RESEARCH ARTICLE OPEN ACCESS

Learning When Not to Use a Battery: Multihorizon Failure Intelligence

Tareq Anwar Shikdar | Hannu Laaksonen

Flexible Energy Resources/Electrical Engineering, School of Technology and Innovations, University of Vaasa, Vaasa, Finland

Correspondence: Hannu Laaksonen (hannu.laaksonen@uwasa.fi)

Received: 12 February 2026 | **Revised:** 18 March 2026 | **Accepted:** 7 April 2026

Academic Editor: Wangqiang Niu

Keywords: battery energy storage system (BESS) | calibrated failure probability | decision-oriented prognostics | energy storage degradation | flexibility dispatch | multihorizon hazard learning | operational reliability | power system flexibility markets | predictive maintenance | reliability-aware control

ABSTRACT

Battery energy storage systems (BESSs) are increasingly dispatched to provide flexibility services in renewable-dominated power systems. However, conventional battery analytics focuses on state-of-health (SOH) and remaining useful life (RUL), which describe long-term degradation rather than short-term operational safety. In practical grid operation, the relevant decision is whether a battery can reliably complete a specific service commitment within a predefined time horizon. This paper reformulates battery prognostics as an operational reliability estimation problem and proposes a decision-oriented framework based on multihorizon discrete-time hazard learning. The method predicts the probability of failure within a service window and applies probability calibration to obtain a trustworthy risk metric. The calibrated risk is integrated into a reliability-aware dispatch policy that regulates participation in flexibility according to an acceptable failure tolerance. The framework is evaluated on 37 lithium-ion batteries using leave-battery-out cross-validation to ensure cross-device generalization. The hazard model achieves strong discrimination performance, reaching an area under the ROC curve (AUC) of 0.944 before calibration for the 20-cycle horizon. After probability calibration, the operational probabilities used for decision-making achieve an AUC of 0.891 with a Brier score of 0.064, demonstrating reliable probabilistic risk estimation across service horizons. When incorporated into dispatch decisions, the reliability-aware policy reduces the operational failure rate from 10.3% to 2.95% and increases delivered energy from 43.4 to 47.0 kWh compared with non-reliability-aware operation. An early-life analysis further shows that operational risk becomes predictable only after degradation observability emerges, distinguishing reliability estimation from long-term lifetime prediction. The results demonstrate that battery management for power systems should transition from predicting lifetime to controlling operational risk. The proposed framework provides an interpretable reliability signal enabling safe participation of storage assets in flexibility markets and bridges the gap between battery diagnostics and operational decision-making.

1 | Introduction

Battery energy storage systems (BESSs) have become essential operational assets in power systems with high renewable penetration, providing peak shaving, renewable smoothing, and ancillary regulation services [1]. In such applications, operators must determine not only how much energy a battery can deliver but also whether it can reliably complete a scheduled service within a predefined duration. Consequently, operational planning requires a reliability-oriented decision variable rather than a purely diagnostic health indicator.

Most battery analytics focus on estimating state-of-health (SOH) and remaining useful life (RUL), which characterize long-term degradation behavior [2–4]. Data-driven approaches using deep learning and sequence models have significantly improved lifetime prediction accuracy [5, 6], while physics-informed models incorporate electrochemical knowledge to enhance robustness [7]. Transfer learning and cross-condition learning further improve generalization across batteries and operating regimes [8–10]. These advances enable improved maintenance planning and lifecycle estimation.

However, lifetime prediction does not directly provide operational safety. A battery with a long predicted RUL may still fail during a high-power dispatch, whereas a degraded battery may safely complete a short service. This mismatch arises because prognostic indicators estimate time-to-failure, while grid operation requires the probability of failure within a specified service window. Uncertainty-aware and probabilistic prognostics quantify prediction confidence [11–13], but they do not directly provide a decision variable aligned with operational dispatch horizons.

Reliability-oriented approaches based on survival and hazard modeling estimate failure probability rather than deterministic lifetime [14, 15]. In parallel, degradation-aware optimization incorporates aging constraints into dispatch planning [16]. Nevertheless, existing methods typically evaluate reliability at a single horizon or rely on heuristic thresholds, preventing direct translation into operational decision-making across multiple service durations.

Therefore, a gap exists between battery prognostics and grid operation: current models predict degradation, whereas power-system control requires a calibrated probability that a battery cannot safely complete a predefined service interval.

This paper addresses this gap by reformulating battery prognostics as an operational reliability estimation problem. Instead of predicting lifetime, the proposed framework estimates failure probability across multiple service horizons using discrete-time hazard learning and converts the calibrated risk into a dispatch constraint.

The main contributions are as follows:

- A decision-oriented paradigm transforming lifetime prediction into service-horizon reliability estimation.
- A multihorizon hazard learning framework producing calibrated operational failure probabilities.
- A reliability-aware dispatch policy enabling controllable trade-off between utilization and safety.
- Cross-battery validation demonstrating transferable probabilistic prediction and improved operational performance.

The remainder of the paper is organized as follows. Section 2 reviews the related work. Section 3 presents the methodology. Section 4 provides experimental validation. Section 5 discusses implications and limitations. Section 6 concludes the paper.

2 | Related Work

Battery prognostics research can be broadly categorized into four directions: health estimation, lifetime prediction, probabilistic reliability modeling, and degradation-aware operation. While each contributes to understanding degradation behavior, none directly provides an operational reliability decision aligned with grid service commitments.

2.1 | Data-Driven Battery Health Estimation

Battery management systems (BMSs) commonly assess degradation using SOH indicators derived from voltage, current, and temperature measurements. Machine learning and deep learning models have significantly improved SOH estimation accuracy under varying operating conditions [17–20]. Attention mechanisms and feature fusion methods enhance robustness across load profiles [21, 22], while transfer learning enables generalization across battery types and domains [23, 24].

Recent work in battery prognostics and health management has emphasized data-driven anomaly detection and failure prediction using system behavior monitoring and degradation trajectory analysis [25, 26]. In addition, recent reviews highlight the importance of combining experimental measurement, modeling, and data-driven techniques for accurate estimation of battery operational states such as state of charge and SOH [27].

Although these approaches provide accurate diagnostic indicators, the estimated health value does not directly determine whether a battery can safely execute a specific service within a predefined duration.

2.2 | Remaining Useful Life Prediction and Uncertainty Modeling

Remaining useful life prediction estimates the time until end-of-life (EOL) using regression-based machine learning, sequence models and hybrid degradation approaches [3, 6, 28, 29]. Early-cycle prediction methods infer lifetime before significant aging occurs [4, 30], and uncertainty-aware methods provide confidence bounds through probabilistic inference and ensemble learning [29–32].

Physics-based degradation models have also been widely used to study capacity fade mechanisms in lithium-ion batteries. For instance, electrochemical models incorporating degradation mechanisms such as SEI layer growth and side reactions have been developed to analyze long-term capacity loss and aging processes [33, 34].

These approaches describe when failure may occur, but operational decision-making requires determining whether failure may occur within the next service interval. Consequently, lifetime prediction does not directly provide actionable operational reliability.

2.3 | Reliability and Survival-Based Modeling

To address deterministic lifetime limitations, several studies model degradation using stochastic reliability and survival analysis [13, 35, 36]. These methods estimate hazard and survival probabilities, offering a probabilistic interpretation of failure behavior. Reliability modeling has also been widely applied in power system risk assessment [37, 38].

Recent studies in prognostics and health management further explore predictive frameworks that combine data-driven monitoring with system-level anomaly detection for battery reliability evaluation [39, 40].

However, existing reliability formulations typically evaluate failure probability at a single prediction horizon. Therefore, they cannot represent operational risk across multiple service durations required in flexibility markets.

2.4 | Degradation-Aware Operation of Energy Storage Systems

Recent research integrates battery aging into energy management and dispatch optimization [12, 37–41]. These approaches improve economic operation and protect the lifetime by imposing degradation constraints. Nevertheless, operational decisions are derived from heuristic thresholds or simplified aging proxies rather than a learned reliability probability. As a result, dispatch remains indirectly linked to degradation rather than derived from calibrated risk estimation.

2.5 | Existing Research Limitations

Existing research can, therefore, be categorized into four directions: health estimation, lifetime prediction, probabilistic reliability modeling, and degradation-aware operation. While each improves understanding of degradation behavior, none directly provides a decision variable answering the operational question:

Can the battery safely perform the next service within a predefined horizon?

