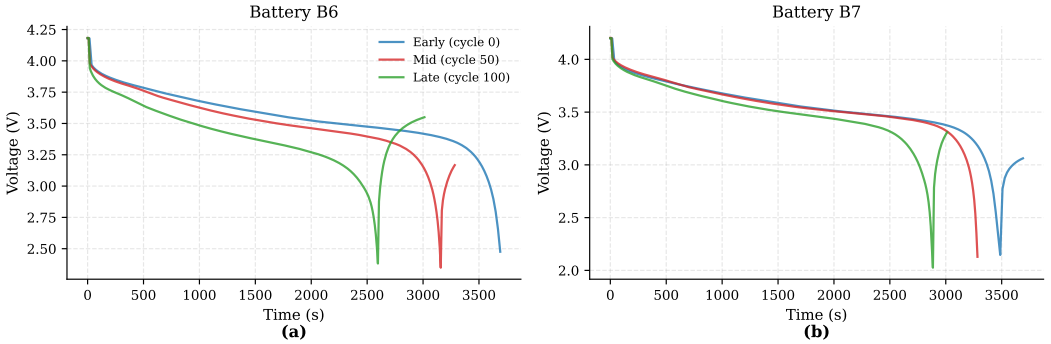

Figure 4. Discharge voltage profiles for test batteries. (**a**) B6 discharge profiles at early, mid-, and late-life stages. (**b**) B7 discharge profiles at early, mid-, and late-life stages. Curves are color-coded by cycle number and illustrate progressive capacity fade and voltage plateau shifts.

4.1.3. Evaluation Metrics

We report Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). MAE serves as the primary comparison metric; RMSE penalizes large deviations more heavily, capturing robustness to outlier cycles; MAPE provides scale-independent relative error.

4.1.4. Baselines

We compare against four categories: (i) recurrent models—LSTM [14] (two-layer bidirectional, 256 hidden units); (ii) hybrid architectures—CNN-BiGRU (1D CNN encoder followed by bidirectional GRU); (iii) attention-based models—Transformer [16] (four layers, eight heads); and (iv) state space models—Samba unimodal [22] (Mamba backbone on discharge curves only). We additionally evaluate two naive multimodal baselines: MultiModal-Concat (direct concatenation of impedance features with the discharge sequence) and MultiModal-Add (element-wise addition of mean-pooled impedance features to discharge tokens). All baselines use matched hyperparameter search budgets.

4.1.5. Implementation Details

All models are trained on a single NVIDIA RTX 4090 GPU (NVIDIA Corporation, Santa Clara, CA, USA) using Python 3.10 (Python Software Foundation, Wilmington, DE, USA) and PyTorch 2.1 (Meta Platforms, Inc., Menlo Park, CA, USA). We use AdamW with learning rate 10^{-4}, weight decay 0.05, and cosine annealing ($T_{\max} = 100$, $\eta_{\min} = 10^{-6}$). Batch size is 32; training runs for 100 epochs with early stopping (patience 20). Our model uses $d_{\text{model}} = 512$, eight Mamba layers [21], three Q-former layers [25] with 16 query tokens, and dropout/drop-path rates of 0.25. The global random seed is fixed at 42.

4.2. Main Results

Table 3 presents the primary comparison across all dataset scales. Our method outperforms every baseline on all three metrics and all three scales, with the magnitude of improvement varying systematically with data availability.

Table 3. Main results on the NASA battery dataset. Δ denotes relative MAE improvement over the unimodal Samba baseline at the same scale.

Model	Scale	MAE	RMSE	MAPE	Δ
Samba (unimodal)	NASA-S	2.964	3.291	3.789%	—
Ours	**NASA-S**	**2.705**	**3.171**	**3.492%**	**+8.7%**
LSTM	NASA-M	9.535	11.418	11.910%	—
CNN-BiGRU	NASA-M	6.607	7.862	8.184%	—
Transformer	NASA-M	11.149	13.754	13.313%	—
Samba (unimodal)	NASA-M	2.024	2.603	2.499%	—
MultiModal-Concat	NASA-M	1.543	2.711	1.873%	+23.8%
Ours	**NASA-M**	**1.457**	**2.241**	**1.795%**	**+28.0%**
Samba (unimodal)	NASA-L	0.983	1.381	1.178%	—
MultiModal-Concat	NASA-L	1.155	1.604	1.386%	−17.5%
Ours	**NASA-L**	**0.887**	**1.352**	**1.142%**	**+9.8%**

Best results in **bold**.

