batteries

MDPI

Review

Toward Trustworthy and Transferable SOH/RUL Estimation for Lithium-Ion Batteries: A Critical Review and Multi-Fidelity Validation Framework from Laboratory Cells to Real-World Packs

Stefan Rizanov [1,*], Anna Stoynova [2] and Georgy Mihov [1]

[1] Department of Electronics, Faculty of Electronic Engineering and Technologies, Technical University of Sofia, 1000 Sofia, Bulgaria
[2] Department of Microelectronics, Faculty of Electronic Engineering and Technologies, Technical University of Sofia, 1000 Sofia, Bulgaria; avstoynova@abv.bg
* Correspondence: srizanov@tu-sofia.bg

Abstract

Reliable state of health (SOH) and remaining useful life (RUL) estimation is essential for lithium-ion battery diagnostics, prognosis, and management across cell, module, and pack levels. Yet the reported performance metrics often remain tied to controlled cell datasets, with batteries degrading due to chemistry shifts, protocol variation, temperature changes, inconsistent and insufficient measurements, and pack-level heterogeneity. This critical review investigates what evidence is required before an SOH/RUL estimator can be considered trustworthy, transferable, and suitable for battery management system deployment. Based on a de-duplicated classified set of 176 scientific works and a supplementary evidence audit workbook, this review synthesizes model-based, machine learning, deep learning, transfer learning, physics-informed, impedance-based, thermographic, relaxation-based, and digital twin approaches through observability, robustness, uncertainty calibration, transferability, and deployment feasibility. A compact mathematical framework formalizes the health inference, domain shift, cross-fidelity degradation, calibrated uncertainty, and BMS-facing validation criteria. The analysis argues that deployment-ready battery health intelligence should be evaluated as an evidence system rather than as a point prediction task. The proposed multi-fidelity validation framework links synthetic cells, controlled aging, module (pack) testing, fleet shadow operation, and closed-loop safety-governed deployment using acceptance criteria, based on worst-domain error, calibration data, warning risk, and computational feasibility.

Keywords: lithium-ion batteries; state of health; remaining useful life; battery diagnostics; battery prognosis; battery management systems; pack-level estimation; uncertainty quantification; digital twins; validation

Academic Editor:Chris Mi

Received: 30 May 2026
Revised: 5 July 2026
Accepted: 7 July 2026
Published: 15 July 2026

1. Introduction

Battery-health estimation has moved from an academic diagnostic exercise to a systems problem with direct implications for fast charging, warranty exposure, maintenance scheduling, safety supervision, second-life rating and usefulness, and grid service availa-

 https://doi.org/10.3390/batteries12070255

bility. In this transition, the classic distinction between state of health and remaining useful life becomes operationally important. SOH is not a single universal scalar metric; rather, depending on the area of application, the relevant targeted health response may be capacity retention, internal resistance growth, power capability, lithium active material loss, or a composite task-specific health index. Likewise, RUL is not a fixed property of a cell but a conditional forecast that depends on the future load conditions, ambient temperature, state of charge (SOC) window, and control policy. Reviews by Berecibar et al., Vidal et al., Tian et al., Yao et al., and Liu et al., and later authors established this conceptual basis, while Severson et al. showed that early-cycle data can contain highly predictive aging-informative signatures under controlled protocols [1–6].

The central challenge today is no longer whether a neural network, Gaussian process, Kalman filter, or equivalent-circuit model can fit one laboratory dataset. The challenge is whether an estimator remains credible when the operating environment changes. Real systems experience temperature swings, current transients, rest period interruptions, balancing events, incomplete charge processes, manufacturing spread, sensor drift, and labeling deficiencies. Sulzer et al. argued that field data fundamentally alters the lifetime prediction problem because labeling is often delayed due to subjective factors, operating histories are irregular, and co-contributing factors cannot be eliminated experimentally. The recent literature on electric vehicles, battery energy storage systems, and digital twins confirms that the gap between cell-benchmark performance and pack-level decision usefulness remains large [7–13].

This conceptual review therefore takes a battery-centered deployment perspective that aligns directly with the scope of battery diagnostics, prognosis, modeling, testing, and battery management systems (BMSs). Rather than organizing the related literature only by model family, it asks three linked questions. Firstly, which sensing modalities and feature spaces most effectively preserve aging-relevant information under realistic observability constraints? Secondly, which estimator families are more structurally robust to domain shift, sparse labeling, and cross-fidelity transfer? Thirdly, what is the minimal validation evidence that is required before an estimator may be considered trustworthy for integration into battery management and lifecycle decision-making? Within the scope of this work, these questions will be addressed separately in different sections; more precisely, Section 3 will discuss the topic of optimal estimator family choice, Section 5 will tackle sensing and feature spaces, and finally, Section 7 will address estimator trustworthiness criteria. In a broader sense, a generalized answer that emerges, for the questions put forward, is that robustness is an ecosystem property that involves sensing, the model structure, uncertainty treatment, metadata structuring, and multi-stage validation—not simply an algorithmic choice [6,8,9,11,14].

The contributions of this review are separated into two levels: review analysis contributions and author-defined conceptual synthesis contributions. Firstly, at the review analysis level, this article provides a critical synthesis of the recent SOH/RUL estimation literature through the lenses of deployment-oriented observability, transferability, uncertainty calibration, and validation maturity. Rather than proposing a new estimator, this work systematically aggregates and structures existing empirical evidence. Secondly, at the conceptual–methodological level, this review formulates battery-health estimation as a partially observed, domain-dependent, and fidelity-dependent inference problem. This framing combines established state estimation, probabilistic inference, uncertainty quantification and robust-learning concepts with the author-defined review synthesis related to lithium-ion battery SOH/RUL assessment. Thirdly, this review operationalizes algorithmic trustworthiness into BMS-facing validation metrics, including worst-domain error, coverage, warning risk, cross-fidelity degradation, and online feasibility. Fourthly, it proposes a multi-fidelity validation framework that links synthetic and controlled cell-level evidence to module, pack, field-shadow, and closed-loop deployment stages. The novelty status of these elements relative to the prior literature is summarized explicitly in the next subsection.

Author-Proposed Conceptual Contributions and Relation to Prior Literature

To improve scientific transparency, this subsection distinguishes among concepts adopted from the prior literature, literature-inspired adaptations, and author-defined conceptual contributions introduced in the present review. The purpose is not to claim that every mathematical element below is new. Several components, such as probabilistic prediction, conditional uncertainty summaries, worst-domain risk, and conformal coverage, are well established in the broader estimation, machine learning, uncertainty quantification, and robust-learning literature. The novelty of this review lies in integrating these elements into a battery-specific, deployment-oriented framework for trustworthy and transferable SOH/RUL estimation across cell, module, pack, fleet, and closed-loop BMS contexts.

Table 1 summarizes the main conceptual components of the manuscript, their literature foundation, and their degree of novelty. The labels "adopted," "adapted," "review synthesis," and "author-defined" are used consistently throughout the manuscript to clarify whether a given construct is a generalization of the existing literature or a new conceptual proposal of the present review.

Table 1. Novelty status of the main conceptual contributions of this review.

Manuscript Element	Relation to Prior Literature	Originality and Contribution
Deployment-oriented review synthesis	Based on prior SOH/RUL, BMS, field-data, and digital-twin studies	Review-analysis contribution
Observable-state and probabilistic SOH/RUL framing	Adapted from battery state estimation, probabilistic regression, and uncertainty-quantification literature	Literature-inspired adaptation
Task-dependent health-target vector	Synthesizes capacity, RUL, power capability, and safety-margin concepts usually treated separately	Author-defined synthesis construct
Robustness, calibration, and worst-domain validation logic	Inspired by physics-informed learning, robust learning, and conformal uncertainty methods	Adapted and integrated framework
Cross-fidelity degradation and BMS-facing validation gates	Motivated by battery testing, validation standards, and deployment constraints	Author-defined validation framework
Multi-fidelity validation roadmap from cells to packs and field operation	Builds on digital-twin and battery-validation literature	Author-proposed conceptual framework

This review does not attempt to exhaustively summarize all SOH/RUL estimation algorithms. Instead, it focuses on the evidentiary conditions under which such algorithms can be considered trustworthy and transferable across sensing modes, operating domains, and fidelity levels. The subsequent sections are organized to address the three research questions introduced above, progressing from observability and estimator design to deployment-oriented validation and digital twin readiness.

2. Review Protocol, Set Classification, and Evidence Synthesis Logic

2.1. Review Design and Research Question

This study is structured as a critical conceptual review utilizing a rigorous, reproducible classification protocol. It deliberately eschews a statistical meta-analysis, as the compiled SOH/RUL literature exhibits substantial heterogeneity regarding cell chemistries, form factors, aging protocols, sensing configurations, target definitions, validation splits, data availability, and reporting completeness.

Instead, this review employs evidence mapping to systematically determine the requisite validation criteria for evaluating whether an SOH/RUL estimator is sufficiently trustworthy, transferable, and robust for deployment within vehicle- or grid-level battery management systems (BMSs). Consequently, the central research question of this study is formulated as follows: what empirical evidence is required to substantiate claims that a lithium-ion battery SOH/RUL estimator remains trustworthy and transferable when scaling from controlled laboratory-cell environments to module, pack, vehicle, stationary storage, or fleet-level operations? This inquiry is systematically examined through five distinct analytical lenses: aging observability, robustness against realistic operational variability, cross-domain and multi-fidelity transferability, uncertainty quantification and calibration quality, and BMS-facing deployment feasibility.

2.2. Set Construction and Supplementary Evidence Audit

The final de-duplicated dataset consisted of 176 scientific and technical publication records, including peer-reviewed journal articles, high-quality conference papers, technical reports, industry standards, dataset publications, software repositories, and selected recent preprints that provided relevant evidence for lithium-ion battery-health estimation, prognostics, validation, or deployment. Of these records, 156 were published from 2020 onward and 108 appeared in 2023 or later, reflecting the rapid expansion of research on machine learning(ML), trustworthy AI, physics-informed estimation, field-data learning, and digital-twin-enabled battery diagnostics.

The research dataset was derived from an initial working set of 183 records before de-duplication. During the submission-clean audit, six duplicate or associated non-publication items were segregated from the primary corpus and documented in Supplementary Table S1. The final main block therefore reports 176 de-duplicated publication records. Associated code or software repositories were retained only as traceability artifacts and were not counted as separate publications.

The full set classification and evidence-audit workbook are provided in Supplementary Table S1, tracking bibliographic metadata, source types, diagnostic modalities, estimator families, validation fidelities, variability regimes, trustworthiness dimensions, deployment contexts, evidence roles, classification confidence metrics, and figure/table traceability. The supplementary file therefore functions as an audit trail connecting the review synthesis to the underlying evidence base.