The missing element is a calibrated probability of failure defined across multiple service horizons and directly usable as a dispatch constraint.

To highlight the distinction between existing approaches and the proposed framework, Table 1 summarizes the operational relevance of major research categories. The comparison emphasizes that prior methods primarily estimate degradation or lifetime, whereas the proposed approach directly produces an operational reliability variable usable for dispatch decision-making.

The comparison demonstrates that existing approaches either estimate degradation or incorporate reliability indirectly into optimization. In contrast, the proposed framework introduces a calibrated multihorizon failure probability that can directly serve as an operational decision variable.

3 | Methodology

The proposed framework integrates prediction and control in a unified pipeline. Instead of estimating degradation indicators for maintenance purposes, the model directly evaluates operational safety over future service horizons. The predicted failure probability is calibrated and converted into a dispatch constraint, enabling batteries to participate in flexibility markets only when reliability requirements are satisfied. The overall operational decision framework is illustrated in Figure 1.

3.1 | Operational Reliability Formulation

Conventional battery prognostics primarily focus on estimating degradation-oriented indicators such as SOH or RUL. While these metrics are suitable for maintenance scheduling, grid-level flexibility services require a different operational decision: determining whether the battery can reliably complete a forthcoming service commitment of predefined duration.

TABLE 1 | Comparison of battery prognostic approaches in terms of operational usability.

Approach Category	Primary output	Time scale	Probabilistic output	Operational decision capability	Representative references
SOH estimation	Health indicator	Long term	Limited	Limited	[17–24, 42]
RUL prediction	Remaining lifetime	Long term	Limited	Limited	[3, 4, 28–34]
Probabilistic RUL	Lifetime distribution	Long term	Available	Limited	[30–35]
Reliability/survival modeling	Failure probability	Single horizon	Available	Partial	[13, 35–37, 39–43]
Degradation-aware operation	Operational constraint	Operational	Not explicit	Indirect	[12, 40–45]
Proposed framework	Failure probability within service window	Multihorizon	Calibrated	Direct	This work

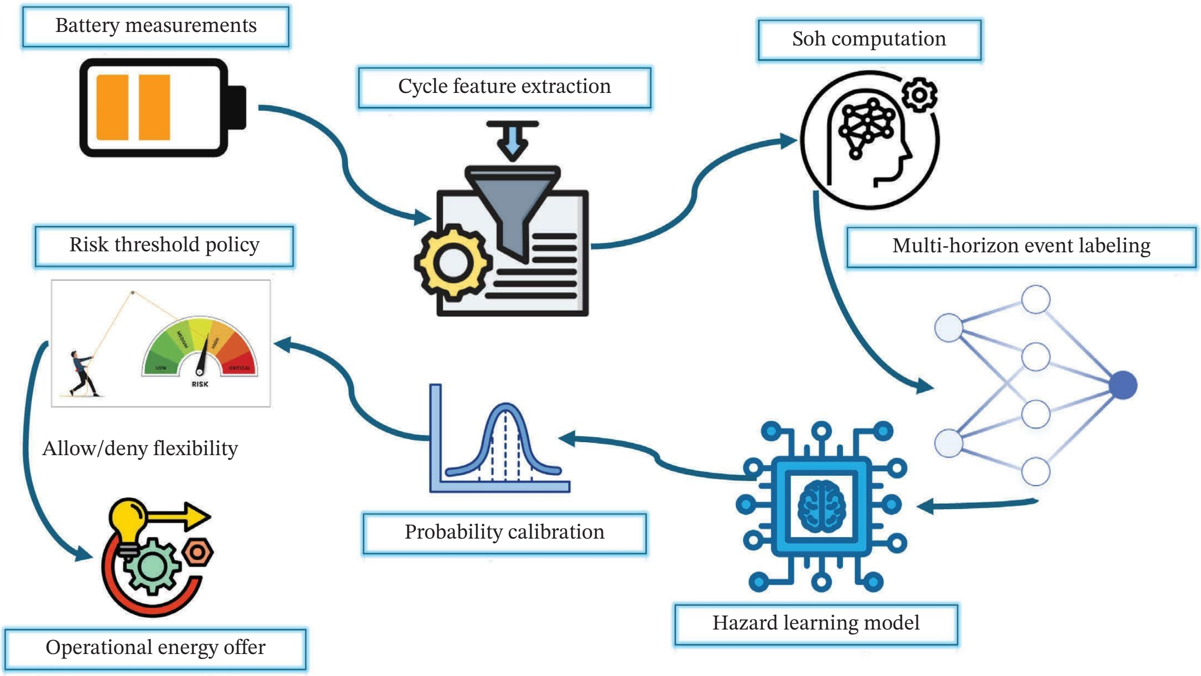


FIGURE 1 | Decision-oriented battery reliability learning framework.

In this work, the term *failure* refers to operational service failure, defined as the inability of a battery to successfully complete a requested service within a specified operational horizon. This definition differs from catastrophic safety failures (e.g., thermal runaway) and from permanent EOL conditions used in degradation studies. Instead, the objective is to evaluate whether the battery can reliably deliver the required energy for a short-term operational commitment.

To address this mismatch, the proposed framework reformulates battery diagnostics as a finite-horizon reliability prediction task rather than a lifetime regression problem.

For a battery observed at cycle t, a binary operational event is defined as follows:

$$y_{t,H} = \begin{cases} 1, & \text{if failure occurs within } (t,\ t+H], \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where H denotes the service horizon. Instead of predicting total lifetime, the objective is to estimate the conditional probability of failure within this horizon.

$$P_{\text{fail}}(t,H) = P\left(T_f \le t+H \,|\, T_f > t, \quad x_t\right), \tag{2}$$

where T_f is the failure cycle and x_t represents the measurements available at the cycle t. This formulation converts battery health assessment into a discrete-time survival analysis problem, directly aligned with operational decision-making. Consequently, the model produces a risk measure interpretable by grid operators, enabling dispatch decisions based on short-term reliability rather than long-term degradation indicators.

3.2 | Feature Representation

The proposed framework utilizes only measurements commonly available in standard BMSs, ensuring practical deployability and facilitating transferability across battery types without requiring specialized electrochemical sensing. For each cycle t, the feature vector is defined as

$$x_t = \left\{SOH_t,\ V_{\text{avg}},\ I_{\text{avg}},\ T_{\text{avg}},\ \text{duration},\ t\right\}. \tag{3}$$

Rather than relying on handcrafted degradation indicators, the method learns degradation dynamics directly from temporal evolution in the measurements. To this end, a rolling historical window is constructed as follows:

$$X_t = [x_{t-w},\ \ldots, x_t], \tag{4}$$

where w denotes the observation window length. This formulation enables the model to capture degradation trajectories and transition patterns over time, instead of inferring reliability from a single static health snapshot. Consequently, operational risk is derived from learned temporal behavior, improving generalization across heterogeneous batteries and operating conditions.

3.3 | Multihorizon Discrete-Time Hazard Learning

Because the operational risk of a battery depends on the duration of the requested service, reliability must be assessed across multiple commitment lengths rather than a single prediction horizon. The framework, therefore, performs multihorizon learning, simultaneously estimating risk for a set of service durations.

$$H \in \{10, 20, 30, 50\}. \tag{5}$$

For each horizon, a conditional discrete-time hazard function is defined as

$$h(t,H) = P\left(T_f = t+H \mid T_f > t, \quad X_t\right). \tag{6}$$

Representing the probability that failure occurs exactly at the future step $t+H$ Given survival up to the cycle t. The corresponding survival probability over the interval $(t, t+H)$ is

$$S(t, H) = \prod_{k=1}^{H} (1 - h(t, k)), \tag{7}$$

and the cumulative probability of failure within the service horizon becomes

$$P_{\text{fail}}(t, H) = 1 - S(t, H). \tag{8}$$

The predictive model learns the mapping

$$f_\theta \colon X_t \longrightarrow P_{\text{fail}}(t, H) \tag{9}$$

by jointly optimizing all horizons using a binary cross-entropy objective

$$\mathcal{L} = -\sum_{t,H} \left[y_{t,H} \log P_{\text{fail}}(t, H) + \left(1 - y_{t,H}\right) \log\left(1 - P_{\text{fail}}(t, H)\right) \right]. \tag{10}$$

Unlike conventional RUL regression, this objective directly minimizes operational risk estimation error, enabling the model to produce decision-relevant reliability probabilities tailored to grid service commitments.