On NASA-M, the gap between traditional sequence models and Mamba-based approaches is striking: LSTM, CNN-BiGRU, and Transformer achieve MAEs of 9.535, 6.607, and 11.149 respectively, while the unimodal Samba baseline reaches 2.024—a 3–5× reduction. This confirms that the linear complexity selective state space mechanism captures

long-range discharge dynamics far more effectively than recurrent or attention-based alternatives at this sequence length ($T = 128$).

Adding impedance via our Q-former alignment yields further gains at every scale. On NASA-S (four batteries, ∼480 training cycles), both modalities are severely data-constrained: the Q-former lacks training diversity to learn stable cross-modal correspondences, and the impedance encoder tends to overfit battery-specific impedance trajectories rather than generalizing degradation mechanisms—yet the alignment mechanism remains conservative enough to avoid introducing harmful noise (+8.7%). NASA-M (five batteries, ∼600 training cycles) represents a sweet spot: training diversity is sufficient for the Q-former to learn generalizable impedance-to-discharge correspondences, yet discharge data alone still cannot resolve fine-grained degradation states, particularly in the mid-life regime where capacity fade decelerates and discharge curves become nearly indistinguishable across batteries with different underlying degradation mechanisms. Impedance features, which are directly sensitive to charge transfer resistance and SEI layer growth in this regime, provide the discriminative signal the discharge curves lack. On NASA-L (13+ batteries, ∼1320 training cycles), the unimodal baseline approaches its performance ceiling (MAE $< 1\%$), leaving less complementary information for impedance to contribute (+9.8%). This data-sufficiency interpretation—multimodal gains peak in the data-sufficient-but-not-saturated regime—is likely a generalizable principle beyond the specific numbers reported here.

A revealing contrast emerges from the MultiModal-Concat baseline. On NASA-M, naive concatenation achieves a respectable +23.8% improvement, suggesting that even unaligned impedance features carry useful signal when data is moderately sized. However, on NASA-L, the same strategy degrades performance by 17.5%. This reversal indicates that as training data grows and the backbone learns richer discharge representations, unaligned impedance tokens act as distributional noise that disrupts temporal modeling. Our Q-former alignment avoids this failure mode by projecting impedance features into the backbone's representation space before fusion.

Cross-Dataset Generalization on CALCE

Table 4 reports results on the CALCE CS2 dataset, which differs from NASA in three key aspects: constant-current cycling (no variable rest or multi-rate discharge), room temperature operation (no thermal variation), and the use of DC internal resistance as an impedance proxy rather than AC EIS. Our multimodal method achieves a 3.4% MAE improvement and a 20.6% RMSE reduction over the discharge-only Samba baseline, indicating that the Q-former alignment can leverage DC internal resistance features—which monotonically reflect SEI growth and capacity fade—as effectively as AC EIS, provided the features encode physically meaningful electrochemical degradation state.

Table 4. Results on the CALCE CS2 dataset (cross-dataset generalization). Test battery CS2-38 (593 cycles), trained on CS2-35/36/37 (1656 cycles). DC internal resistance is used as an impedance proxy (no AC EIS available). Δ_{RMSE} is relative RMSE improvement over unimodal Samba.

Model	MAE (%)	RMSE (%)	MAPE (%)	Δ_{RMSE}
Samba (unimodal, discharge only)	0.290	0.447	0.326	—
Ours	**0.280**	**0.355**	**0.311**	**−20.6%**

The absolute error values on CALCE are lower than those on NASA. This is expected and consistent with the controlled experimental conditions: strict 1C CC cycling at fixed temperature produces highly regular discharge curves whose shape is almost entirely determined by degradation state, with minimal confounding variability. NASA batteries,

by contrast, experience variable rest durations, intermittent EIS pulses, and non-monotonic capacity recovery (particularly at elevated temperature), all of which constitute irreducible noise for the regression model. The relative multimodal improvement is the comparable quantity across datasets, and it is consistent: Q-former alignment reduces the largest prediction errors (measured by RMSE) in both settings. Figure 5 compares the main model families on NASA-M and NASA-L, and Figure 6 summarizes the data-scaling trend across the three NASA settings.