2.3. Eligibility, Classification Axes, and Evidence Scoring

Records were screened in two stages. Titles, abstracts, keywords, and metadata were first used to remove clearly irrelevant studies. Full texts or source records were then assessed if they contributed to SOH/RUL estimation, battery diagnostics, aging observability, transferability, uncertainty, validation methodology, pack-level analysis, field-data prediction, BMS implementation, or digital-twin architectures. Sources were retained when they addressed realistic operating variability, transfer across chemistry or protocol, sparse observability, scaling from cells to modules or packs, mechanistically meaningful diagnostics, uncertainty-aware or interpretable estimation, or deployment-oriented validation.

Each retained source was classified along five primary axes: diagnostic modality, estimator family, validation fidelity, variability mode, and trustworthiness dimension. Diagnostic modalities included raw electrical curves, partial-charge features, incremental capacity analysis (ICA), differential voltage analysis (DVA), electrochemical impedance spectroscopy (EIS), distribution of relaxation times (DRT), relaxation-voltage descriptors, thermal or thermographic features, metadata, and multimodal fusion. Estimator families included model-based estimation, equivalent circuit models (ECMs), electrochemical models, feature-based ML, Gaussian process and Bayesian methods, deep sequence learning, transfer learning, domain adaptation, physics-informed learning, hybrid estimators, uncertainty-aware methods, and digital-twin architectures.

To avoid mixing methodological levels, a distinct "methodological role" axis was established to differentiate among estimator families, system-level architectures, and cross-cutting trustworthiness layers. Estimator families include model-based estimation, equivalent-circuit and electrochemical models, feature-based machine learning, Gaussian-process and Bayesian prediction models, deep sequence learning, transfer learning, domain adaptation, physics-informed learning, and hybrid estimators. Digital twins were coded as system-level architectures because they integrate models, data streams, validation artifacts, updating logic, and operational feedback rather than constituting a single SOH/RUL estimator family. Uncertainty quantification, calibration, out-of-distribution (OOD) detection, interpretability, and trustworthy-AI methods were coded as cross-cutting trustworthiness layers because they can be applied on top of several estimator families and deployment architectures.

Because direct numerical comparison across heterogeneous studies would be misleading, evidence synthesis was performed using a conservative ordinal assessment rather than a pooled meta-analysis. The ordinal scores used in the comparative tables are not measurements of algorithmic performance and are not rankings of individual papers. They are evidence-maturity judgments assigned to method families, sensing modalities, system architectures, or trustworthiness layers.

To reduce subjectivity, each score was assigned according to three criteria: the breadth of supporting literature, the realism of the validation conditions, and the consistency of reported behavior across studies. A score of 1 indicates weak or mostly conceptual evidence; 2 indicates preliminary evidence, typically limited to narrow datasets or controlled examples; 3 indicates moderate evidence across several studies but with incomplete transfer, uncertainty, or deployment validation; 4 indicates strong evidence under multiple operating conditions or external holdouts; and 5 indicates deployment-oriented evidence involving realistic variability, pack- or field-facing evaluation, calibrated uncertainty, or explicit BMS feasibility constraints. When evidence was mixed, the lower score was assigned unless strong independent evidence supported the higher level.

The scores should therefore be interpreted as conservative review-level comprehensiveness indicators. They summarize how convincingly the literature supports a given capability, not how well a single algorithm performs on a particular benchmark. This scoring policy was applied consistently across estimator family, diagnostic modality, and validation readiness comparisons. Table 2 shows the evidence maturity scoring rubic.

Table 2. Evidencematurity scoring rubric used in the review-level comparative assessment.

Score	Evidence Maturity Level	Interpretation
1	Weak	Limited evidence; isolated studies or poorly validated claims
2	Preliminary	Early evidence with restricted validation scope
3	Moderate	Consistent evidence across multiple studies under controlled conditions
4	Strong	Robust evidence across different operating conditions and datasets
5	Deployment-oriented	Extensive evidence including transferability, uncertainty evaluation, and deployment-relevant validation

The main manuscript cites the sources most directly used in the narrative synthesis, while additional classified records are retained in Supplementary Table S1 to document the evidence map, research set composition, and methodological traceability.

2.4. Protocol Limitations

The term "reproducible" is used here to indicate the transparency of search logic, eligibility criteria, set coding, evidence scoring, and synthesis rules rather than the statistical reproducibility of pooled data. While the protocol improves auditability, it does not claim PRISMA-level exhaustiveness. Furthermore, the classification involves expert judgment—especially for studies spanning multiple estimator families, diagnostic modalities, or validation regimes, and remains limited by incomplete reporting in the primary literature.

Figure 1 summarizes the final de-duplicated dataset by publication year and methodological family. It should be interpreted as an evidence map overview rather than as a pooled meta-analytic dataset.

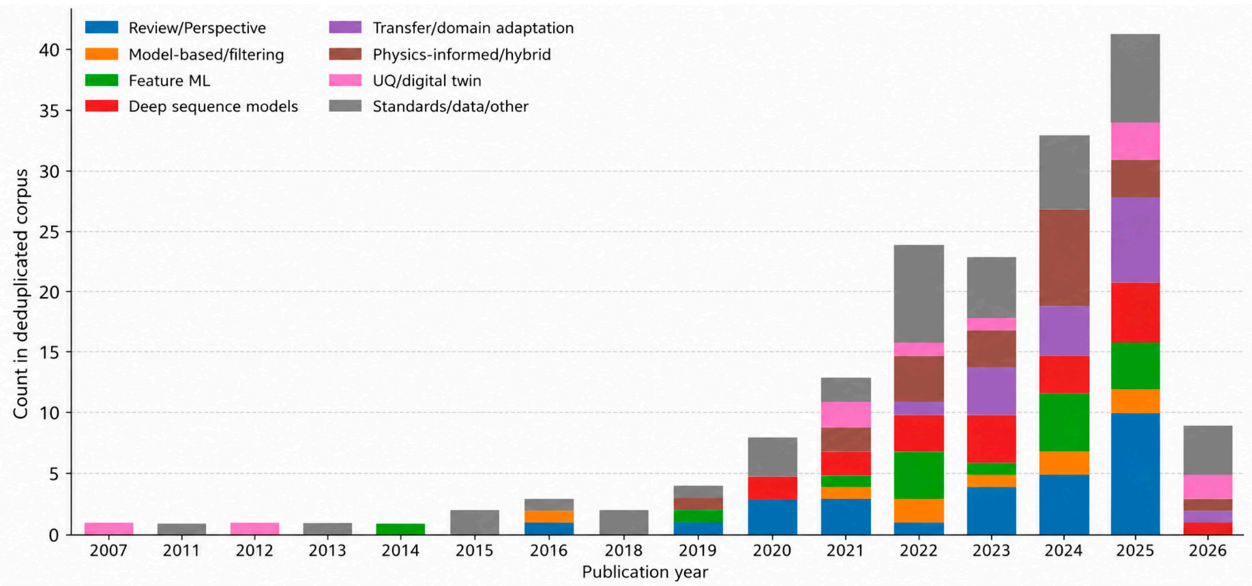

Figure 1. Publication timeline of the final de-duplicated research set.

Figure 1 shows the temporal expansion of the de-duplicated evidence corpus and the diversity of the methodological families. The sharp growth after 2020 confirms that estimator design, transfer learning, uncertainty-aware methods, digital twins, and deployment-oriented validation have advanced rapidly, while shared validation standards for robustness, calibration, and cross-fidelity transfer remain less mature.

Before comparing estimator families and sensing modalities, it is also necessary to establish a consistent vocabulary for the main targets, assumptions, and realism constructs used throughout the review. Table 3 therefore defines the core concepts adopted in the paper, including SOH, RUL, observability, deployment domain, fidelity level, realistic variability, trustworthiness, transferability, and BMS-facing validation.

Table 3. Definitions and realism constructs used in this review.

Construct	Operational Definition Used in This Review	Importance
SOH target	Capacity retention, power fade, resistance growth, or task-specific health index	Single scalar SOH is often insufficient for fast-charge, thermal, or power-limited use cases.
RUL target	Time/cycles/throughput until an application-specific end-of-life threshold	RUL depends on duty cycle, ambient conditions, and remaining control policy.
Realistic variability	Temperature, C-rate, SOC window, rest, current transients, aging path, chemistry, manufacturing spread, balancing state	High laboratory accuracy under a narrow protocol rarely implies field robustness.
Observability regime	Full cycling data, partial charge windows, relaxation snippets, EIS sweeps, sparse pack sensors, maintenance labels	Estimator choice should be matched to observability rather than benchmark convenience.
Validation unit	Cell, module, pack, vehicle/fleet, or digital twin	Evidence does not transfer automatically across fidelities.
Trustworthiness target	Calibration, drift awareness, OOD detection, interpretable failure modes, rollback-safe integration	Accuracy without quantified risk is insufficient for safety-relevant BMS decisions.

Table 3 is not only a definitional aid. It clarifies why apparently similar SOH/RUL studies may lead to different conclusions when their targets, sensing assumptions, validation units, and deployment domains differ. By making these assumptions explicit, the table supports the main methodological argument of this review: trustworthy battery-health estimation requires alignment between the claimed health target, the available observability, the validation fidelity, and the operational decision context.

3. Mathematical Framing for Trustworthy and Transferable SOH/RUL Estimation

For *Batteries* readers, the central issue is not whether SOH/RUL can be predicted in principle, but how the inference framework for the battery state should be formulated under partial observability, domain shifting, and implementation constraints. A realistic SOH/RUL estimator must handle three coupled mismatches. The first is a physics-to-signal mismatch: degradation mechanisms such as loss of lithium inventory, loss of active material, impedance growth, gas formation, and thermal non-uniformity are only partially observable in terminal voltage, current, temperature, or sparse auxiliary diagnostics. The second is a domain mismatch: laboratory protocols sample only a small part of the operating dynamics encountered in real-world applications. The third is a fidelity mismatch: evidence obtained from single cells does not automatically transfer to modules, packs, or fleets [2,6,8–11,14–16].

These mismatches imply that SOH and RUL should be framed as partially observed, context-dependent inference problems. Capacity-based SOH remains useful, but it is not always sufficient for the deployment decisions that are critical for readers of *Batteries*. Fast-charge de-rating, power limitation, thermal risk management, balancing policy, and maintenance scheduling often depend at least as much on resistance growth, power capability, and uncertainty margins as on nominal capacity. Any review that seeks deployment relevance should therefore formalize battery health in a way that accommodates multiple task-dependent targets [2,6,17–20].

The literature also shows that 'realistic variability' is not one variable but a hierarchical structure. At the lowest level are controlled perturbations of temperature, current, SOC window, or rest time. At higher levels are chemistry changes, manufacturer differences, calendar-and-cycling interaction, sensor degradation, pack thermal gradients, control-policy shifts, and user-dependent duty cycles. A transferable estimator is one whose useful representation survives at least part of this hierarchy. Accordingly, the equations below should be read as compact review synthesis formalizations. When an expression is adapted from an established literature stream, the source is cited explicitly; when an expression is introduced in this review, that is stated directly [7–13,21–23].