3.4 | Probability Calibration

Raw machine learning probabilities are often miscalibrated and, therefore, cannot be directly interpreted as operational risk. To obtain reliable risk estimates, a monotonic calibration function is learned using isotonic regression as follows:

$$\widehat{P}_{cal} = g\left(\widehat{P}_{fail}\right), \tag{11}$$

where $g(\cdot)$ is a nonparametric monotonic mapping fitted on validation data. The calibration enforces probabilistic consistency as follows:

$$P\left(\text{failure} \mid \widehat{P}_{cal} = p\right) \approx p. \tag{12}$$

After calibration, the predicted probability represents a true empirical failure likelihood rather than a classification score. This property is essential because the prediction is used as a control variable in the reliability-aware dispatch policy.

3.5 | Reliability-Aware Dispatch Policy

To integrate reliability prediction into operational decision-making, the estimated failure probability is calibrated and compared against a predefined risk tolerance threshold τ. A service request is considered admissible only if the calibrated probability satisfies

$$\widehat{P}_{\text{cal}}(t, H) \le \tau. \tag{13}$$

Based on this condition, the battery dispatch is converted into a reliability-aware constraint. The allowable dispatch energy at time t is defined as

$$E_{\text{allowed}}(t) = \begin{cases} E_{\text{req}}(t), & \widehat{P}_{\text{cal}}(t, H) \le \tau, \\ 0, & \text{otherwise}, \end{cases} \tag{14}$$

where $E_{\text{req}}(t)$ denotes the requested service energy. This formulation transforms battery diagnostics from a passive monitoring indicator into an active control variable, enabling participation in flexibility markets only when operational reliability requirements are satisfied. Consequently, battery health estimation directly informs dispatch decisions, linking predictive analytics with real-time grid operation.

3.6 | Evaluation Protocol

The model is trained on multiple batteries and tested on an unseen battery using leave-battery-out cross-validation. Performance is assessed using discrimination (area under the ROC curve [AUC]), probabilistic accuracy (Brier score), calibration consistency, and operational energy-risk trade-off metrics. The validation methodology is illustrated in Figure 2. To evaluate cross-device generalization, a leave-battery-out cross-validation protocol is adopted. For a dataset containing $N_{\text{batteries}}$ cells, the model is iteratively trained on $N-1$ batteries and tested on the remaining unseen battery. This procedure prevents information leakage across degradation trajectories and assesses the ability of the learned representation to transfer to previously unobserved devices.

Model performance is quantified using complementary reliability-oriented metrics as follows:

- AUC: measures discrimination capability between safe and unsafe operation.
- Brier score: evaluates probabilistic accuracy of predicted failure risk.
- Calibration curve: assesses consistency between predicted probabilities and observed outcomes.
- Operational trade-off analysis: examines the relationship between delivered energy and realized failure rate.

Together, these metrics characterize not only predictive accuracy but also decision reliability, ensuring the model is suitable for safety-critical operational deployment.

3.7 | Dataset and Implementation Details

The proposed framework is evaluated using the NASA Prognostics Center of Excellence lithium-ion battery aging dataset [45]. The dataset contains multiple commercial 18,650 lithium-ion cells cycled under controlled charge–discharge conditions under different operating regimes. Each cycle includes measurements of voltage, current, temperature, and capacity degradation throughout the battery lifetime.

Following common practice in battery prognostics research, the dataset defines EOL when the available capacity drops below 70% of the nominal capacity. In this work, this EOL point is used only to determine the reference failure cycle T_f required for constructing the survival labels in equation (1). The proposed model does not directly predict EOL. Instead, the framework predicts the probability that the battery will fail to complete a requested service within a finite operational horizon.

Therefore, the prediction task focuses on operational reliability [46], while the EOL definition is used only as a reference point to construct failure events for supervised learning.

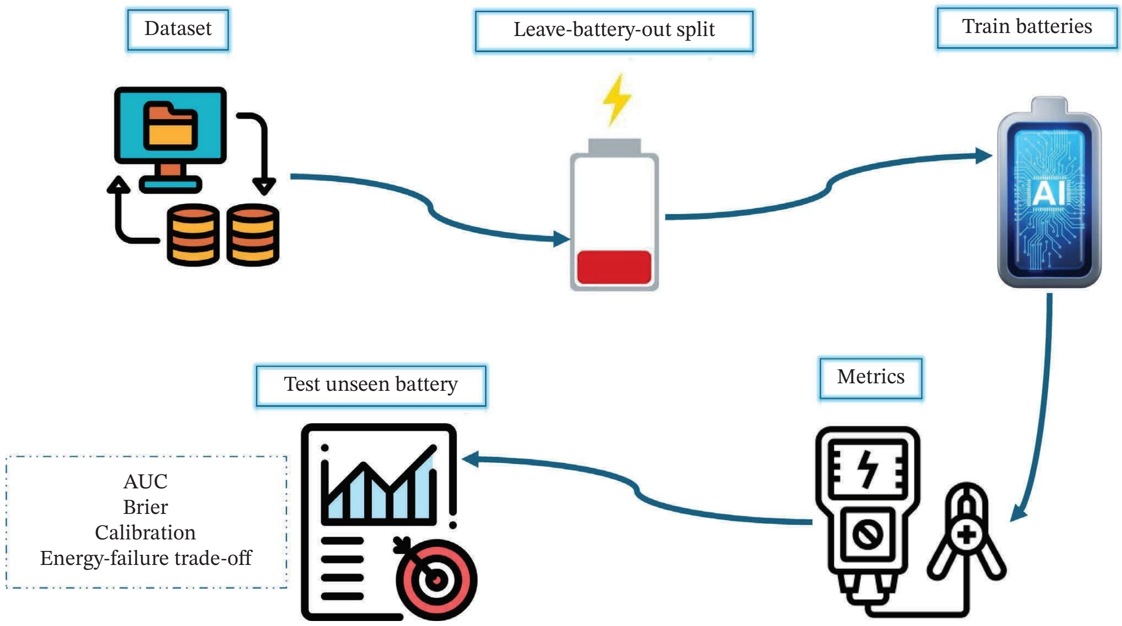


FIGURE 2 | Cross-battery generalization evaluation protocol.

The hazard learning model is implemented in Python using gradient boosted decision trees (XGBoost). Model parameters are optimized using the gradient boosting procedure with logistic loss and second-order gradient updates. Early stopping based on the validation Brier score is used to prevent overfitting.

Probability calibration is performed using isotonic regression fitted on validation predictions. All experiments are repeated across batteries, and reported results represent the mean performance across cross-validation folds. Training configurations are summarized in Table 2.

4 | Results

4.1 | Why Lifetime Prediction is Not an Operational Indicator

Conventional battery prognostics rely on SOH and RUL to quantify battery condition. However, operational grid services require determining whether a battery can safely complete an upcoming service window rather than estimating its total lifetime.

Figure 3(a) presents SOH trajectories for representative cells. Although all cells follow a global decreasing trend, their short-term degradation behavior differs substantially. Cells with similar SOH values exhibit different local degradation dynamics, including temporary recovery, accelerated decay, and irregular fluctuations. Consequently, SOH alone does not uniquely determine near-term reliability since batteries at comparable health levels may experience different short-term failure risks.

Figure 3(b) shows the corresponding RUL evolution. RUL decreases smoothly and monotonically even when degradation behavior becomes unstable in the SOH trajectories. This smoothing effect masks short-term degradation acceleration and, therefore, cannot represent operational safety within a limited service horizon.

These observations demonstrate that lifetime indicators describe long-term aging but do not capture short-term operational risk. Therefore, battery diagnostics are reformulated as a finite-horizon failure probability estimation problem, where the objective is to predict whether failure occurs within a predefined service interval rather than estimating total lifetime.

4.2 | Probability Calibration and Reliability of Predictions

For operational deployment, predicted risk values must correspond to actual failure frequencies rather than acting as relative confidence scores. Figure 4(a) presents the calibration relationship between predicted failure probability and observed event frequency. After calibration, predicted probabilities align with empirical occurrence rates, indicating that the model outputs statistically consistent risk estimates.