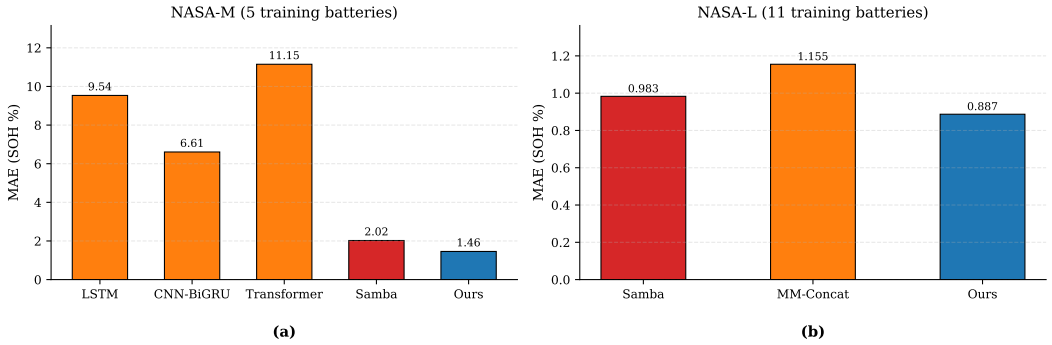

Figure 5. MAE comparison across methods. (**a**) Results on NASA-M. (**b**) Results on NASA-L. Mamba-based approaches achieve 3–5× lower error than traditional sequence models, and our multimodal method further reduces error at both scales.

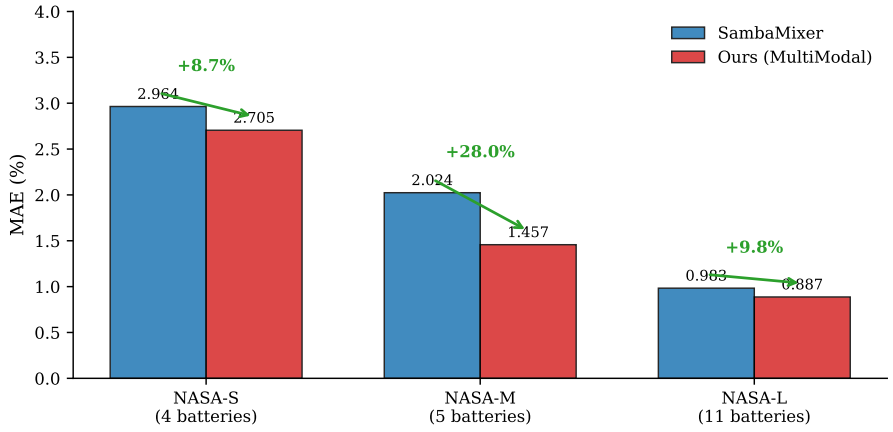

Figure 6. Data scaling behavior. Our method consistently outperforms the unimodal Samba baseline across all dataset sizes, with the largest relative gain at NASA-M (+28.0%) and positive improvements at both NASA-S (+8.7%) and NASA-L (+9.8%).

4.3. Why the Method Works: Ablation Analysis

We isolate the contribution of each design choice through two sets of ablations on NASA-L, where the larger test set provides the most reliable signal.

4.3.1. Fusion Strategy Matters More than Fusion Itself

Table 5 compares five fusion variants, all sharing the same Mamba backbone. The results reveal a clear hierarchy. Naive concatenation without Q-former alignment degrades MAE by 20.5% relative to the unimodal baseline, confirming that directly appending heterogeneous features disrupts the backbone's temporal representations. Adding Q-former alignment but collapsing queries via mean pooling (Concat + Mean Pooling) still underperforms by 17.5%, because compressing all impedance information into a single vector discards the multi-aspect structure of EIS data. Element-wise addition (Add Fusion) achieves a +6.9% improvement, demonstrating that even a simple fusion operator can be effective when the impedance encoder is properly trained—but it remains limited by

the single-vector bottleneck. Prepend fusion with 32 queries and two Q-former layers performs near baseline (−0.2%), suggesting that too many shallow queries introduce redundancy. Our final configuration—16 queries with three Q-former layers—achieves the best result (+9.8%), indicating that fewer, more deeply refined queries produce higher-quality cross-modal alignment.

Table 5. Ablation on fusion strategies (NASA-L). All variants share the same Mamba backbone. Δ is relative to the unimodal Samba baseline.

Fusion Strategy	MAE	RMSE	Δ
Samba (unimodal, no impedance)	0.983	1.381	—
No Q-former (naive concat)	1.185	1.540	−20.5%
Concat + Mean Pooling	1.155	1.604	−17.5%
Add Fusion	0.915	1.298	+6.9%
Prepend (32 queries, 2 layers)	0.985	1.444	−0.2%
Ours (16 queries, 3 layers)	**0.887**	**1.352**	**+9.8%**

4.3.2. Q-Former Depth and Query Count

Table 6 further dissects the prepend fusion design space. Across query counts at fixed depth (two layers), 16 queries strike the best balance: eight queries limit representational capacity (MAE 0.952), while 32 queries dilute the alignment signal through redundant queries (MAE 0.985). Across depths at fixed query count (16), performance improves monotonically up to three layers (MAE 0.887), then slightly regresses at four layers (MAE 0.901), likely because the limited dimensionality of the 10-feature impedance input does not warrant deeper cross-attention. The optimal configuration (16 queries, three layers) adds only 6.5 M parameters over the 19.7 M unimodal baseline—a 33% increase that yields 9.8% MAE reduction.