The novelty status of the mathematical components introduced in this section is summarized in Table 1.Accordingly, equations derived from established probabilistic, robust learning or uncertainty quantification principles are identified as adopted or adapted, whereas the battery-specific validation indices, deployment gates, and multi-fidelity synthesis framework are identified as author-defined review contributions.

3.1. State, Observability, and Context

A review paper does not need to impose one universal battery health model, but it should define the minimum mathematical protocol that an estimator must satisfy. Let x_t denote the observable diagnostic state assembled from the sensing stack that is actually available in practice, including current, voltage, temperature, occasional impedance measurements, relaxation signatures, and metadata on the operating regime. Equation (1) adopts an augmented observable-state notation adapted from state and parameter estimation formulations used in model-based battery diagnostics and joint SOC/SOH estimation [14–16].

$$x_t = [\, I_{0:\mathrm{t}}, V_{0:\mathrm{t}}, T_{0:\mathrm{t}}, Z_t, R_t, m_t \,] \tag{1}$$

where I, V, and T are the time histories of current, voltage, and temperature up to time t, Z_t denotes EIS-derived descriptors when they are available, R_t denotes relaxation/rest-response descriptors, and m_t denotes contextual metadata such as chemical composition specifics, SOC window, pack topology, cooling mode, and age history. The health target is not necessarily a single scalar but a task-dependent latent vector.

Equation (2) is a review synthesis target vector introduced here to unify capacity, remaining life, power capability, and safety margin perspectives that are often discussed separately in battery health and PHM reviews; it is not transcribed verbatim from one prior paper [2,6,24].

$$y_t = [\, \mathrm{SOH}_t, \mathrm{RUL}_t, P_t, S_t]\text{\^{}T} \tag{2}$$

where $P_t$can represent power capability and $\mathrm{S_t}$can represent a safety-relevant state or margin. This formulation is particularly relevant for *Batteries* because many deployment failures arise when capacity-centered SOH metrics are used outside the application for which they were validated. However, Equation (2) defines a target space rather than a claim of full online observability. In practice, each component of the vector must be accompanied by an explicit statement of its sensor regime, identifiability status, and validation evidence. For example, capacity-related SOH may be partially identifiable from repeated partial-charge or reference events, resistance and power capability may be proxy-identifiable from current–voltage–temperature transients, RUL is a conditional forecast rather than a directly measured state, and safety-relevant indicators are usually inferred through conservative anomaly or warning indicators rather than uniquely estimated as physical degradation modes.

Practical Identifiability Under BMS Sensor Constraints

Not all latent degradation dimensions are directly observable within practical battery management systems. The multidimensional health vector in Equation (2) should not be interpreted as implying that all battery health components are simultaneously observable from every sensing configuration. In a production BMS, the routinely available measurements are typically pack or cell-group current, terminal voltage, limited temperature channels, balancing information, time stamps, SOC estimates, and operational metadata. Direct measurements of capacity loss, loss of lithium inventory, loss of active material, charge-transfer resistance, lithium plating, gas formation, or spatial thermal non-uniformity are generally not available during normal operation. Therefore, the proposed health vector should be interpreted as a decision-target space, whose individual components have different degrees of practical identifiability depending on the sensor stack, operating excitation, validation fidelity, and intended BMS decision.

In this review, a component of the health vector is considered practically identifiable under a sensor set S only when it can be estimated with bounded error, calibrated uncertainty, and acceptable robustness on the relevant deployment domain. Components that can influence the measurements but cannot be uniquely separated from other degradation or operating effects are treated as proxy-observable rather than directly identifiable. Components that require external reference tests, service-bay diagnostics, EIS/DRT, thermography, or destructive/post-failure analysis are not considered directly identifiable from standard online BMS data.

Table 4 summarizes the practical observability of the main health vector components under typical BMS sensing constraints.

Table 4. Practical observability of the main healthvector components.

HealthVector Component	Identifiability with Typical BMS I,V,TI,V,TI,V,T and Metadata	Practical Interpretation
Capacity-related SOH	Partially identifiable when repeated charge/discharge windows or reference events exist	Reliable only when operating windows contain sufficient aging information; otherwise, uncertainty should be widened
Resistance/power capability	Proxy-identifiable from voltage response, current transients, SOC, and temperature	Useful for power limitation and derating, but strongly confounded by temperature, SOC, and pulse history
RUL	Not directly observable; conditionally forecastable	Should be reported as a predictive distribution conditional on future duty cycle, temperature, SOC window, and control policy
Safety-relevant margin	Weakly proxy-observable from abnormal voltage, temperature, imbalance, or residual patterns	Suitable mainly for conservative warning logic, OOD detection, and fallback decisions, not precise mechanism identification
Mechanistic degradation modes, e.g., LLI, LAM, charge-transfer resistance, diffusion limitation, LI plating	Generally not uniquely identifiable from standard BMS data alone	Requires auxiliary diagnostics such as ICA/DVA, EIS/DRT, relaxation tests, thermal imaging, reference capacity tests, or post-mortem evidence

3.2. Probabilistic Estimator and Uncertainty

Under realistic variability, the estimator should be defined as a conditional predictive distribution rather than as a deterministic point map. This makes uncertainty, confidence bounds, and calibration first-class outputs rather than optional post-processing. Equation (3) follows the standard probabilistic regression formulation used in predictive uncertainty and probabilistic forecasting [25,26].

$$p_\theta(y_t \mid x_t, d, h) \tag{3}$$

In this notation, d denotes the deployment domain, including chemistry, manufacturer, thermal environment, control policy, and user behavior, while h denotes the fidelity level of the evidence used to train or validate the model. A practical point estimate and its uncertainty summary can then be written as Equation (4), which uses the standard conditional mean and conditional variance summaries of the predictive distribution in Equation (3) [25,26].

$$\hat{y}_t = \mathrm{E}[y_t \mid x_t, d, h], \quad U_t = \mathrm{Var}[y_t \mid x_t, d, h] \tag{4}$$

This simple probabilistic framing clarifies a central message of the review: an estimator cannot be called trustworthy if it reports only a point value but offers no information about confidence, sensitivity to domain shift, or the expected width of the error under sparse sensing.

3.3. Robustness, Transferability, and Physics Consistency

Robust deployment of battery prognostic models requires more than achieving high predictive accuracy on a specific dataset. In a practical application, models are expected to operate under varying temperatures, duty cycles, charging strategies, aging histories, battery chemistries, and sensor configurations. Consequently, robustness should be understood not only as resistance to measurement noise but also as the ability to maintain acceptable performance under operational variability and domain shift.

Transferability has therefore emerged as a critical requirement for next-generation SOH and RUL estimation frameworks. A model developed using one battery chemistry, operating profile, or testing protocol should ideally preserve its predictive capability when applied to related but previously unseen conditions. However, successful transfer cannot be evaluated solely through prediction error metrics. Low prediction error does not necessarily mean that the learned representations remain physically meaningful or reliable outside the original training domain.

For battery prognostics, transferability should be accompanied by physics consistency. Physics consistency refers to the extent to which model predictions and learned representations remain compatible and physics-informed in relation to established degradation mechanisms and electrochemical behavior. This requirement is particularly important for machine learning and deep learning approaches, where latent representations may capture statistical correlations that do not necessarily correspond to physically plausible degradation processes.

Several practical criteria may be considered when evaluating the physical consistency of transferred representations:

- Monotonicity consistency: estimated SOH should not exhibit systematic improvement during progressive aging unless a physically justified explanation exists.
- Temperature consistency: degradation trends should remain compatible with known temperature-dependent aging mechanisms.
- Impedance consistency: health indicators should exhibit physically plausible relationships with resistance growth and impedance evolution.
- Relaxation consistency: features derived from voltage relaxation analysis should remain compatible with expected diffusion and relaxation dynamics.
- Uncertainty consistency: predictive uncertainty should increase when models operate outside their validated domain rather than remaining artificially overconfident.

The proposed criteria do not constitute a formal framework but provide practical indicators for assessing whether transferred knowledge remains physically credible. Their importance becomes increasingly evident when models are deployed across different chemistries, manufacturers, or operating environments, where purely statistical transfer may result in negative transfer and reduced reliability.

https://doi.org/10.3390/batteries12070255

The learning or inference objective should balance nominal accuracy with robustness across domains, agreement with basic battery physics, and uncertainty calibration. A compact review-level objective can therefore be expressed as Equation (5). This objective is proposed here as a review synthesis formulation; it is inspired by physics-informed battery estimation and hybrid learning studies [27–34] together with worst-group/domain-robust learning formulations [35], rather than taken verbatim from a single source.

In the notation below, calligraphic symbols denote review-level mathematical objects: J is an objective functional, H is the set of fidelity levels, D denotes data distributions or deployment-domain sets, L is a loss function, and R_{rob} is a robustness-risk functional. They are used for compactness and do not introduce new battery-specific operators.

$$J(\theta) = \sum_{\mathrm{h}=1}^{H} \alpha_h \mathrm{E}_{(\mathrm{x,y})\sim D_h}$$

$$[\,\mathrm{L}(y, f_\theta(x, d, h))\,]$$

$$+\,\lambda_{\mathrm{phys}}\Omega_{\mathrm{phys}} + \lambda_{\mathrm{cal}}\Omega_{\mathrm{cal}} + \lambda_{\mathrm{tr}}\Omega_{\mathrm{tr}} \tag{5}$$

The first term aggregates prediction loss across the fidelity ladder. Ω_{phys} penalizes physically implausible behavior, such as non-monotonic long-horizon aging trends or inconsistent coupling between resistance growth and power fade. Ω_{cal} penalizes miscalibrated predictive intervals. Ω_{tr} penalizes poor transfer across chemistries, protocols, or sensing stacks. In conceptual terms, this means that a model with the lowest average error on one cell dataset is still suboptimal if it fails under temperature shift, partial charging, or pack heterogeneity.

The weighting coefficients in Equation (5) should not be interpreted as universal constants. They are application-dependent hyper parameters that determine the operational behavior of the estimator and should therefore be selected according to the deployment objective, sensing regime, and acceptable risk level. In practice, their role is to encode the relative importance of nominal prediction accuracy, physical plausibility, uncertainty calibration, and transfer robustness.

A practical coefficient selection procedure can be organized in three steps. The first step is to define that safety-critical and physically mandatory requirements should be treated as constraints rather than as weak penalties. For example, implausible long-term aging trends, violation of known voltage, temperature, or power limits, and unsafe warning behavior should not be compensated by a lower prediction error. As a second step, the remaining weights should be selected using validation domains that are separated by cell, protocol, temperature, chemistry, or fidelity level, rather than by random cycle-level splits. As a third and final step, the selected coefficient set should be subjected to sensitivity analysis to determine whether the estimator remains stable when the relative weights are perturbed.