This property is essential because the prediction is used as a control variable in dispatch decisions. A miscalibrated classifier may rank unsafe conditions correctly but still produce unreliable probability magnitudes, which would lead to incorrect operational constraints.

Figure 4(b) further illustrates prediction uncertainty over the degradation trajectory. The predicted median trajectory follows the degradation trend while the quantile band captures increasing uncertainty near the EOL. The uncertainty interval expands as failure approaches, reflecting the stochastic nature of degradation dynamics rather than deterministic lifetime behavior.

Together, calibration consistency and uncertainty awareness demonstrate that the model produces interpretable operational probabilities instead of classification scores, enabling its use in reliability-aware control.

4.3 | Multihorizon Failure Prediction Performance

The proposed model estimates the probability that a battery fails within predefined service horizons rather than predicting

TABLE 2 | Training configuration of the multihorizon hazard learning model.

Component	Setting	Value
Learning framework	Model type	Gradient boosted trees
Implementation	Library	XGBoost (sklearn API)
Prediction target	Task	Binary hazard classification
Loss function	Objective	Logistic loss (binary cross-entropy)
Number of estimators	Trees	300
Maximum tree depth	Depth	4
Learning rate	η	0.05
Row sampling	Subsample	0.8
Feature sampling	Colsample	0.8
Regularization	Minimum child weight	5
Multihorizon setup	Horizons (H)	{10, 20, 30, 50 cycles}
Probability calibration	Method	Isotonic regression
Validation scheme	Cross-validation	Leave-battery-out (5 folds)
Evaluation metrics	Metrics	AUC, Brier score, ECE
Dispatch policy	Risk threshold (\tau)	0.05–0.50

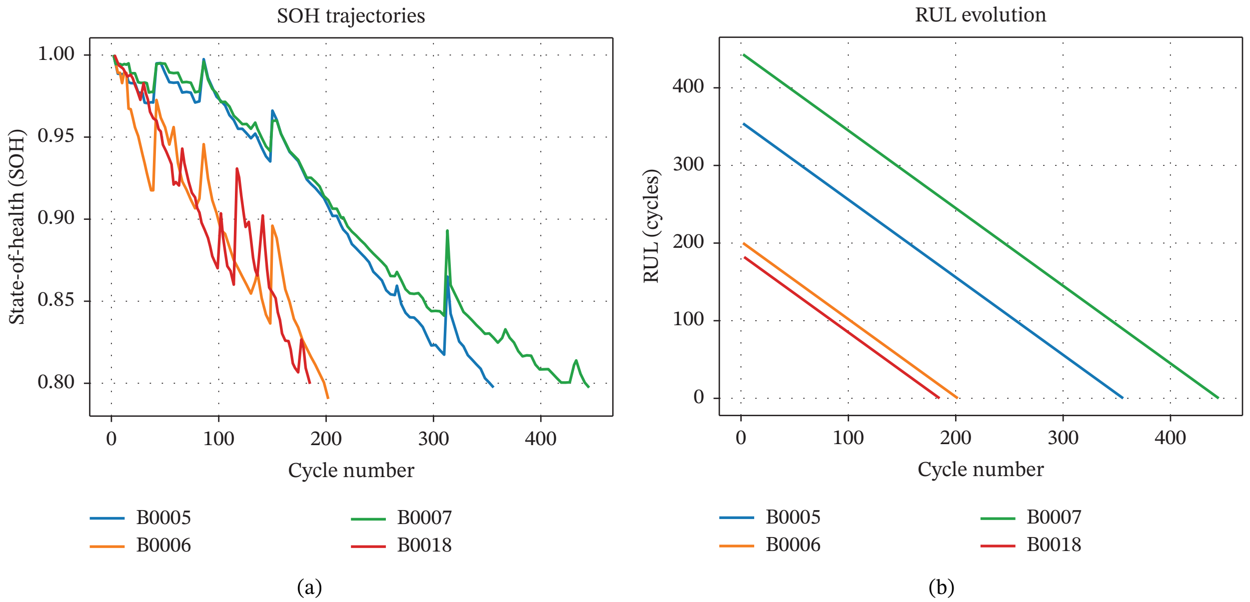


FIGURE 3 | SOH and RUL limitations for operational reliability. (a) SOH trajectories. (b) RUL evolution.

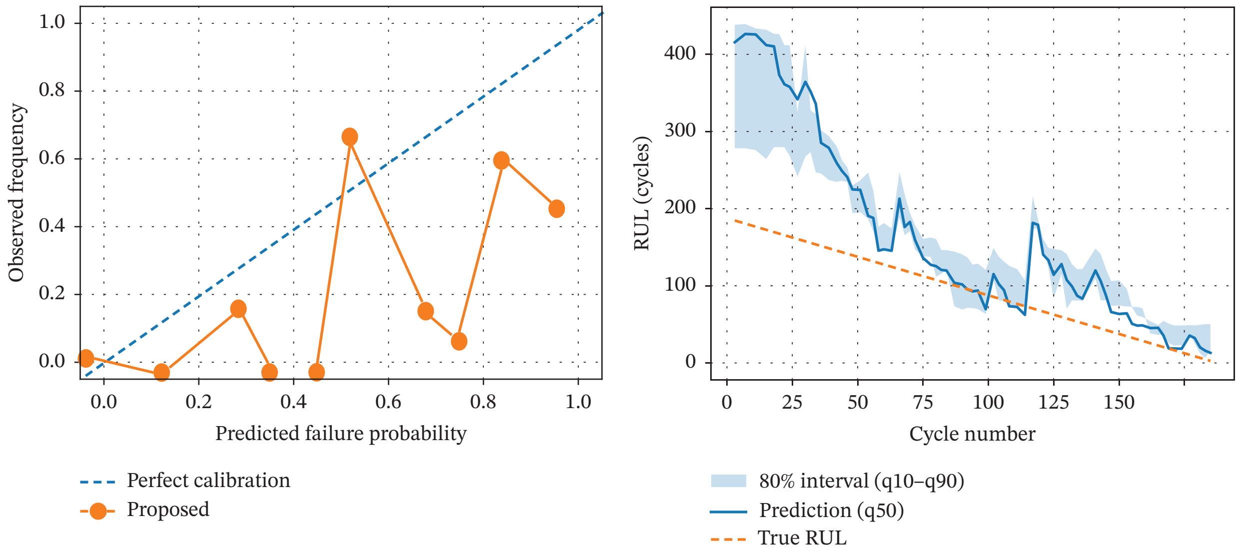
a single lifetime value. Figure 5(a) presents the survival probability for a representative cell at a 20-cycle horizon. The survival estimate remains close to unity during normal operation and drops sharply shortly before EOL, indicating that the model detects imminent operational risk instead of gradual aging.

Figure 5(b) shows predicted failure probabilities for multiple horizons. Short horizons detect risk earlier and more abruptly, while longer horizons increase more gradually and reach higher probability values near the EOL. This behavior is consistent with reliability theory: the closer the requested service duration approaches the remaining lifetime, the higher the operational risk.

Table 3 summarized the predictive performance of the multihorizon hazard model across service horizons. Raw AUC corresponds to the discrimination capability of the hazard model before probability calibration, while calibrated metrics correspond to the probabilities used in the operational dispatch policy.