Table 6. Q-former architecture ablation (NASA-L, prepend fusion). The optimal configuration balances representational capacity against the limited dimensionality of impedance features.

Queries	Layers	MAE	Params (M)
8	2	0.952	24.1
16	2	0.921	25.3
32	2	0.985	27.8
16	1	0.935	24.8
16	2	0.921	25.3
16	3	0.887	26.2
16	4	0.901	27.1

4.4. Robustness and Generalization

Aggregate metrics can mask heterogeneous behavior across batteries. We therefore examine per-battery performance, prediction error distributions, and learned representations to assess whether the gains are robust.

4.4.1. Per-Battery Breakdown

Table 7 decomposes the NASA-L results by test battery. On B6, which exhibits smooth monotonic degradation, our method improves MAE by 5.0% (0.920 vs. 0.968). On B7, which follows a similar regular pattern, the improvement is substantially larger at 39.3% (0.682 vs. 1.124). Figure 7a,b shows that both models track B6 closely, but on B7, the unimodal baseline accumulates systematic bias in late-life cycles that our method corrects by conditioning on impedance drift.

The most striking result concerns B28, an out-of-sample battery with only 28 discharge cycles operated at elevated temperature (43 °C). B28 was never seen during training and differs from the training distribution in both cycle count and operating conditions. The unimodal Samba baseline produces an MAE of 6.967—nearly 7× its performance on B6/B7. Our method reduces this to 2.417, a 65.3% improvement (Figure 7c). This demonstrates that impedance features provide a direct window into electrochemical state that compensates for the absence of long-term discharge history, enabling reliable SOH estimation even in severely data-constrained regimes.

Table 7. Per-battery MAE on NASA-L. B6 and B7 are standard test batteries (∼168 cycles each); B28 is an out-of-sample battery (28 cycles, 43 °C) not seen during training.

Model	B6	B7	Avg (B6 + B7)	B28
Samba (unimodal)	0.968	1.124	1.046	6.967
Ours	0.920	0.682	0.801	2.417

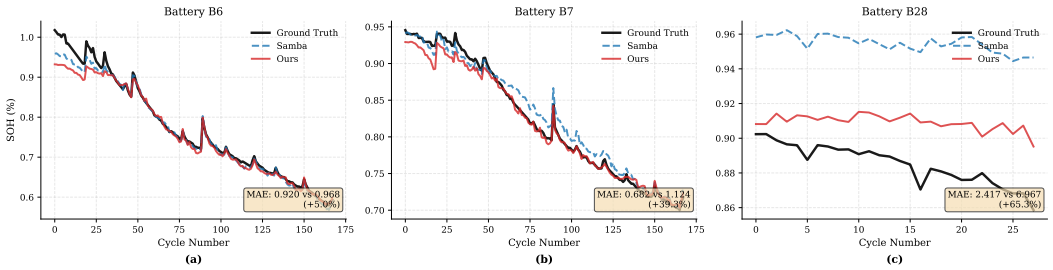

Figure 7. SOH prediction curves for individual test batteries on NASA-L. (**a**) B6: smooth degradation with close tracking by both models. (**b**) B7: our method corrects the late-life bias of the unimodal baseline (39.3% MAE improvement). (**c**) B28: out-of-sample battery (28 cycles); our method achieves 65.3% improvement by leveraging impedance to compensate for limited discharge history. Ground truth (black), Samba (blue dashed), ours (red solid).

4.4.2. Error Distribution

Figure 8 compares prediction error distributions on NASA-L. Our method produces a tighter, more symmetric distribution centered closer to zero, with reduced variance ($\sigma = 1.35$ vs. 1.38 for Samba). The heavier positive tail in the Samba distribution corresponds to systematic under-prediction in late-life cycles—precisely the regime where impedance drift provides the strongest corrective signal.

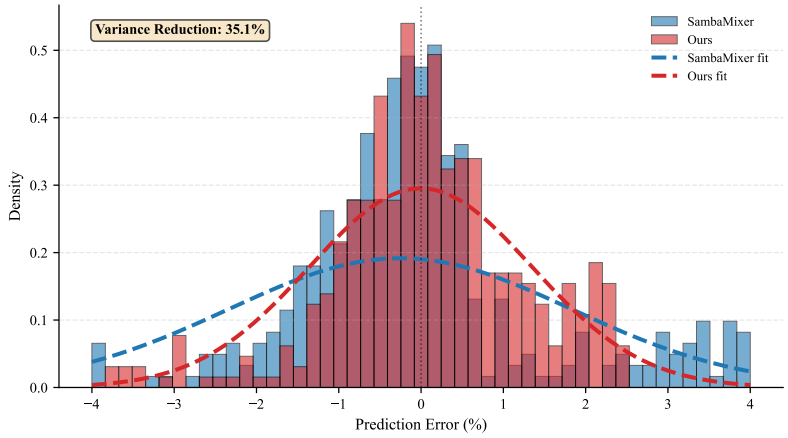

Figure 8. Prediction error distributions on NASA-L. Our method (red) produces a tighter distribution centered closer to zero compared to Samba (blue). Gaussian fits (dashed) confirm reduced variance and bias.