For practical use, the prediction-loss weight may dominate in controlled laboratory benchmarking, whereas the physics-consistency and calibration weights should increase when extrapolation outside of the training protocol is expected. The transfer robustness weight should increase when the estimator is intended for cross-chemistry, cross-temperature, cross-protocol, or cell-to-pack deployment. Conversely, if the estimator is used only within a narrow and well-characterized operating domain, the transfer penalty may be lower, provided that the deployment boundary is explicitly stated. Thus, Equation (5) has to be understood and used as an operational template whose weights must be selected, validated, and reported together with the intended BMS use case. Table 5 shows a practical guidance for selecting the weight coefficients of the previous numerical description.

https://doi.org/10.3390/batteries12070255

Table 5. Practical guidance for selecting the weighting coefficients in Equation (5).

Objective Term	When the Weight Should be Increased	Practical Selection Logic
Prediction loss	Controlled datasets, narrow operating windows, benchmark comparison	Tune for lowest validation error only after leakage-safe splits
Physics-consistency penalty	Extrapolation, sparse labels, long-horizon RUL, hybrid or physics-informed models	Treat severe physical violations as constraints; tune mild violations as penalties
Calibration penalty	RUL intervals, maintenance planning, warranty, safety-related decisions	Select using coverage error, interval width, and risk-conditioned validation
Transfer-robustness penalty	Cross-temperature, cross-protocol, cross-chemistry, cell-to-pack, or field deployment	Tune on external domains or worst-domain validation rather than pooled average error
Computational/deployment penalty, if included	Embedded BMS implementation, limited memory, real-time inference	Enforce latency, memory, and sensing constraints as deployment gates

A complementary deployment-oriented notion of robustness is the worst-domain risk. Equation (6) adapts the worst-group or worst-domain risk idea from distribution-robust learning to the deployment-domain setting considered here [35].

$$\mathrm{R_{rob}}(\theta) \;=\; \max_{\mathrm{d\ in\ D\_deploy}} \mathrm{E}[\,\mathrm{L}(y, f_\theta(x, d, h)) \;\mid\; d\,] \tag{6}$$

where D_deploy denotes the set of anticipated deployment domains, including different battery chemistries, operating temperatures, usage patterns, and aging trajectories; L is the prediction loss function; and f_θ represents the prognostic model parameterized by θ.

Equation (6) quantifies the largest expected prediction error across all relevant deployment scenarios and does not rely solely on pooled test-set performance. This perspective is particularly relevant for battery applications because engineering acceptability is often determined by the most challenging operating condition rather than by average-case behavior. A prognostic model exhibiting low mean prediction error may still be unsuitable for deployment if substantial performance degradation occurs under extreme temperatures, fast charging conditions, or previously unseen aging pathways.

From a trustworthiness perspective, worst-domain risk complements conventional accuracy metrics by explicitly accounting for robustness under domain variability and transfer conditions. Together, Equations (5) and (6) highlight that trustworthy SOH and RUL estimation requires the balancing of predictive performance, uncertainty awareness, computational feasibility, transferability, and robustness to deployment-domain shifts.

Future trustworthy battery prognostic systems are therefore expected to evaluate transferability not only through prediction accuracy but also through physics consistency, uncertainty calibration, and robustness under domain variability. Such a perspective establishes a stronger foundation for deployment-oriented SOH and RUL estimation and directly supports the development of trustworthy battery digital twins and multi-fidelity validation frameworks. The practical implications of observability limitations, domain shift, failure modes, continual adaptation, and deployment-oriented validation are discussed in Section 6.

3.4. Multi-Fidelity Validation Criteria

The proposed multi-fidelity roadmap can also be written in quantitative form. Let h = 1, 2, 3, 4 represent laboratory cells, variability-enriched cell studies, module or pack test benches, and field or fleet deployment, respectively. A transferable estimator should control performance degradation between adjacent fidelity levels. Equation (7) is introduced in this review as a compact cross-fidelity degradation index motivated by staged validation and battery digital-twin road-mapping studies [8–10,12,36].

$$\Delta_{\mathrm{h}\to\mathrm{h}+1} = \mathrm{E}[\mathrm{L}_{\mathrm{h}+1}] - \mathrm{E}[\mathrm{L}_{h}] \tag{7}$$

Small values of$\Delta_{\mathrm{h}\to\mathrm{h}+1}$ indicate that the learned representation survives the transition from controlled evidence to more realistic evidence. Trustworthiness also requires calibrated coverage. For a target tolerance ε and nominal confidence $1 - \mu$, a deployment-ready estimator should satisfy Equation (8). This inequality is an author-defined deployment acceptance criterion inspired by tolerance- and threshold-based logic in battery testing standards and validation manuals[17–20].

$$\Pr(|\widehat{\mathrm{SOH}} - \mathrm{SOH}| \le \varepsilon_{\mathrm{SOH}}) \ge 1 - \mu \tag{8}$$

$$\Pr(\mathrm{RUL\ in\ C}_{\sigma}(x_t)) \ge \sigma \tag{9}$$

where $\mathrm{C}_{\sigma}(x_t)$ is a predictive interval or set for RUL at coverage level β. Equation (9) adopts the standard predictive-set coverage guarantee used in conformal prediction and distribution-free uncertainty quantification [37]. These criteria are deliberately generic: they can be implemented with Bayesian models, ensemble methods, conformal prediction, Gaussian processes, physics-informed filters, or hybrid digital twins. Their value in a review article is not algorithmic specificity, but a shared language for comparing evidence across *Batteries*-relevant studies [36–40].

Taken together, Equations (1)–(9) formalize the central thesis of this review: trustworthy and transferable battery-health estimation is not defined by point accuracy alone, but by observability-aware inference, cross-fidelity stability, physical plausibility, and calibrated uncertainty under realistic operating variability.

3.5. BMS-Facing Metrics and Deployment Readiness Rules

From a battery management perspective, the mathematical framing becomes substantially more actionable when it is translated into quantities that a battery management engineer can audit, threshold, and compare across fidelity levels. Equation (10) uses the conventional signed-error notation employed throughout estimation and prognostics studies and is included here for notation completeness [6,14].

$$e_t^{\mathrm{SOH}} = \widehat{\mathrm{SOH}}_t - \mathrm{SOH}_t,\ e_t^{\mathrm{RUL}} = \widehat{\mathrm{RUL}}_t - \mathrm{RUL}_t \tag{10}$$

Let e_t^{SOH} and e_t^{RUL}be the signed health-estimation errors. A deployment-oriented validation vector can then be defined for each fidelity-domain pair (h,d) as Equation (11). This vector is a review synthesis construct proposed in this paper, combining accuracy, calibration, warning risk, and computational criteria motivated by battery management and trustworthy AI requirements [11,17–20,36,37,39,40].

$$\mathrm{m}_{(\mathrm{h,d})}(\theta) = [\mathrm{MAE}, \mathrm{RMSE}, \mathrm{ICE}, \Delta_{\mathrm{h}\to\mathrm{h}+1}, \mathrm{FAR}, \mathrm{MDR}, t_{\mathrm{inf}}, c_{\mathrm{mem}}] \tag{11}$$

where $\mathrm{ICE}_{\beta}(h,d) = |\Pr\left(\mathrm{y\ in\ C}_{\beta}(\mathrm{h,d})\right) - \beta|$is the interval-coverage error, FAR is the false-alarm rate for health-warning logic, MDR is the missed-detection rate, t_{inf} is online inference latency, and c_{mem} is memory or implementation burden. This notation makes explicit that a model may look accurate in offline benchmarking while still being unsuitable for BMS deployment because it is poorly calibrated, computationally heavy, or too risky in warning decisions.

A compact deployment rule may therefore be written as Equation (12). This deployment gate is an author-defined hard-threshold acceptance rule inspired by standards-based acceptance thinking and BMS-oriented go/no-go criteria [17–20].

$$\mathrm{A}(\theta) = 1 \text{ iff } \max_{(\mathrm{h,d})\ \mathrm{in\ H} \times \mathrm{D\_deploy}} \mathrm{m}_{(\mathrm{h,d})}^{(\mathrm{j})} \le \tau_j,$$

$$j \text{ in } J_{\mathrm{hard}} \tag{12}$$

https://doi.org/10.3390/batteries12070255

with additional soft objectives for average performance and interpretability. In words, an estimator is deployment-ready only if it clears predefined thresholds not merely on pooled average error, but on the worst relevant fidelity-domain combinations. This is especially important for cell-to-pack transfer, where pack thermal gradients, balancing behavior, sparse sensing, and partial windows can create failure modes that are invisible in cell-level random splits.

To translate the methodological discussion into deployment-oriented evaluation logic, Table 6 summarizes the BMS-facing validation metrics and acceptance criteria that determine whether an SOH/RUL estimator is credible for practical usage.

Table 6 makes explicit a weakness that remains widespread in the literature: validation is still too often treated as a retrospective reporting exercise rather than as a forward-looking deployment filter. In many published studies, estimators are judged almost exclusively by MAE, RMSE, or correlation metrics, even though these quantities say little about whether the model can be trusted when sensing becomes sparse, operating regimes drift, or decisions must be taken under asymmetric risk. The metrics assembled in Table 6 challenge this narrow-view practice by framing validation as a multi-criterion acceptance problem [11,17–20,36,37,39,40].

Figure 2 shows a BMS-facing acceptance gate diagram for SOH/RUL estimators, illustrating the translation of the deployment validation vector into sequential pass/no-pass rules. An estimator should proceed to shadow operation or closed-loop BMS implementation only when all validation gates are fully satisfied; otherwise, the required corrective action involves revalidation, recalibration, supplementary safeguards, sensing enhancements, or model simplification.