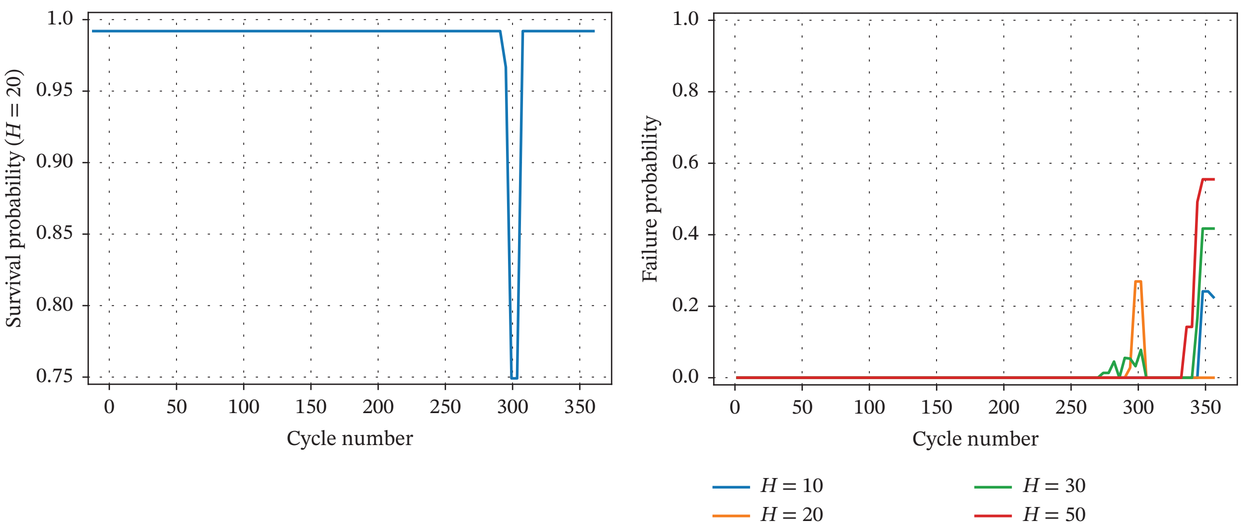
The predictive performance of the multihorizon hazard model is summarized in Table 3. It is important to note that two sets of performance metrics are reported: the discrimination performance of the raw model predictions and the performance of the calibrated probabilities used for operational decision-making.

The raw hazard model achieves an AUC of 0.944 for the 20-cycle horizon, indicating strong discrimination between safe and unsafe operational states. After probability calibration using isotonic regression, the AUC becomes 0.891 while the Brier score improves slightly. This change occurs because calibration modifies the probability distribution to ensure probabilistic consistency with observed outcomes.



(a) (b)

FIGURE 4 | Reliability calibration and predictive uncertainty. (a) Calibration relationship between predicted failure probability and observed event frequency. (b) Prediction uncertainty over the degradation trajectory.



(a) (b)

FIGURE 5 | Multihorizon operational risk prediction. (a) Survival probability for a representative cell. (b) Predicted failure probabilities for multiple horizons.

TABLE 3 | Predictive performance of the multihorizon hazard model.

Horizon (cycles)	AUC (raw)	AUC (calibrated)	Brier score (calibrated)
10	0.928	0.891	0.032
20	0.944	0.891	0.064
30	0.942	0.898	0.058
50	0.908	0.868	0.090

Since the calibrated probabilities are used in the reliability-aware dispatch policy, the calibrated metrics reported in Table 3 represent the operationally relevant performance of the proposed framework.

4.4 | Operational Value of Risk-Based Dispatch

To evaluate operational usefulness, the calibrated failure probability was converted into a dispatch constraint. Figure 6 illustrates the relationship between delivered energy and observed failure rate for different decision policies.

Conventional rule-based strategies exhibit clear limitations. The always-dispatch policy results in the highest failure rate (10.3%), indicating that operating the battery without reliability constraints exposes the system to significant operational risk. The SOH-based rule slightly reduces failure occurrence but still operates in a relatively unsafe region since a single health threshold cannot capture short-term degradation dynamics.

The RUL threshold policy eliminates failures in the evaluated dataset, achieving a failure rate of 0% while delivering 47.53 kWh of energy. However, this policy relies on a deterministic lifetime threshold and, therefore, does not account for uncertainty in degradation trajectories or variations across operational horizons. In practice, such fixed thresholds may be difficult to tune across heterogeneous batteries and operating conditions.

In contrast, the proposed risk-threshold policy uses calibrated failure probabilities to regulate dispatch decisions. The resulting policy achieves a failure rate of 2.9% while maintaining 46.97 kWh of delivered energy. Although the deterministic RUL rule eliminates failures in this particular dataset, the probabilistic framework provides a more flexible reliability signal that can adapt to different service horizons and reliability requirements.

These results demonstrate that probabilistic reliability estimation can be translated into an operational control variable, enabling grid operators to balance reliability and utilization while maintaining consistent participation in flexibility services.

4.5 | Reliability-Aware Dispatch Behavior

The proposed framework enables operators to directly control participation through a reliability tolerance parameter.

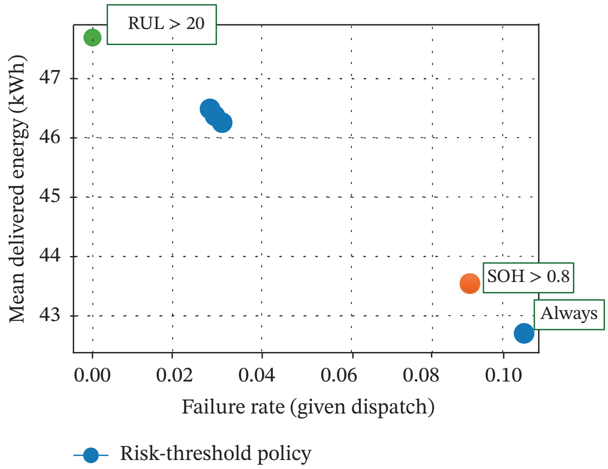


FIGURE 6 | Operational energy–risk trade-off across dispatch policies.

Figure 7(a) shows the mean offered energy as a function of the risk threshold τ. Lower thresholds enforce conservative operation, while increasing τ gradually increases energy contribution. The relationship is continuous rather than binary, allowing adjustable reliability levels instead of fixed acceptance rules.

Figure 7(b) illustrates dispatch behavior over the degradation trajectory. The reliability-aware index follows the available energy during early operation and progressively limits participation as degradation advances. Unlike threshold rules that abruptly disable the battery, the proposed policy produces a smooth reduction in utilization near EOL.

This behavior demonstrates that predicted failure probability functions as a controllable operational variable. Grid operators can tune reliability tolerance while maintaining stable participation in flexibility services.

4.6 | Generalization to Early-Life Batteries

To evaluate predictive generalization, the model was trained using only early-life data and tested across the full lifetime of the battery. Figure 8 compares discrimination performance under limited degradation exposure and full trajectory availability. Operational benchmark comparisons are summarized in Table 4.

When trained on early-life data, performance decreases but remains above random prediction, indicating the model captures meaningful degradation behavior even before significant aging occurs. Using full lifetime data substantially improves discrimination capability, confirming that additional degradation observations refine the learned reliability representation.

These results indicate that the model does not rely solely on late-stage degradation patterns but instead learns transferable indicators of operational risk. Consequently, the framework can support reliability assessment even for relatively new batteries, which is essential for practical deployment where long historical records are not yet available.

5 | Ablation Analysis of the Operational Reliability Framework

To better understand the contribution of individual components in the proposed operational reliability framework, an ablation analysis was conducted. The framework consists of three key elements: (i) multihorizon hazard learning, (ii) probability calibration, and (iii) a risk-aware operational decision policy. While the previous sections evaluated the complete framework, the ablation analysis isolates the effect of these components by comparing alternative decision strategies.

The quantitative comparison is summarized in Table 5. The table includes the baseline always-dispatch policy, the raw hazard-based policy, the calibrated hazard-based policy used in the proposed framework, and the deterministic RUL threshold benchmark. Since the purpose of the ablation figure is to show the incremental contribution of the proposed model components, Figure 9 focuses only on the progression from the baseline policy to hazard learning and then to hazard learning with calibration. The RUL threshold is retained in Table 5 as an external benchmark rather than as a component of the proposed framework.