4.4.3. Learned Representations

Figure 9 visualizes fused representations via t-SNE for B6 and B7. The learned space exhibits a smooth degradation trajectory: healthy cycles (high SOH) cluster at one end, degraded cycles (low SOH) at the other, with the gray trajectory connecting cycles in SOH order confirming monotonic structure. Both batteries follow similar spatial patterns despite being independent test units, indicating that the Q-former alignment learns a generalizable degradation manifold rather than battery-specific shortcuts. Analysis of Q-former attention weights reveals query specialization: different queries attend preferentially to charge transfer resistance (R_{ct}), bulk real impedance ($\overline{Z_{\text{re}}}$), or imaginary impedance features, validating the use of multiple queries over a single aggregated representation. Figure 9 provides the corresponding low-dimensional visualization of the learned fused representations.

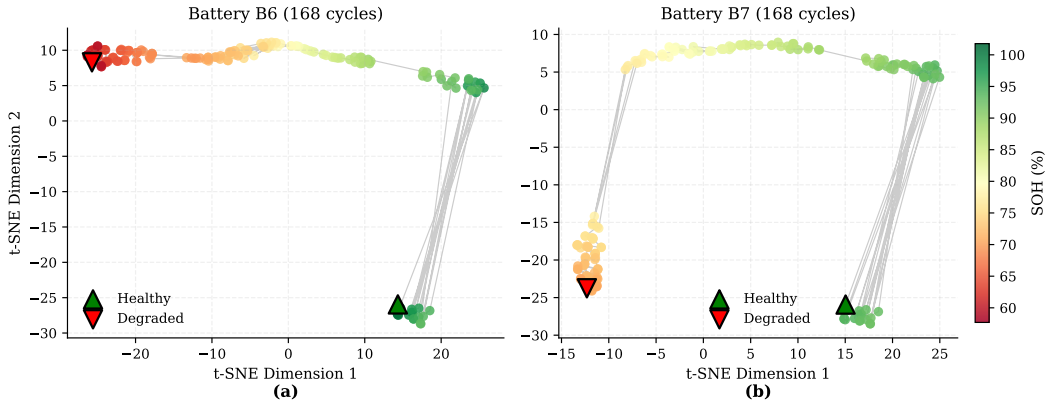

Figure 9. t-SNE visualization of fused representations. (**a**) B6 representations colored by SOH (%). (**b**) B7 representations colored by SOH (%). The gray trajectory connects cycles in decreasing SOH order. Triangle markers indicate healthy start (△, green) and degraded end (▽, red). The consistent spatial structure across both batteries demonstrates a generalizable degradation manifold.

4.5. Efficiency and Limitations

Table 8 compares computational costs. Our model adds 6.5 M parameters and 0.9 G FLOPs over the unimodal Samba baseline, reducing throughput by 14% (156 vs. 182 samples/s). This overhead is modest compared to the accuracy gains, and the model remains 2.3× faster than the Transformer baseline while achieving 8× lower MAE. The Q-former and prepend fusion account for the bulk of the additional cost; the impedance MLP is negligible .

Table 8. Computational efficiency measured during **inference** (forward pass only, FP16, batch size 32, sequence length 128, single RTX 4090, averaged over 1000 batches after 100 warmup batches). ‡ LSTM uses a two-layer bidirectional architecture (256 hidden units/direction) with cuDNN acceleration; lower throughput than our method reflects the sequential dependency in bidirectional recurrence at this configuration.