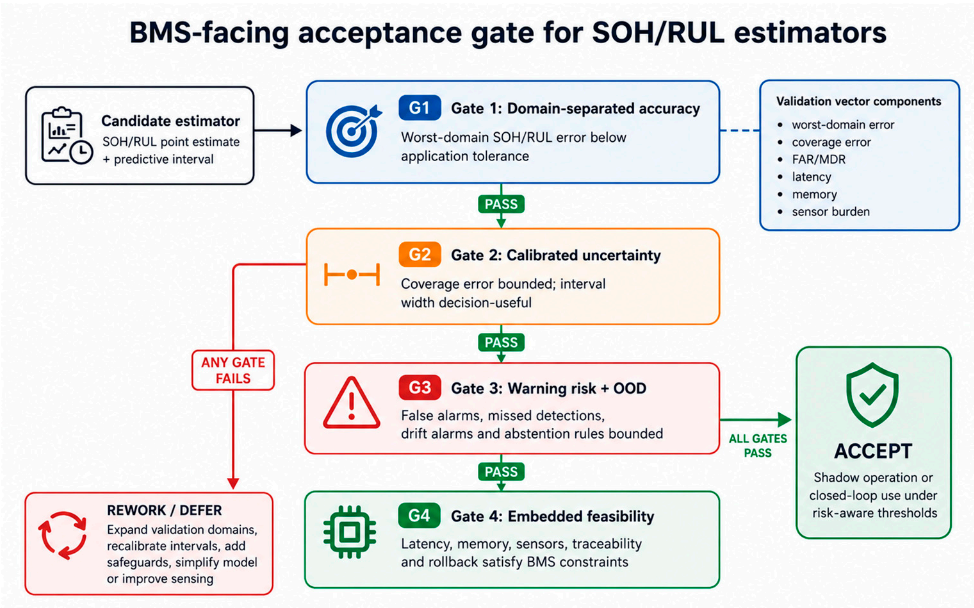

Figure 2. BMS-facing acceptance gates for SOH/RUL estimators

Table 6. BMS-facing validation metrics and deployment readiness gates.

Metric Family	Representative Definition	Why It Matters for BMS	Typical Acceptance Logic
Worst-domain health error	max_dMAE_d or max_dRMSE_d across relevant domains	Avoids hiding failure in difficult chemistries, temperatures, or pack states	Must remain below application-specific SOH tolerance
Cross-fidelity degradation	Delta_(h→h + 1) = E[L_(h + 1)] − E[L_h]	Quantifies loss when moving from cell data to module/pack/field evidence	No sharp performance collapse between adjacent fidelity levels

Coverage/calibration	|Pr(y in C_beta(x)) − beta|	Ensures uncertainty intervals mean what they claim	Coverage error bounded by a predefined margin
Warning risk	FAR and MDR for threshold-triggered health alarms	Prevents excessive false alarms or missed dangerous degradation	Both rates bounded for critical decisions
Online feasibility	t_inf, memory footprint, and sensing burden	Determines whether the estimator can run inside or alongside the BMS	Must satisfy runtime, memory, and sensor-availability constraints

3.6. Chapter Synthesis

As concluding remarks to this section, it is important to note that this chapter addressed one of the main paradigm questions: how can estimator robustness be defined under domain shifts? It is conjectured and formalized that robustness can be quantitatively defined by considering the application domain, cross-fidelity stability, physical plausibility, and calibrated uncertainty under realistic operating variability. Defining the deployment readiness of a given system or estimator necessitates a carefully considered approach to weighting and prioritizing accuracy, uncertainty calibration, computational feasibility and transferability.

4. Estimator Families Under Realistic Variability

Pure model-based approaches remain attractive because they explicitly encode electrochemical or equivalent-circuit structures.

Extended Kalman filtering, particle filters, and dual-estimation schemes can jointly infer states and slowly varying parameters, especially when embedded within an ECM or reduced-order physics models. Their main strengths include physical consistency, low online computational burden, and transparency of state evolution. Conversely, their primary weaknesses are weak parameter identifiability under sparse sensing, sensitivity to model mismatch, and difficulty in representing heterogeneous aging paths without frequent recalibration. For this reason, purely mechanistic approaches still play an important role in constrained systems; however, they are rarely sufficient as standalone, robust estimators under wide operating variability [14–16,41,42].

Feature-based machine learning expanded the field by exploiting observable descriptors, extracted from charge or discharge curves, incremental capacity peaks, differential voltage signatures, resistance estimates, entropy-like indicators, or handcrafted statistical summaries. This family includes support vector regression, random forests, gradient boosting, Gaussian process regression, sparse Bayesian models, and shallow neural networks. Their advantages include a strong performance with a modest data volume and relatively simple deployment when the feature pipeline is stable. However, feature stability itself is often conditional and dependent upon protocol alignment. ICA and DVA are informative but their peak positions, amplitudes, and smoothness vary with charge/discharge rates, temperature, and data preprocessing. Models trained exquisitely on ideally aligned laboratory curves can therefore fail when transferred to partial-charge field windows or to cells with different underlying kinetics [43–49].

Deep sequence models—such as convolutional networks, recurrent networks, temporal convolutional networks, transformers, and graph-enhanced temporal models—address a key weakness of feature engineering by learning directly from current–voltage–temperature curves. They can capture irregular histories and interaction terms that handcrafted features typically miss. Over the last several years, this family of computational methods has become dominant in benchmark publications because it scales well with dataset size and frequently improves average accuracy. Yet two limitations persist. First, deep models tend to learn protocol-specific shortcuts unless training and testing splits are

https://doi.org/10.3390/batteries12070255

carefully designed. Second, their success relies on large and representative datasets that are rare and difficult to acquire under field conditions. In short, sequence models are powerful function approximators, but do not serve as a substitute for robust validation design [18,19,50–54].

Transfer learning and domain adaptation have become crucial because cross-chemistry and cross-platform generalization are now recognized as first-order difficulties rather than corner cases. Domain adversarial methods, instance reweighting, feature alignment, meta-learning, and parameter-efficient fine-tuning have all been explored. Their attraction is clear: they attempt to preserve the learned aging structure while correcting for dataset or platform shift. The main unresolved issue lies in whether the aligned representation remains physically meaningful or merely statistical. This can be evaluated by testing whether transferred features preserve degradation-relevant associations. Such representations must remain disentangled from the reversible operating context, follow plausible aging trends, and remain consistent with independent diagnostic signatures, such as ICA/DVA, EIS/DRT, relaxation voltage, or thermal indicators. Transfer methods can fail silently when the target domain differs not only in covariates but also in degradation mechanism; therefore, physical consistency checks should accompany target-domain accuracy metrics [55–61].Table 7 summarizes representative criteria that may be used to evaluate the physical consistency of transferred representations. These criteria are intended as practical diagnostic checks that complement conventional performance metrics. Their common purpose is to reduce the risk that apparently successful transfer learning models achieve predictive accuracy through spurious correlations that do not remain valid under changing operating conditions, chemistries, or deployment environments.

Table 7. Practical checks for physical consistency of transferred representations.

Check	Practical Metric or Procedure	Interpretation
Health association	Rank correlation or mutual information with capacity, resistance, power capability, or validated diagnostic features	Transferred features should remain linked to degradation-relevant quantities
Context disentanglement	Residual correlation or regression of latent features against temperature, C-rate, SOC window, or protocol after conditioning on health	Low residual dependence suggests the representation is not merely encoding operating context
Agingtrend consistency	Monotonicity violation rate or trend agreement across aging sequence	Useful for checking whether features evolve plausibly with degradation
Diagnosticsignature preservation	Agreement with ICA/DVA peaks, EIS/DRT descriptors, relaxation parameters, or thermal indicators when available	Supports mechanistic plausibility of the transferred representation
Counterfactual robustness	Perturb temperature, current, SOC window, or sensing window and inspect whether SOH/RUL changes remain physically plausible	Detects shortcut learning and implausible sensitivity
Target-domain holdout validation	Apply the above checks on unseen chemistry, protocol, temperature, pack, or field data	Confirms that interpretability survives transfer rather than only source-domain training

To make the physical interpretability of transferred representations auditable, transfer learning and domain adaptation studies should report not only target-domain prediction error but also physical consistency checks on the learned features or latent variables.

In this review, a transferred representation is considered physically consistent only if it satisfies three conditions: it remains associated with degradation-relevant health quantities, it is not dominated by the reversible operating context, and its response to controlled perturbations agrees with known battery behavior.

Several practical checks can be used. First, the transferred features should show a stable rank correlation or mutual information with independently measured health indicators such as capacity, resistance, power capability, or diagnostic features derived from ICA/DVA, EIS/DRT, relaxation voltage, or thermal response. Second, their residual dependence on temperature, C-rate, state of charge (SOC) window, or protocol should be tested after conditioning on health. A representation that mainly encodes the usage context rather than degradation should not be treated as physically interpretable. Third, monotonicity or trend consistency checks should be applied where the target quantity is expected to evolve gradually with aging. Fourth, counterfactual or perturbation tests should verify that small changes in operating conditions do not produce physically implausible jumps in SOH/RUL, while changes in degradation-sensitive inputs produce directionally plausible responses. Finally, transferred representations should be evaluated under external target-domain holdouts to determine whether physical consistency is preserved outside the training set domain.

These checks provide objective evidence that the transferred representation is aligned with battery-relevant degradation behavior and is not merely exploiting protocol-specific or dataset-specific correlations.

Physics-informed and hybrid estimators currently offer the most convincing compromise between accuracy, data efficiency, and extrapolation. These models use electrochemical constraints, monotonicity priors, latent degradation laws, or equivalent-circuit dynamics to regularize learning. Their primary utility stems not from ideological purity but from their structural bias; they significantly reduce the chance that a model will exploit purely incidental correlations. When successful, they improve data efficiency, promote smoother extrapolation, and simplify data interpretation and explanation. The main challenge to their utilization lies in balancing realism with tractability and calibration [27–34,62,63].

Digital twins should be viewed not as an additional estimator class but as a systems architecture for integrating models, data, uncertainty, and operations. A health-aware twin can host online estimators, periodic parameter identification routines, thermal submodels, maintenance logs, and validation artifacts inside a singular lifecycle-aware stack. The opportunity is substantial: digital twins can support cross-fidelity transfer, fleet shadow evaluation, anomaly triage, and what-if maintenance planning. The risk is also equally real; the phrase 'digital twin' can become an empty branding shell unless it is continually subjected to model updating, data lineage, and rigorously specified and tailored acceptance criteria [8–10,64–67].

The following comparative tables use the evidence-weighted synthesis procedure described in Section 2.3.

Table 8 shows that estimator families differ not only in predictive accuracy but also in the type of variability they are structurally able to absorb. Table 8 should be interpreted as a comparison of methodological roles under realistic variability rather than as a list of mutually exclusive algorithm classes. The first five rows correspond primarily to estimator families, whereas digital twins represent system-level integration architectures and trustworthy uncertainty quantification (UQ) methods represent cross-cutting reliability layers. Purely model-based methods retain value because of their physical consistency, transparency, and low online computational cost, yet they become less reliable when parameter identifiability degrades under sparse sensing or when aging trajectories become strongly heterogeneous. Feature-based machine learning methods remain attractive when stable and informative descriptors can be extracted, but their robustness is often conditional on protocol alignment and preprocessing consistency. Deep sequence models offer greater representational flexibility and can learn directly from current–voltage–temperature histories, yet their apparent performance is highly sensitive to training and testing design and dataset representativeness. Transfer learning and domain adaptation approaches are especially relevant under chemistry, manufacturer, and platform shifts, but their practical credibility depends on whether the transferred representation remains physically meaningful rather than merely statistically aligned. Overall, Table 8 supports a central conclusion of this review: under realistic variability, the decisive issue is not which estimator family performs best on average, but which family preserves reliability, interpretability, and calibration when evidence is transferred from controlled laboratory settings to deployment-oriented battery management contexts.