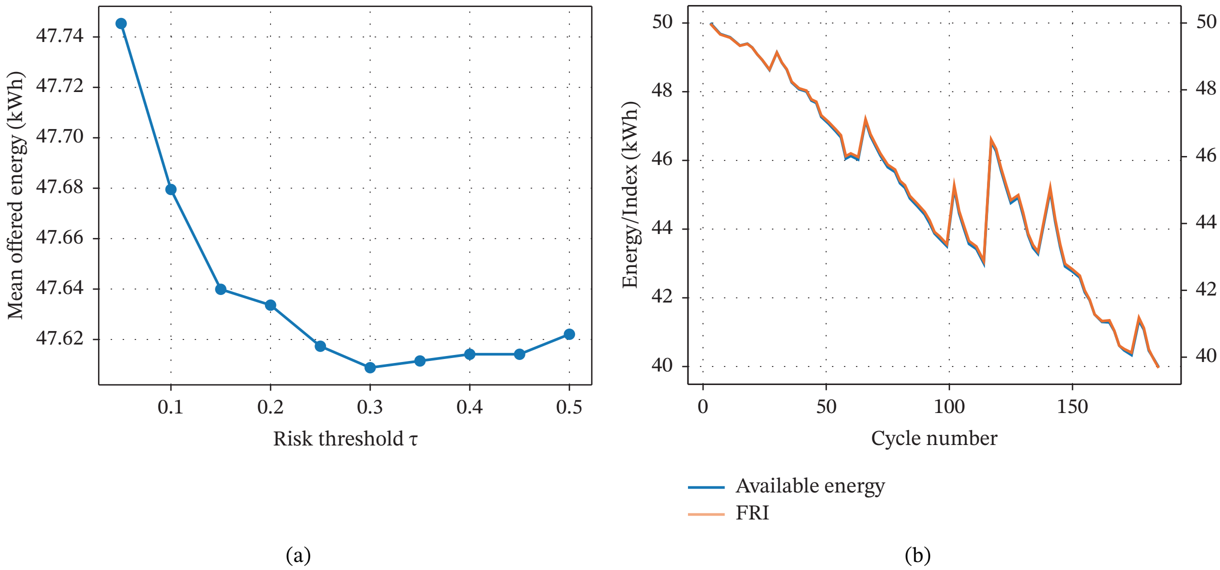


FIGURE 7 | Reliability-aware dispatch behavior. (a) Offered energy as a function of the risk threshold. (b) Dispatch behavior over the degradation trajectory.

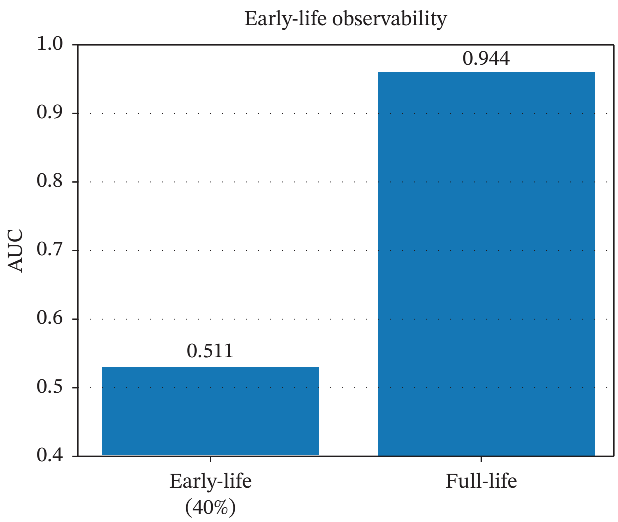


FIGURE 8 | Generalization to early-life batteries.

TABLE 4 | Operational benchmark comparison.

Policy	Failure rate	Delivered energy (kWh)
Always dispatch	0.103	43.38
SOH threshold	0.089	43.07
RUL threshold	0.000	47.53
Proposed risk policy	0.029	46.97

The baseline policy yields a failure rate of 10.31%, showing that unrestricted operation exposes the system to substantial operational risk. Introducing hazard-based reliability prediction produces the largest improvement: when raw hazard probabilities are used to filter risky operation, the failure rate decreases to 3.00%, while delivered energy increases to 46.94 kWh.

Applying probability calibration further refines the operational decision variable. The calibrated hazard policy, which represents the proposed framework, reduces the failure rate slightly further to 2.95% while maintaining nearly identical delivered energy (46.97 kWh). Although the numerical improvement relative to the raw hazard policy is modest, calibration is important because it converts model scores into empirically consistent probabilities that can be used directly in operational control.

The RUL threshold policy achieves 0% observed failures and 48.22 kWh of delivered energy in the evaluated dataset. However, this benchmark is based on a fixed deterministic lifetime rule and does not provide calibrated probabilistic risk estimates across multiple operational horizons. Therefore, its role in this study is as a benchmark policy rather than as part of the proposed decision framework.

To visualize the contribution of the proposed model components, Figure 9 presents the failure-rate reduction from the baseline policy to the raw hazard model and then to the calibrated hazard model.

The figure highlights that the transition from the baseline policy to hazard-based decision-making produces the largest reduction in failure rate, while calibration provides an additional improvement in the reliability of the decision variable. Overall, the ablation analysis confirms that the proposed operational reliability framework achieves a strong balance between reliability and operational flexibility. Hazard-based risk estimation significantly reduces failure exposure, and calibration improves the trustworthiness of the probability estimates used for operational control.

6 | Discussion

6.1 | Interpretation of Findings

The results indicate that battery reliability is inherently horizon dependent. Conventional indicators such as SOH and RUL

TABLE 5 | Ablation analysis of operational reliability policies.

Configuration	Failure rate	Delivered energy (kWh)
Always operate (baseline)	10.31%	43.38
Raw hazard probability	3.0%	46.94
Calibrated hazard probability (proposed)	2.95%	46.97
RUL threshold policy	0.00%	48.22

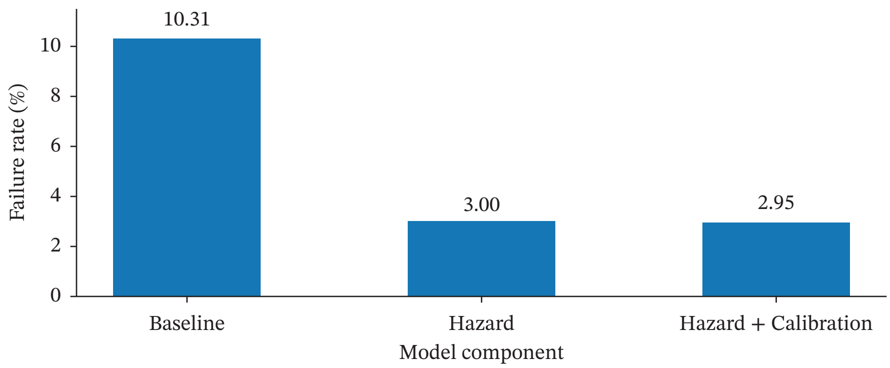


FIGURE 9 | Ablation comparison of operational failure rates across different decision policies.

describe long-term degradation progression but do not quantify short-term operational safety. As demonstrated in the results, batteries with comparable health states may exhibit substantially different near-term failure risk, while lifetime predictions remain smooth even close to failure.

The proposed hazard formulation reframes battery diagnostics from a degradation estimation task into an operational reliability assessment. Rather than predicting when a battery will reach EOL, the model estimates whether the battery can safely complete a specified service commitment. Consequently, the output becomes directly interpretable in operational decision-making.

6.2 | Novelty of the Proposed Framework

The contribution of this work is not a new degradation predictor but a new operational perspective on battery prognostics. The framework is as follows:

1. Transforms prognostic indicators into a decision-level reliability variable,
2. Introduces multihorizon failure intelligence aligned with service duration,
3. Produces calibrated probabilities interpretable as operational risk, and
4. Enables reliability-aware flexibility dispatch.

Most existing studies target maintenance planning, whereas the proposed method produces a dispatch-ready control signal compatible with grid operation.

6.3 | Comparison With Existing Work

Traditional methods provide maintenance indicators but require heuristic thresholds for operational use. Reliability models improve interpretability but typically operate at a fixed horizon. The proposed method uniquely produces a calibrated, horizon-aware reliability measure directly usable in control decisions. Comparisons with existing works with their approaches and operational usabilities are summarized in Table 6.

6.4 | Practical Implications

The framework enables aggregators and grid operators to incorporate battery health into operational planning. Instead of binary acceptance rules, the system allows risk-aware participation by:

- Deciding whether to accept a flexibility bid,
- Limiting unsafe dispatch,
- Reducing unexpected failure events, and
- Gradually retiring degraded batteries rather than abrupt removal.

This capability bridges predictive maintenance and operational control, enabling batteries to participate in flexibility markets while respecting reliability constraints.

6.5 | Limitations

Despite promising results, several limitations remain. First, failure is defined using an SOH threshold rather than physical failure mechanisms, which may not fully represent real safety events. Second, the dataset consists of laboratory cycling data and does not capture the full variability of field BESS operation. Third, the dispatch policy is evaluated offline without market interaction or system-level optimization. Finally, safety-critical phenomena such as thermal runaway and protection system faults are not explicitly modeled.

6.6 | Future Research Directions

Future work should extend the framework toward practical deployment by integrating reliability prediction with power

TABLE 6 | Comparison with existing battery prognostics approaches.