Model	Params (M)	FLOPs (G)	Throughput (Samples/s, Inference)
LSTM ‡	12.3	2.1	145
Transformer	18.5	8.7	68
Samba (unimodal)	19.7	3.2	182
Ours	26.2	4.1	156

Several limitations warrant discussion. First, although our revised manuscript adds CALCE CS2 evaluation (a second $LiCoO_2$ chemistry dataset), generalization to substantially different chemistries such as $LiFePO_4$ and NMC, and to operational scenarios with partial charging, variable C-rates, and field temperature profiles, remains to be validated. We view

cross-chemistry transfer as the most important open challenge for practical deployment. Second, EIS acquisition requires dedicated hardware and rest periods, which may not be available in all deployment scenarios—though periodic impedance testing is increasingly integrated into modern battery management systems [24]. Third, the B28 result, while encouraging, is based on a single out-of-sample battery; broader cross-battery evaluation would strengthen the generalization claims. Finally, our method assumes that the most recent preceding EIS measurement is informative for the current cycle; in scenarios with very sparse EIS (e.g., once per 50 cycles), the staleness of impedance features could degrade performance. A systematic study of model robustness under injected measurement noise (e.g., additive Gaussian noise with $\sigma \in \{0.01, 0.05, 0.1\}$ on discharge channels) and under non-stationary operating conditions (e.g., rest period-induced recuperation effects) remains important future work; the B28 out-of-sample result (43 °C, 28 cycles) provides preliminary evidence of robustness to distributional shift, but is insufficient for a systematic characterization.

Future directions include domain adaptation for cross-chemistry transfer, uncertainty quantification for safety-critical applications, online adaptation to individual degradation trajectories, and physics-informed constraints to improve robustness under extreme operating conditions. A rigorous significance analysis using leave-one-battery-out cross-validation would further strengthen the statistical claims; we leave this for future work, noting that the consistent directional improvement across all three dataset scales and both test batteries provides qualitative evidence against random variation.

5. Conclusions

We presented a multimodal architecture for lithium-ion battery State of Health prediction that integrates Mamba-based [21] temporal modeling of discharge curves with cross-modal fusion of electrochemical impedance features [23]. Our key contributions are (1) a Q-former module [25] that bridges the semantic gap between high-dimensional temporal sequences and low-dimensional static impedance features through learnable query tokens, (2) a prepend fusion strategy that injects aligned impedance queries as prefix tokens to enable dynamic integration of internal mechanism signals with external response patterns, and (3) comprehensive experimental validation on the NASA battery dataset [26] demonstrating consistent improvements across all data scales: 8.7% on small-scale data, 28.0% on medium-scale data, and 9.8% on large-scale data compared to unimodal baselines [22].

Our ablation studies reveal several important insights. First, Q-former-based alignment is essential—naive concatenation degrades performance by 20.5%, demonstrating that heterogeneous modalities cannot be directly fused without proper semantic alignment. Second, prepend fusion significantly outperforms mean pooling and concatenation-based approaches by preserving query diversity and enabling the Mamba backbone to dynamically attend to impedance information throughout temporal processing. Third, multimodal fusion provides consistent gains across all data scales, with the largest improvements in medium-data regimes where discharge curve data alone is insufficient, and meaningful gains even in small-data regimes where the alignment mechanism effectively leverages sparse impedance information. Cross-dataset evaluation on CALCE CS2 confirms that these findings generalize beyond the NASA dataset to a different cycling protocol.

5.1. Broader Impact

Accurate battery health prediction has significant implications for electric vehicle safety, grid-scale energy storage optimization, and sustainable battery life cycle management. By enabling early detection of degradation and more precise remaining useful life estimates, our method could help prevent battery failures, optimize maintenance schedules,

and facilitate second-life battery applications. However, over-reliance on predictive models without proper validation and uncertainty quantification could lead to safety risks if models fail silently under out-of-distribution conditions. We emphasize the importance of human oversight and conservative safety margins when deploying such systems in practice.

5.2. Future Work

Several promising directions remain for future research. First, extending our framework to handle missing modalities (e.g., when impedance measurements are unavailable) through modality dropout during training or learned imputation mechanisms. Second, incorporating physics-informed constraints such as capacity fade models or electrochemical degradation equations as soft constraints in the loss function to improve robustness under extreme conditions [27]. Third, developing domain adaptation techniques to enable cross-battery and cross-chemistry generalization without requiring extensive labeled data for each new battery type [36]. Fourth, integrating uncertainty quantification through Bayesian neural networks or deep ensembles to provide confidence bounds on predictions [29]. Finally, exploring online learning and continual adaptation to individual battery degradation trajectories [38], enabling personalized health monitoring that adapts to each battery's unique operating history.

In conclusion, our work demonstrates, using the NASA and CALCE datasets across multiple data scales, that structured multimodal fusion through cross-modal alignment and prepend fusion can consistently improve battery health prediction by bridging macroscopic discharge dynamics with microscopic impedance mechanisms. These results are a systematic step toward, rather than a definitive solution for, the broader challenge of multimodal battery health estimation, and further validation on additional chemistries and operational scenarios remains essential. We hope this framework inspires future research on multimodal learning for energy systems and other domains where heterogeneous sensor data must be integrated to capture complex physical processes.