Table 8. Comparative assessment of estimator families, system architectures, and trustworthiness layers.

Estimator Family	Temp/Rate Variability	Cross-Chemistry Transfer	Sparse Windows	Pack-to-Field Scaling	Counterfactual Reasoning	Calibrated Risk Estimates	Critical Review Note
Pure model-based	3	2	2	3	4	2	Strong extrapolation logic, but parameter identifiability degrades under sparse field sensing.
Feature ML	3	3	2	2	2	2	Competitive on curated datasets; most brittle to protocol shift and leakage.
Deep sequence models	4	3	2	2	2	3	Good for complex current/voltage histories, but data-hungry and often weakly interpretable.
Transfer/domain adaptation	4	4	4	3	2	3	Key enabler for chemistry/protocol transfer if assumptions are explicitly tested.
Physics-informed/hybrid	4	4	3	4	4	4	Best compromise between extrapolation, data efficiency, and interpretability.
Digital twin	5	4	4	5	5	4	Useful system architecture for model fusion, validation traceability, and lifecycle updating.
Trustworthy/UQ layer	3	3	4	5	2	5	Not a standalone estimator; essential overlay for calibrated deployment decisions.

Score interpretation: 1 = very low; 2 = low; 3 = moderate; 4 = high; 5 = very high.

The practical value of these estimator families depends on the diagnostic signals available to them. Therefore, Section 5 complements the algorithm-level comparison by examining the sensing and feature-space assumptions under which different estimator families are most likely to be reliable, transferable, and BMS-feasible.

5. Observability and Diagnostic Modalities

The estimator families discussed in Section 4 and the diagnostic modalities summarized in Section 5 should not be selected independently. In practical BMS development, the most suitable algorithm depends strongly on the available signal type, sensing frequency, excitation quality, operating variability, and required decision output. Table 9 therefore links the main diagnostic modalities to estimator families that are most compatible with their information structure and deployment constraints.

Table 9. Recommended pairing between diagnostic modalities and estimator families for SOH/RUL estimation.

Diagnostic Signal or Feature Space	Most Suitable Estimator Families	Practical Rationale for BMS Use
Raw current–voltage–temperature trajectories	Deep sequence models, hybrid observers, uncertainty-aware models	Suitable for online BMS data streams, but requires context handling to avoid learning usage patterns instead of degradation
Partial-charge or short-window features	Feature ML, transfer learning, Gaussian-process or Bayesian models, conformal prediction	Practical for EV and BESS logs where full cycles are unavailable; uncertainty reporting is important because observability is incomplete
ICA/DVA-derived features	Feature ML, model-data fusion, physics-informed learning	Mechanistically informative under controlled conditions, but requires careful preprocessing and validation under rate, temperature, and window variation
Relaxation-voltage descriptors	Feature ML, CNN/sequence models, transfer learning, hybrid models	Useful for opportunistic diagnostics during rest or charging interruptions; sensitive to prior current and thermal history
EIS/DRT descriptors	Model-based estimation, Gaussian-process regression, interpretable ML, hybrid physics-informed models	Mechanistically rich and valuable for calibration or service-bay diagnostics, but less suitable for continuous embedded operation
Thermal or thermographic signals	Multimodal fusion, anomaly detection, graph or spatial–temporal models	Useful for heterogeneity, cooling, imbalance, and safety-related monitoring; usually requires fusion with electrical data
Multimodal diagnostic stacks	Physics-informed hybrid models, digital twins, transfer learning, uncertainty-aware fusion	Most appropriate when deployment requires robustness, calibration, and traceability across cell, module, pack, and field levels

The matrix should be interpreted as a practical matching guide rather than as a strict ranking. Online signals, such as current, voltage, temperature, and partial-charge windows, favor embedded, uncertainty-aware, and transfer-capable estimators. Mechanistically richer but less continuously available modalities—such as EIS/DRT, ICA/DVA, relaxation tests, and thermal imaging—are especially useful as calibration anchors, validation references, or digital-twin synchronization events. Consequently, the most deployment-relevant architecture is often not a single estimator applied to a single signal, but a layered system in which online electrical estimation is periodically corrected or validated using richer diagnostic modalities.

The pairings in Table 9 are based on the information structure of each diagnostic signal rather than on algorithm popularity alone. Dense current–voltage–temperature trajectories contain temporal, thermal, and load-history information; therefore, deep sequence models, hybrid observers, and uncertainty-aware estimators are appropriate when the training data cover realistic operating variability. Their main advantage is direct compatibility with embedded BMS data streams, while their main limitation is shortcut learning, where the model may encode the usage pattern, charger behavior, or temperature history instead of irreversible degradation. Partial-charge and short-window features are more intermittent and lower-dimensional; they are therefore better matched with feature-based ML, Gaussian-process or Bayesian regression, transfer learning, and conformal uncertainty layers, because these methods can explicitly handle incomplete observability and uncertainty in the unobserved part of the cycle. ICA/DVA-derived features should be paired with feature ML, model-data fusion, or physics-informed learning because they retain a clearer electrochemical interpretation, but they become fragile under a high current, coarse sampling, temperature variation, and incomplete voltage windows. EIS/DRT, relaxation, thermal, and multimodal signals should not be viewed only as alternative inputs to the same estimator; they play different BMS roles. EIS/DRT and thermal data are most useful as service-bay, calibration, or digital-twin synchronization signals, while raw electrical and partial-charge signatures remain the most practical online sources. Thus, for BMS developers, the recommended combination is the simplest estimator that matches the available signal quality, preserves physically meaningful aging information, and satisfies uncertainty, computational, and domain-shift constraints.

Raw electrical data remain the lowest-friction information source because every BMS already measures current, terminal voltage, and at least one temperature channel. This makes curve-based learning attractive for embedded health estimation. The limitation is that voltage and temperature are multiplexed observables, meaning that they reflect both degradation and the current operating context. Without carefully designed training distributions or context-aware models, curve-based estimators can overfit towards usage style rather than health observables [68–71].

ICA and DVA remain among the most informative laboratory-friendly diagnostic tools because they sharpen phase-transition signatures and connect the observed curve morphology to degradation mechanisms. In well-controlled environments, they can reveal the loss of lithium inventory, active material loss, and resistance-related distortion more transparently than raw data curves. Their weakness lies in their fragility under noise, partial windows, rate variation, and coarse sampling. This is why transfer of ICA/DVA logic from laboratory cells to in-service packs is scientifically attractive but technically difficult [49,68,72–75].

Relaxation-voltage methods have progressed rapidly because they offer a compromise between laboratory richness and operational feasibility. Short relaxation windows encode information about internal diffusion, polarization effects, and open-circuit-voltage recovery while requiring less instrumentation overhead than EIS. They are especially promising for fast diagnostics during short, reduced-consumption periods or charging interruptions. Their main open question is robustness under temperature variation and irregular preceding current histories [75–80].

Partial-charge and small-window approaches are now central to realistic applications because complete charge–discharge cycles are uncommon in field service data. Estimators that depend on full curves therefore enjoy a strong but often unrealistic observability assumption. Recent work shows that useful SOH information can still be extracted from partial windows, but only if the model accounts for path dependence, temperature, and uncertainty in the unobserved parts of the cycle [61,69,70].

Partial-charge methods are therefore promising, but their applicability is bounded by the information content of the available window. The minimum useful window length is not universal as it depends on chemistry, SOC region, sampling rate, temperature stability, current excitation, sensor resolution, and the target being estimated. A short window located in a flat open-circuit voltage (OCV) region or under highly variable current may contain an insufficient degradation-sensitive structure, whereas a similarly short window crossing a voltage plateau, an ICA/DVA-sensitive region, or a repeatable charging segment may be informative. Consequently, partial-charge estimators should report the observation-window definition explicitly, including SOC span, time duration, sampling rate, current profile, temperature range, and preprocessing assumptions.

A practical validation procedure is to perform window-ablation testing, in which the available SOC or time window is progressively shortened and shifted across the charge curve. The minimum usable window can then be defined as the shortest window for which the estimator still satisfies predefined error, calibration, and robustness thresholds. Measurement error should also be tested explicitly by injecting realistic voltage, current, temperature, timing, and SOC-alignment noise into the validation data. This is especially important for ICA/DVA-based reconstruction because derivative-based features amplify measurement noise, which may create artificial peak shifts or suppress degradation signatures. Under very short, noisy, or non-repeatable partial windows, the output should therefore be treated as a high-uncertainty estimate or used only for screening, drift detection, or triggering a richer diagnostic procedure.

The partial-charge literature does not support a universal minimum window because the useful information depends on chemistry, voltage/SOC region, current rate, temperature, sampling quality, and whether the target is SOH regression, failure classification, or second-life sorting. However, several practical design anchors can be extracted. For voltage-window methods, 0.1 V windows placed near IC-sensitive regions have been demonstrated across NCA, LCO, LCO/NCO, and LFP cells, while broader Ah-counting or optimized NMC windows such as 3.55–3.80 V or 3.127–4.027 V trade a higher information content against lower availability during real charging [81–83]. For time-window methods, arbitrary 400 s charging fragments at 1 C and sequence models trained under realistic sampling/noise conditions provide a practical lower bound for short-window deep learning approaches [82,84]. In terms of measurement quality, comparative evidence suggests that robust partial-curve estimators can remain below about 2% average MAPE under 3 mV voltage noise or 60 s sampling intervals, while SOC-based windowing can incur capacity errors of about 5.16% of nominal capacity under ±5% SOC uncertainty [81,82]. These values should be interpreted as indicative starting points for BMS design, not universal thresholds—each implementation should verify them by window ablation, noise injection, temperature-stratified validation, and chemistry-specific holdout testing.

EIS and DRT-based methods remain unmatched in mechanistic richness. By resolving the frequency-dependent response, they provide insights into charge-transfer resistance, diffusion processes, contact changes, and interfacial evolution that are difficult to infer uniquely from time-domain trajectories alone. Their primary weakness is deployment friction: excitation design, hardware overhead, time cost, and sensitivity to operating conditions complicate continuous in-field use. For this reason, EIS is best viewed not as an always-on flag but as a periodic calibration anchor within a broader health estimation stack [85–97]. Table 10 shows a quantitative assessment of partial-charge estimation of SOH.

Table 10. Quantitative anchors for partial-charge SOH estimation.