Approach	Output	Operationally usable	Multihorizon	Calibrated risk
SOH estimation [2, 39]	Health level	No	No	No
RUL prediction [13, 33]	Lifetime	No	No	No
Probabilistic RUL [16, 44]	Distribution	Limited	No	Partial
Reliability models [18, 20]	Failure rate	Limited	Single horizon	Limited
Proposed method	Failure probability	Yes	Yes	Yes

system optimization, modeling multibattery aggregation behavior, and enabling real-time market bidding. Incorporating physics-informed failure definitions and safety event datasets would further strengthen reliability interpretation.

7 | Conclusion

This paper presented a decision-oriented reliability framework for BESSs based on multihorizon discrete-time hazard learning. Instead of estimating degradation indicators such as SOH or RUL, the proposed approach predicts the probability that a battery cannot safely complete a predefined service commitment. This reformulation aligns battery diagnostics with the operational requirements of modern power systems, where short-term reliability rather than lifetime prediction determines dispatch feasibility.

Experimental evaluation across multiple lithium-ion batteries demonstrated that lifetime indicators do not reliably represent near-term operational safety. The proposed model produced calibrated failure probabilities that remained stable across different service horizons and generalized to unseen batteries. When integrated into a risk-threshold dispatch policy, the predicted reliability measure enabled a controllable trade-off between delivered energy and failure occurrence, outperforming conventional SOH- and RUL-based decision rules.

The results show that battery prognostics can be transformed from a maintenance-oriented prediction task into a real-time operational decision variable. By directly linking health estimation to dispatch control, the framework enables reliability-aware participation of battery storage in flexibility services.

Future work will focus on integrating the reliability prediction into power system optimization and market operation, extending the approach to aggregated storage fleets, and incorporating physics-based failure definitions to support field deployment.

Author Contributions

Conceptualization: Tareq Anwar Shikdar; methodology: Tareq Anwar Shikdar; software: Tareq Anwar Shikdar; validation: Tareq Anwar Shikdar; formal analysis: Tareq Anwar Shikdar; investigation: Tareq Anwar Shikdar; writing–original draft: Tareq Anwar Shikdar; writing–review and editing: Hannu Laaksonen; supervision: Hannu Laaksonen.

Acknowledgments

Use of AI-Assisted Language Tools: The authors used AI-based language tools, including Grammarly, QuillBot, and ChatGPT, solely for grammar correction and language refinement during manuscript preparation. The scientific concepts, methodology, simulations, analyses, and conclusions are entirely the authors' original work, and the authors take full responsibility for the content of the manuscript.

Funding

This research received no external funding and was conducted as part of the authors' academic research activities at the University of Vaasa: Vaasan yliopisto. FinELib. Open access publishing facilitated by Vaasan yliopisto, as part of the Wiley - FinELib agreement.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The dataset used in this study is publicly available from the NASA Prognostics Center of Excellence battery dataset repository. The processed data and code used to generate the results are available from the corresponding author upon reasonable request.

References

1. Z. Almutairi, H. A. Bheyan, H. Al-Ansary, and A. M. Eltamaly, "Reducing Lithium-Ion Battery Testing Costs Through Strategic Sample Optimization," *Processes* 13, no. 7 (2025): 2030, https://doi.org/10.3390/pr13072030.

2. M. Li, W. Xu, S. Zhang, et al., "State of Health Estimation and Battery Management: A Review of Health Indicators, Models and Machine Learning," *Materials* 18, no. 1 (2025): 145, https://doi.org/10.3390/ma18010145.

3. B. Zhao, W. Zhang, Y. Zhang, C. Zhang, C. Zhang, and J. Zhang, "Lithium-Ion Battery Remaining Useful Life Prediction Based on Interpretable Deep Learning and Network Parameter Optimization," *Applied Energy* 379 (2025): 124713, https://doi.org/10.1016/j.apenergy.2024.124713.

4. B. Celik, R. Sandt, L. C. P. dos Santos, and R. Spatschek, "Prediction of Battery Cycle Life Using Early-Cycle Data, Machine Learning and Data Management," *Batteries* 8, no. 12 (2022): 266, https://doi.org/10.3390/batteries8120266.

5. A. Fouka, K. Lepenioti, A. Bousdekis, and G. Mentzas, "A Machine Learning Framework for Li-Ion Battery Lifetime Prognostics," in *2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA)* (Corfu, Greece, 2022), 1–8, https://doi.org/10.1109/IISA56318.2022.9904407.

6. B. Song, G. Yue, D. Guo, et al., "Prediction of the Remaining Useful Life of Lithium–Ion Batteries Based on Mode Decomposition and ED-LSTM," *Batteries* 11, no. 3 (2025): 86, https://doi.org/10.3390/batteries11030086.

7. A. Tang, Y. Huang, S. Liu, Q. Yu, W. Shen, and R. Xiong, "A Novel lithium-ion Battery State of Charge Estimation Method Based on the Fusion of Neural Network and Equivalent Circuit Models," *Applied Energy* 348 (2023): 121578, https://doi.org/10.1016/j.apenergy.2023.121578.

8. F. Wang, Z. Zhai, Z. Zhao, Y. Di, and X. Chen, "Physics-Informed Neural Network for lithium-ion Battery Degradation Stable Modeling and Prognosis," *Nature Communications* 15, no. 1 (2024): 4332, https://doi.org/10.1038/s41467-024-48779-z.

9. E. H. E. Bayoumi, M. De Santis, and H. Awad, "A Brief Overview of Modeling Estimation of State of Health for an Electric Vehicle's Li-Ion Batteries," *World Electric Vehicle Journal* 16, no. 2 (2025): 73, https://doi.org/10.3390/wevj16020073.

10. Z. Ye and J. Yu, "State-Of-Health Estimation for lithium-ion Batteries Using Domain Adversarial Transfer Learning," *IEEE Transactions on Power Electronics* 37, no. 3 (March 2022): 3528–3543, https://doi.org/10.1109/TPEL.2021.3117788.

11. T. Xiahou, Y. Liu, Z. Zeng, and M. Wu, "Remaining Useful Life Prediction with Imprecise Observations: An Interval Particle Filtering Approach," *IISE Transactions* 55, no. 11 (2023): 1075–1090, https://doi.org/10.1080/24725854.2022.2125602.

12. A. Hassan, G. V. Hollweg, W. Su, X. Zhou, and M. Wang, "Degradation-Aware Bi-Level Optimization of Second-Life Battery Energy Storage System Considering Demand Charge Reduction," *Energies* 18, no. 15 (2025): 3894, https://doi.org/10.3390/en18153894.

13. R. Ibraheem, T. I. Cannings, T. Sell, and G. dos Reis, "Robust Survival Model for the Prediction of Li-ion Battery Lifetime Reliability and Risk Functions," *Energy and AI* 19 (2025): 100465, https://doi.org/10.1016/j.egyai.2024.100465.

14. M. Iftikhar, M. Shoaib, A. Altaf, et al., "A Deep Learning Approach to Optimize Remaining Useful Life Prediction for Li-ion Batteries," *Scientific Reports* 14, no. 1 (2024): 25838, https://doi.org/10.1038/s41598-024-77427-1.

15. C. Hu, M. Geng, H. Yang, et al., "A Review of Capacity Fade Mechanism and Promotion Strategies for Lithium iron Phosphate Batteries," *Coatings* 14, no. 7 (2024): 832, https://doi.org/10.3390/coatings14070832.

16. F. C. Mushid and M. F. Khan, "Battery Energy Storage for Ancillary Services in Distribution Networks: Technologies, Applications, and Deployment Challenges—A Comprehensive Review," *Energies* 18, no. 20 (2025): 5443, https://doi.org/10.3390/en18205443.

17. C. H. Parimala, P. S. Varma, R. S. Kumar, A. R. Singh, B. Mohapatra, and H. K. Addis, "Uncertainty Aware Hybrid Learning Framework for Fast and Safe Charging of lithium-ion Batteries Using multi-fidelity Observers," *Scientific Reports* 16, no. 1 (2026): 2452, https://doi.org/10.1038/s41598-025-31976-1.

18. K. Tang, B. Luo, D. Chen, et al., "The State of Health Estimation of Lithium-Ion Batteries: A Review of Health Indicators, Estimation Methods, Development Trends and Challenges," *World Electric Vehicle Journal* 16, no. 8 (2025): 429, https://doi.org/10.3390/wevj16080429.