Author Contributions: Conceptualization, Y.M. and M.Z.; methodology, Y.M. and Q.S.; software, Y.M. and A.B.; validation, Q.S., Z.Y. and Q.Y.; formal analysis, Y.M. and J.X.; investigation, A.B. and Z.Y.; resources, J.X., Z.W. and M.Z.; data curation, Y.M. and Q.S.; writing—original draft preparation, Y.M.; writing—review and editing, Q.S., J.X., Q.Y. and Z.W.; visualization, Y.M. and A.B.; supervision, J.X., Z.W. and M.Z.; project administration, M.Z.; funding acquisition, Z.W. and M.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the "Key-Area Research and Development Program of Guangdong Province" under Grant No. 2024B1111080003; the Civil Aviation Safety Capacity Building Project of China (No. MHAQ2024035); and the General Program of Civil Aviation Flight University of China, Grant No. 25CAFUC03065.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Li, Q.; Song, R.; Wei, Y. A review of state-of-health estimation for lithium-ion battery packs. *J. Energy Storage* **2025**, *118*, 116078. [CrossRef]
2. Yang, S.; Zhang, C.; Jiang, J.; Zhang, W.; Zhang, L.; Wang, Y. Review on state-of-health of lithium-ion batteries: Characterizations, estimations and applications. *J. Clean. Prod.* **2021**, *314*, 128015. [CrossRef]
3. Chang, P.; Liu, Z.; Xi, M.; Guo, Y.; Wu, T.; Ding, J.; Liu, H.; Huang, Y. Frustrated lewis pairs regulated solid polymer electrolyte enables ultralong cycles of lithium metal batteries. *Adv. Powder Mater.* **2025**, *4*, 100263. [CrossRef]

4. Wang, X.; Li, Z.; Mao, Q.; Wu, S.; Cheng, Y.; Qin, Y.; Chen, Z.; Peng, Z.; Yao, X.; Wang, D. Electrolyte-independent and sustained inorganic-rich layer with functional anion aggregates for stable lithium metal electrode. *Adv. Powder Mater.* **2025**, *4*, 100261. [CrossRef]
5. Huang, H.; Liu, C.; Liu, Z.; Wu, Y.; Liu, Y.; Fan, J.; Zhang, G.; Xiong, P.; Zhu, J. Functional inorganic additives in composite solid-state electrolytes for flexible lithium metal batteries. *Adv. Powder Mater.* **2024**, *3*, 100141. [CrossRef]
6. Agubra, V.; Fergus, J. Lithium ion battery anode aging mechanisms. *Materials* **2013**, *6*, 1310–1325. [CrossRef] [PubMed]
7. O'Kane, S.E.; Ai, W.; Madabattula, G.; Alonso-Alvarez, D.; Timms, R.; Sulzer, V.; Edge, J.S.; Wu, B.; Offer, G.J.; Marinescu, M. Lithium-ion battery degradation: How to model it. *Phys. Chem. Chem. Phys.* **2022**, *24*, 7909–7922. [CrossRef]
8. Wang, S.; Jin, S.; Deng, D.; Fernandez, C. A critical review of online battery remaining useful lifetime prediction methods. *Front. Mech. Eng.* **2021**, *7*, 719718. [CrossRef]
9. Fu, S.; Fan, H.; Jin, Z.; Ji, F.; Tao, Y.; Dong, Y.; Chen, X.; Shao, M.; Yuan, S.; Wang, Y.; et al. Recent progress in state of health estimation for lithium-ion batteries: From laboratory to practical application. *Renew. Sustain. Energy Rev.* **2026**, *226*, 116323. [CrossRef]
10. Shen, M.-C.; Gao, Q. A review on battery management system from the modeling efforts to its multiapplication and integration. *Int. J. Energy Res.* **2019**, *43*, 5042–5075. [CrossRef]
11. Su, L.; Xu, Y.; Dong, Z. State-of-health estimation of lithium-ion batteries: A comprehensive literature review from cell to pack levels. *IET Energy Convers. Econ.* **2024**, *5*, 224–242. [CrossRef]
12. Wang, Y.; Guo, S.; Cui, Y.; Deng, L.; Zhao, L.; Li, J.; Wang, Z. A comprehensive review of machine learning-based state of health estimation for lithium-ion batteries: Data, features, algorithms, and future challenges. *Renew. Sustain. Energy Rev.* **2025**, *224*, 116125. [CrossRef]
13. Zhang, Y.; Xiong, R.; He, H.; Pecht, M.G. Remaining useful life prediction of lithium-ion batteries based on health indicators and deep learning. *IEEE Trans. Ind. Electron.* **2019**, *66*, 1585–1597. [CrossRef]
14. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
15. 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]
16. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
17. Chen, Z.; Zhao, H.; Zhang, Y.; Shen, S.; Shen, J.; Liu, Y. Transformer network for remaining useful life prediction of lithium-ion batteries. *IEEE Access* **2022**, *10*, 19621–19628. [CrossRef]
18. Dao, T.; Fu, D.Y.; Ermon, S.; Rudra, A.; Ré, C. FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness. In Proceedings of the Advances in Neural Information Processing Systems, New Orleans, LA, USA, 28 November–9 December 2022.
19. Katharopoulos, A.; Vyas, A.; Pappas, N.; Fleuret, F. Transformers are RNNs: Fast Autoregressive Transformers with Linear Attention. In Proceedings of the International Conference on Machine Learning, Virtual, 13–18 July 2020.
20. Gu, A.; Goel, K.; Ré, C. Efficiently Modeling Long Sequences with Structured State Spaces. In Proceedings of the International Conference on Learning Representations, Virtual, 25–29 April 2022.
21. Gu, A.; Dao, T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. In Proceedings of the International Conference on Learning Representations, Vienna, Austria, 7–11 May 2024.
22. Olalde-Verano, J.I.; Kirch, S.; Pérez-Molina, C. SambaMixer: State of Health Prediction of Li-Ion Batteries Using Mamba State Space Models. *IEEE Access* **2024**, *12*, 189586–189600. [CrossRef]
23. Pastor-Fernández, C.; Uddin, K.; Chouchelamane, G.H.; Widanage, W.D.; Marco, J. A comparison between electrochemical impedance spectroscopy and incremental capacity-differential voltage as Li-ion diagnostic techniques to identify and quantify the effects of degradation modes within battery management systems. *J. Power Sources* **2017**, *360*, 301–318. [CrossRef]
24. Waag, W.; Fleischer, C.; Sauer, D.U. Critical review of the methods for monitoring of lithium-ion batteries in electric and hybrid vehicles. *J. Power Sources* **2014**, *258*, 321–339. [CrossRef]
25. Li, J.; Li, D.; Savarese, S.; Hoi, S. BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models. In Proceedings of the International Conference on Machine Learning. PMLR, Honolulu, HI, USA, 23–29 July 2023; pp. 19730–19742.
26. Saha, B.; Goebel, K.; Poll, S.; Christophersen, J. Prognostics methods for battery health monitoring using a Bayesian framework. *IEEE Trans. Instrum. Meas.* **2007**, *58*, 291–296. [CrossRef]
27. Plett, G.L. Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 3. State and parameter estimation. *J. Power Sources* **2004**, *134*, 277–292. [CrossRef]
28. Li, Y.; Liu, K.; Foley, A.M.; Zülke, A.; Berecibar, M.; Nanini-Maury, E.; Van Mierlo, J.; Hoster, H.E. Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review. *Renew. Sustain. Energy Rev.* **2019**, *113*, 109254. [CrossRef]
29. Hu, X.; Xu, L.; Lin, X.; Pecht, M. Battery health prognosis for electric vehicles using sample entropy and sparse Bayesian predictive modeling. *IEEE Trans. Ind. Electron.* **2015**, *63*, 2645–2656. [CrossRef]