BMS Situation	Quantitative Anchor from Partial-Charge Studies	Recommended Estimator Logic and Deployment Caution
Repeatable voltage window during CC charging	Use chemistry-specific windows near aging-sensitive regions. Reported anchors include 0.1 V windows: 3.70–3.80 V for NCA, 3.90–4.00 V for LCO, 3.75–3.85 V for LCO/NCO, and 3.35–3.45 V for LFP [82]. Broader NMC windows such as 3.55–3.80 V and 3.127–4.027 V have also been reported [81,83].	EVTS/ETVS sequences, Ah-counting, GPR/SVR/XGBoost/LSTM, or simple regression is suitable. The window must be reselected for chemistry, charger protocol, current rate, and temperature; broader windows improve observability but reduce real-use availability.
Short, random, or SOC-defined charging fragment	Arbitrary 400 s fragments at 1 C supported maximum- and remaining-capacity estimation with RMSE below 12.68mAh and 11.71 mAh for 0.74 Ah cells [84]. IES features from 20%, 40%, 60%, and 80% SOC starting points achieved MAE below 0.75% and RMSE below 0.8% [96]. Random capacity-increment features with sparse GPR gave maximum errors of 2.88%, 2.52%, and 1.51% for three chemistries [97].	CNN/LSTM/BiLSTM, IES-based models, and sparse GPR are suitable when charging start/end points are uncontrolled. The main risk is dependence on representative training data and accurate SOC, current, and energy integration.
Derivative or IC-based partial features	Partial IC features can support failure recognition around the 80% capacity threshold, with elastic-net feature selection and LDA used to balance precision and recall [98]. Under heavily partial high-voltage charging, IC features may disappear; voltage entropy and multistage feature fusion are therefore needed [99].	Use ICA/DVA features when the window contains an identifiable aging-sensitive peak or transition. Add smoothing, feature-selection, and non-derivative backup features when operating at high voltage, high current, or coarse sampling.
Noisy or low-rate BMS sampling	Comparative partial-curve evidence reports LSTM MAPE of 0.61%, maximum error of 3.94%, and average MAPE below 2% even with 3 mV voltage noise or 60 s sampling intervals [81]. SOC-window methods can show errors up to 5.16% of nominal capacity under ±5% SOC uncertainty [81].	Sequence models and statistically robust features are preferable. BMS implementation should include noise-injection, sampling-rate ablation, and SOC-error sensitivity tests before claiming transferability.
Low-cost screening, sorting, or diagnostic charging policy	Five features from one partial segment reduced retired-cell sorting detection time to 19.96% and energy use to 20.35% of conventional testing [100]. Optimized partial charging can reduce high-SOC residence and was simulated to extend usable life by up to 150% [101,102].	SOM/subtractive clustering, LDA, or other lightweight classifiers are appropriate for grouping/screening, not precise RUL. Diagnostic windows should be collected opportunistically; the BMS should not force high-SOC exposure only to obtain SOH features.

Thermal and thermographic approaches contribute information that electrical observables alone cannot provide, such as spatial heterogeneity, localized defects, tab imbalance, and cooling non-uniformity. At cell and module levels, infrared imaging can reveal degradation asymmetry and precursors to safety-critical behavior. At the pack level, however, sensing density, occlusion, and cost become decisive constraints. Thermal information is therefore most valuable when fused with electrical features or used selectively during diagnostic events [72–75].

Across modalities, the key lesson is that the best signal depends on where in the lifecycle the estimator operates. EIS/DRT and thermography may dominate observability in laboratory or service-bay contexts, whereas raw electrical curves, short-window charging signatures, and sparse temperature channels dominate online deployment. Robust SOH/RUL estimation is therefore not a winner-takes-all competition among modalities but an observability allocation problem across fidelity levels [72–80,85–95].

To complement the temporal overview of the literature and integrate the trade-off perspective within the modality analysis, Figure 3 presents a composite assessment of the principal sensing and feature-space families used for SOH/RUL estimation. Figure 3a compares modalities in terms of aging observability, online deployability, instrumentation burden, interpretability, and robustness under realistic operating variability, while Figure 3b maps their trade-off between diagnostic information richness and deployment readiness.

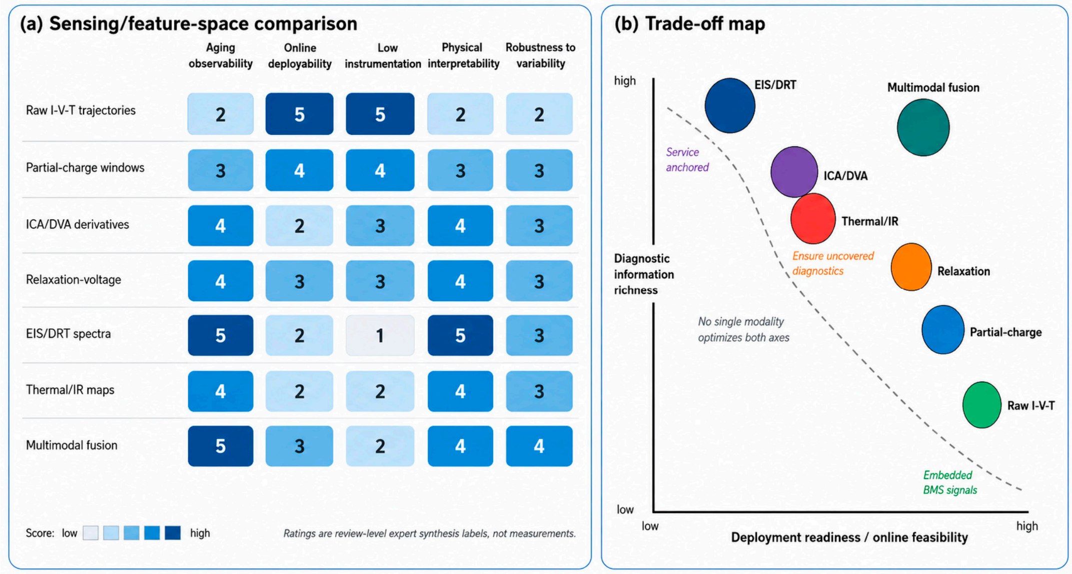

Figure 3. Composite sensing/feature-space assessment and deployment trade-off for robust SOH/RUL estimation. (**a**) Review-level comparison across observability, deployability, instrumentation burden, interpretability, and robustness under realistic variability. (**b**) Trade-off between diagnostic information richness and deployment readiness across diagnostic families.

Figure 3 shows that no single modality simultaneously optimizes diagnostic information, implementation feasibility, and robustness to domain shift. Raw electrical curves are highly deployable but vulnerable to hidden confounding; EIS/DRT, ICA/DVA, and thermal/infrared (IR) diagnostics are richer mechanistically but more difficult to deploy continuously. Relaxation-voltage, partial-charge, and multimodal approaches occupy the strategic middle ground for practical BMS use.

To complement the visual synthesis in Figure 3, Table 11 provides a structured comparison of the main sensing and feature-space families considered in this review, together with their diagnostic strengths, deployment limitations, and relevance for robust SOH/RUL estimation.

Table 11. Comparative assessment of sensing and feature-space families.

Modality Family	Observability	Online Deployability	Instrumentation Burden	Interpretability	Robustness to Duty-Cycle Shift	Critical Review Note
Raw electrical trajectories	4	5	1	3	2	Best online deployability, but vulnerable to protocol leakage and hidden confounding.
ICA/DVA	5	2	4	4	3	Strong aging observability; limited by signal quality, rate dependence, and partial-window constraints.
Relaxation voltage	4	4	2	3	4	Useful when only tens of seconds of rest are available; sensitive to thermal history.
Partial-charge windows	4	3	2	4	4	Practical in EV and BESS logs if models handle truncation explicitly.
Thermal/thermography	3	2	3	4	3	Reveals heterogeneity and safety-relevant gradients, but instrumentation burden is nontrivial.
EIS/DRT	5	2	5	4	3	Mechanistically rich and transferable; hardest to deploy continuously in products.
Hybrid multimodal fusion	4	3	4	5	5	Most promising route when sparse modalities are fused with uncertainty-aware architecture.

Score interpretation: 1 = very low; 2 = low; 3 = moderate; 4 = high; 5 = very high.

Table 11 highlights that the practical utility of a sensing modality depends on how it balances diagnostic richness against measurement realism. Electrical curve data analysis remains attractive because it is directly available in standard BMS data streams; however, its health sensitivity is often entangled with operating context and usage history. More structured features derived from ICA/DVA, relaxation segments, partial charging windows, thermal fields, or impedance spectra can reveal degradation more explicitly, but they also introduce stricter requirements for signal quality, sensing hardware, excitation design, or preprocessing consistency. The table therefore supports a key conclusion of this review: evidence suggests that robust SOH/RUL estimation is unlikely to depend on a single, universally dominant sensing modality. Instead, multimodal, uncertainty-aware, and validation-aware frameworks appear to be a promising direction, particularly when estimators must operate under realistic sensing constraints, partial observability, and changing deployment conditions [61,69,70,72–80,85–95].

6. Trustworthiness, Transferability, and Deployment Challenges

6.1. Sources of Domain Shift in Battery Prognostics

A major obstacle to reliable battery prognostics is the discrepancy between laboratory conditions and real-world operation. Models trained or calibrated using laboratory datasets often assume controlled temperature ranges, predefined cycling protocols, and relatively homogeneous cell populations. In practical deployments, however, batteries experience varying duty cycles, environmental conditions, charging behaviors, and manufacturing variability. These factors introduce domain shifts that may significantly degrade estimation performance. Cross-chemistry and cross-manufacturer transfer is more challenging than cross-protocol transfer because the latent degradation mechanisms themselves can change. Graphite/NMC, LFP/graphite, silicon-containing blends, and highnickel chemistries exhibit different OCV behaviors, resistance growth pathways, and temperature sensitivities. Transfer is still possible, but it typically requires either physically anchored representations or explicit domain adaptation [56–60,91,92].

Not all domain shifts are equivalent. Temperature and C-rate variations often induce covariate shifts in observable data, whereas chemistry, manufacturer, and electrode-design changes may induce mechanism shifts that alter the mapping between diagnostic features and degradation states. This distinction is essential because statistical feature alignment may be sufficient for some covariate shifts but insufficient when the underlying degradation pathway changes.

Temperature variability and charge/discharge conditions remain the most persistent sources of performance degradation. Many apparent SOH features are not health-specific but health-plus-context mixtures. Peak spacing, voltage hysteresis, relaxation curve shaping, and impedance arcs all shift with temperature and current. Models that do not explicitly encode context or learn invariances can therefore misinterpret reversible operating effects as irreversible aging. This is one reason why cross-condition validation is more informative than pooled random splitting [60,69–71,89].

Recent studies have therefore increasingly emphasized transferability, uncertainty awareness, and robustness as essential requirements for deployment-oriented SOH and RUL estimation. Rather than optimizing performance exclusively for a specific dataset, future prognostic frameworks must maintain acceptable reliability across heterogeneous battery populations and operating environments.