19. C. Liu, H. Li, K. Li, Y. Wu, and B. Lv, "Deep Learning for State of Health Estimation of Lithium-Ion Batteries in Electric Vehicles: A Systematic Review," *Energies* 18, no. 6 (2025): 1463, https://doi.org/10.3390/en18061463.

20. G. Liu, Z. Deng, Y. Xu, et al., "Lithium-Ion Battery State of Health Estimation Based on CNN-LSTM-Attention-FVIM Algorithm and Fusion of Multiple Health Features," *Applied Sciences* 15, no. 13 (2025): 7555, https://doi.org/10.3390/app15137555.

21. Q. Wang, M. Ye, X. Cai, D. U. Sauer, and W. Li, "Transferable data-driven Capacity Estimation for lithium-ion Batteries with Deep Learning: A Case Study from Laboratory to Field Applications," *Applied Energy* 350 (2023): 121747, https://doi.org/10.1016/j.apenergy.2023.121747.

22. Y. Wang and B. Jiang, "Attention Mechanism-Based Neural Network for Prediction of Battery Cycle Life in the Presence of Missing Data," *Batteries* 10, no. 7 (2024): 229, https://doi.org/10.3390/batteries10070229.

23. K. A. Severson, P. M. Attia, N. Jin, et al., "Data-Driven Prediction of Battery Cycle Life Before Capacity Degradation," *Nature Energy* 4, no. 5 (2019): 383–391, https://doi.org/10.1038/s41560-019-0356-8.

24. G. S. Abebe and B. Gou, "Cross-Domain Transfer Learning for Lithium-Ion Battery SOH Estimation with Partial Charging and Limited Data," in *2025 IEEE 20th Conference on Industrial Electronics and Applications (ICIEA)* (Yantai, China, 2025), 1–6, https://doi.org/10.1109/ICIEA65512.2025.11148887.

25. H. S. Kouhestani, L. Liu, R. Wang, and A. Chandra, "Data-Driven Prognosis of Failure Detection and Prediction of lithium-ion Batteries," *Journal of Energy Storage* 70 (2023): 108045, https://doi.org/10.1016/j.est.2023.108045.

26. L. Liu, "Data-Driven Prognosis of Multiscale and Multiphysics Complex System Anomalies: Its Application to lithium-ion Batteries Failure Detection," *Journal of the Electrochemical Society* 170, no. 5 (2023): 050525, https://doi.org/10.1149/1945-7111/acd300.

27. X. Liu, Y. Gao, K. Marma, Y. Miao, and L. Liu, "Advances in the Study of Techniques to Determine the lithium-ion Battery's State of Charge," *Energies* 17, no. 7 (2024): 1643, https://doi.org/10.3390/en17071643.

28. H. S. Kouhestani, X. Yi, G. Qi, et al., "Prognosis and Health Management (PHM) of solid-state Batteries: Perspectives, Challenges, and Opportunities," *Energies* 15, no. 18 (2022): 6599, https://doi.org/10.3390/en15186599.

29. A. Giuliano, Y. Wu, J. Yawney, and S. A. Gadsden, "Transformer-Based Transfer Learning for Battery State-of-Health Estimation," *Energies* 18, no. 20 (2025): 5439, https://doi.org/10.3390/en18205439.

30. Z. Fan, Y. Guo, L. Song, J. Du, H. Wang, and X. Li, "BatteryTSFM: Generalizable long-horizon Degradation Prediction Across Conditions and Chemistries via Time Series Foundation Models," *Energy and AI* 22 (2025): 100646, https://doi.org/10.1016/j.egyai.2025.100646.

31. M. J. M. A. Rasul, A. Abbas, J. Baek, and J. Kim, "A Hybrid Ensemble Learning Framework with Uncertainty Quantification for state-of-health Estimation in lithium-ion Batteries," *Measurement* 266 (2026): 120528, https://doi.org/10.1016/j.measurement.2026.120528.

32. W. Lin, Y. Chai, X. Chen, Q. Liu, and F. Feng, "A Generalized Remaining Useful Life Prediction Method Based on Hybrid Model and Sparse Variational Bayesian," *IEEE Transactions on Industrial Informatics* 21, no. 3 (March 2025): 2254–2263, https://doi.org/10.1109/TII.2024.3495776.

33. L. Liu, J. Park, X. Lin, A. M. Sastry, and W. Lu, "A thermal-electrochemical Model that Gives Spatial-dependent Growth of Solid Electrolyte Interphase in a Li-ion Battery," *Journal of Power Sources* 268 (2014): 482–490, https://doi.org/10.1016/j.jpowsour.2014.06.050.

34. X. Lin, J. Park, L. Liu, Y. Lee, W. Lu, and A. M. Sastry, "A Comprehensive Capacity Fade Model and Analysis for Li-ion Batteries," *Journal of the Electrochemical Society* 160, no. 10 (2013): A1701–A1710, https://doi.org/10.1149/2.040310jes.

35. G. Zhiyong, L. Jiwu, and W. Rongxi, "Prognostics Uncertainty Reduction by right-time Prediction of Remaining Useful Life Based on Hidden Markov Model and Proportional Hazard Model," *Eksploatacja i Niezawodność – Maintenance and Reliability* 23, no. 1 (2021): 154–164, https://doi.org/10.17531/ein.2021.1.16.

36. W. Zhang, M. Zhao, X. Du, Z. Gao, and P. Ni, "Probabilistic Machine Learning Approach for Structural Reliability Analysis," *Probabilistic Engineering Mechanics* 74 (2023): 103502, https://doi.org/10.1016/j.probengmech.2023.103502.

37. Q. Wang, H. Yan, Y. Wang, et al., "Probabilistic State of Health Prediction for Lithium-Ion Batteries Based on Incremental Capacity and Differential Voltage Curves," *Energies* 18, no. 20 (2025): 5450, https://doi.org/10.3390/en18205450.

38. P. Sachan and P. Bharadwaj, "Incorporating Uncertainty and Reliability for Battery Temperature Prediction Using Machine Learning Methods," *IEEE Journal of Emerging and Selected Topics in Industrial Electronics* 5, no. 1 (January 2024): 234–241, https://doi.org/10.1109/JESTIE.2023.3327052.

39. S. Zhou, A. Xu, Y. Tang, et al., "Fast Bayesian Inference of Reparameterized Gamma Process with Random Effects," *IEEE Transactions on Reliability* (2023).

40. M. F. H. Shiblee and H. Laaksonen, “Battery-Insight-PSO: A Machine Learning Model for Accurate Prediction of State of Health and Remaining Useful Life in lithium-ion Batteries,” *Future Batteries* 8 (2025): 100114, https://doi.org/10.1016/j.fub.2025.100114.

41. S. S. Madani, Y. Shabeer, F. Allard, et al., “A Comprehensive Review on Lithium-Ion Battery Lifetime Prediction and Aging Mechanism Analysis,” *Batteries* 11, no. 4 (2025): 127, https://doi.org/10.3390/batteries11040127.

42. A. Roy, S. Meshram, R. B. Patil, S. Arun, and A. Kore, “Life Cycle Testing and Reliability Analysis of Prismatic lithium-iron-phosphate Cells,” *International Journal of Sustainable Energy* 43, no. 1 (2024): 2337439, https://doi.org/10.1080/14786451.2024.2337439.

43. H. Xu, L. Cheng, D. Paizulamu, and H. Zheng, “Safety and Reliability Analysis of Reconfigurable Battery Energy Storage System,” *Batteries* 11, no. 1 (2025): 12, https://doi.org/10.3390/batteries11010012.

44. A. Pavão, A. Marot, J. Sintes, et al., “AI Challenge for Safe and Low Carbon Power Grid Operation,” *Energy and AI* 22 (2025): 100564, https://doi.org/10.1016/j.egyai.2025.100564.

45. B. Saha and K. Goebel, “Battery Data Set,” *NASA Ames Prognostics Data Repository, NASA Prognostics Center of Excellence (PCoE)* (Moffett Field, CA, USA, 2007), https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/.

46. J. Zhang, B. Wang, H. Ma, et al., “A Fast Reliability Evaluation Strategy for Power Systems Under High Proportional Renewable Energy—A Hybrid Data-Driven Method,” *Processes* 12, no. 3 (2024): 608, https://doi.org/10.3390/pr12030608.