30. 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]
31. Ng, M.F.; Zhao, J.; Yan, Q.; Conduit, G.J.; Seh, Z.W. Predicting the state of charge and health of batteries using data-driven machine learning. *Nat. Mach. Intell.* **2020**, *2*, 161–170. [CrossRef]
32. Smith, J.T.; Warrington, A.; Linderman, S.W. Structured state space models for in-context reinforcement learning. In Proceedings of the Advances in Neural Information Processing Systems, New Orleans, LA, USA, 28 November–9 December 2022; Volume 35, pp. 13022–13034.
33. Dao, T.; Fu, D.Y.; Saab, K.K.; Thomas, A.W.; Rudra, A.; Ré, C. Hungry hungry hippos: Towards language modeling with state space models. *arXiv* **2022**, arXiv:2212.14052.
34. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning transferable visual models from natural language supervision. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 8748–8763.
35. Gretton, A.; Borgwardt, K.M.; Rasch, M.; Schölkopf, B.; Smola, A.J. A Kernel Two-Sample Test. *J. Mach. Learn. Res.* **2012**, *13*, 723–773.
36. 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]
37. Center for Advanced Life Cycle Engineering. CALCE Battery Research Group Dataset. CALCE Battery Research Group, University of Maryland. 2017. Available online: https://calce.umd.edu/battery-data (accessed on 13 April 2026).
38. He, H.; Xiong, R.; Zhang, X.; Sun, F.; Fan, J. State of charge estimation for electric vehicle batteries using unscented Kalman filtering. *Microelectron. Reliab.* **2013**, *53*, 840–847. [CrossRef]