The challenge becomes particularly pronounced when moving from laboratory cells to battery modules and packs. At higher system levels, cell-to-cell variability, balancing strategies, thermal gradients, and measurement aggregation introduce additional uncertainties that are often absent in cell-level studies. Consequently, deployment-oriented battery prognostics should be evaluated not only in terms of prediction accuracy but also in terms of robustness under domain variability.

6.2. Practical Observability Constraints

A fundamental limitation of battery prognostics is that many degradation processes are only partially observable through standard BMS measurements. While voltage, current, and temperature are directly measurable, the underlying electrochemical degradation states remain latent and must be inferred indirectly through models or data-driven estimators.

Consequently, estimation accuracy is not determined solely by algorithmic sophistication but also by the extent to which the target degradation mechanism is observable from the available sensing information. This issue becomes increasingly important when transitioning from laboratory cells to modules and full battery packs, where measurement aggregation, cell-to-cell variability, and limited sensor availability further reduce observability.

The relative observability presented in Table 12 underlines the critical difference between physical compatibility and practical utility. Although mechanisms such as lithium inventory loss, active material loss, and lithium plating play a central role in battery degradation, their signatures are often weak, indirect, or confounded by other processes when only conventional BMS measurements are available. Consequently, estimation uncertainty generally increases as prognostic models attempt to infer progressively deeper electrochemical states. This limitation becomes more pronounced at module and pack levels, where aggregated measurements may obscure cell-level degradation heterogeneity.

Table 12. Relative observability of representative degradation states.

Latent State	Cell	Module	Pack
Capacity fade	High	Medium	Medium
Resistance growth	High	Medium	Low
Lithium inventory loss	Medium	Low	Low
Active material loss	Low	Low	Very low
Lithium plating	Low	Very low	Not directly observable

Furthermore, interpretability requires a more rigorous standard. Feature importance plots and saliency maps are informative only if the highlighted structures correspond to plausible battery behavior rather than to confounded artifacts of protocol or preprocessing. For *Batteries*, interpretable models are not a public relations accessory; they represent a methodology to verify whether the estimator responds to known signatures such as peak migration, relaxation kinetics, or impedance growth rather than to dataset idiosyncrasies [14,27–34,85–95].

Finally, label scarcity and temporal drift reshape the deployment challenge. In real-world fleets, capacity measurements are rare, maintenance events are selective, and user behavior drifts over time. Estimators must therefore operate under weak, delayed or proxy labels, while simultaneously monitoring whether the data distribution itself has shifted. Under these conditions, strategies such as shadow deployment, continual calibration checks, and selective abstention become essential [12,13,36,37,39,40].

Consequently, trustworthy SOH and RUL estimation should be viewed not only as a modeling challenge but also as an observability-constrained inference problem. In practical deployments, improving observability may yield larger gains in prognostic reliability than increasing model complexity alone.

6.3. Failure Modes Under Domain Shift

The presence of domain shift affects various prognostic methodologies in different ways. While certain approaches are primarily limited by model mismatch, others exhibit greater vulnerability to data distribution changes or sensor-related uncertainties.

Figure 4 shows a mapping that links dominant sources of deployment shift and which emphasizes that acceptable SOH/RUL deployment evidence requires not only low error but also context-aware validation.

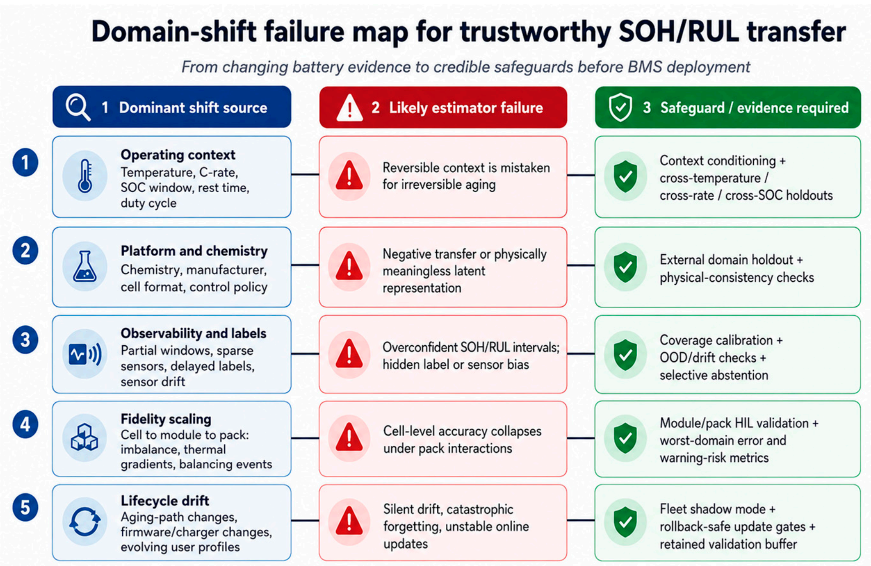

Figure 4. Domain-shift failure map for trustworthy SOH/RUL transfer.

https://doi.org/10.3390/batteries12070255

A critical implication of Table 13 is that method families should not be compared solely by average reported accuracy; their practical value depends heavily on the specific type of domain shift under consideration. Model-based and physics-informed approaches are generally more defensible under moderate extrapolation when their underlying assumptions remain valid. However, they can fail when parameters become unidentifiable or degradation mechanisms are incomplete. Conversely, feature-based and deep learning approaches may achieve a superior benchmark performance, but they exhibit higher exposure to protocol leakage, partial-window mismatch, and shortcut learning. While transfer learning approaches are useful when source and target domains share a degradation structure, they may fail under mechanism shift. Digital twins and uncertainty-aware layers should therefore be treated as deployment-support structures rather than guarantees of robustness. This distinction is central to a critical interpretation of the reviewed literature.

Transfer learning offers a viable mechanism for improving cross-domain generalization, but it may introduce negative transfer when source and target domains differ substantially. Similarly, digital twins may accumulate systematic errors if calibration procedures are not continuously updated using field evidence.

Understanding these failure mechanisms is essential for designing trustworthy battery prognostic systems because reliability depends not only on average prediction accuracy but also on predictable behavior under adverse operating conditions.

Table 13. Method effectiveness and likely failure modes under representative domain shifts.

Methodological Role	Most Effective When	Likely Failure Conditions Under Domain Shift
Equivalent-circuit and reduced-order model-based estimators	Cell chemistry, temperature range, and operating regime are known; online computation must be lightweight	Parameter drift, chemistry change, poor excitation, sparse sensing, or un-modeled degradation pathways
Electrochemical/physics-based models	Degradation mechanisms are known and parameters can be identified or constrained	Missing degradation mechanisms, uncertain parameters, high computational burden, or pack-level heterogeneity
Feature-based ML	Diagnostic features are stable across protocols, preprocessing, and SOC/temperature windows	Protocol shift, partial-window mismatch, rate/temperature variation, feature leakage, or preprocessing inconsistency
Deep sequence models	Large and representative current–voltage–temperature histories are available	Small datasets, shortcut learning, train/test leakage, unseen duty cycles, sparse labels, or weak interpretability
Transfer learning/domain adaptation	Source and target domains share degradation structure and only moderate covariate shift is present	Mechanism shift, chemistry change, manufacturer differences, negative transfer, or statistically aligned but physically meaningless features
Physics-informed/hybrid models	Useful physical priors exist but data-driven flexibility is still needed	Incorrect or incomplete physics constraints, over-regularization, poor calibration, or unvalidated extrapolation
Multimodal fusion	Complementary signals are available and synchronized with sufficient quality	Missing modalities, sensor disagreement, inconsistent sampling, high instrumentation burden, or conflicting diagnostic evidence
Digital-twin architectures	Models, data streams, uncertainty estimates, validation artifacts, and feedback loops are continuously maintained	Outdated calibration, poor data lineage, missing field feedback, model drift, or treating the twin as a static model

Uncertainty-aware/trustworthy-AI layers	Prediction intervals, OOD detection, calibration, and abstention are validated under deployment-like shifts	Mis-calibration under unseen domains, overconfident intervals, untested warning thresholds, or unreliable OOD detection

6.4. Continual Learning, Update Governance, and Lifecycle Adaptation

Sparse and partial measurements are not a niche issue; they represent the default field condition. Pack-level systems rarely expose cell-level full curves, and maintenance calibration labels may become available after months of delay. This reality favors architectures that can update themselves under delayed supervision, heterogeneous metadata, and event-triggered diagnostics rather than only under dense, synchronized labels [61,69–71].

Battery degradation is inherently non-stationary. The operational environment of deployed systems evolves throughout the battery lifetime because of seasonal temperature changes, changing charge behavior, firmware updates, cell imbalance, sensor drift, and gradual shifts in user or grid-service profiles. Consequently, a static estimator that is well calibrated at deployment may gradually lose accuracy and become overconfident as the data distribution moves away from the original validation domain.

Continual learning is therefore a promising strategy for lifecycle battery-health estimation, but it should not be understood as unrestricted online retraining. Uncontrolled model updates may lead to catastrophic forgetting, loss of calibration, or unexpected safety-related behavior. For battery management applications, adaptation mechanisms must conceptualize and preserve previously acquired knowledge while incorporating new information in a controlled manner. Therefore, continual learning must be governed by explicit update rules, drift detection, validation gates, and rollback mechanisms.

Pack-to-field scaling introduces another level of difficulty because thermal gradients, cell imbalance, cell balancing approaches, and contact resistances create interactions that are absent or hidden at a cell level. Consequently, a health estimator that is well calibrated on isolated cells can become overconfident at the pack level even if its average error remains acceptable. This mismatch is precisely why pack-aware transfer studies and field-facing diagnostics are so critical [55,61,72–75].

A practical lifecycle-adaptation workflow can be organized into four stages. First, the deployed estimator should monitor distribution shift using residual trends, novelty scores, disagreement between model families, calibration drift, or changes in operating data. Second, when a shift is detected, the system should enter a shadow update mode in which candidate model updates are evaluated passively on new data without influencing BMS control decisions. Third, updates should be accepted only if they improve or preserve the worst-domain error, calibration, warning-risk behavior, and computational feasibility. This evaluation must be conducted over a validation buffer that includes both recent field data and retained historical cases. Fourth, if an update degrades performance or produces unstable uncertainty estimates, the BMS should retain or restore the previously validated model.

Continual learning can also be implemented at varying levels of aggressiveness. Conservative adaptations may recalibrate uncertainty intervals or correction factors without altering the primary estimator. Moderate adaptations may update domain-specific normalization, transfer layers, or battery-population priors. More aggressive adaptations may retrain feature extractors or sequence models, but this requires more rigorous safeguards and external validation. For safety-relevant SOH/RUL estimation, conservative or modular adaptation is often preferable to unrestricted end-to-end online learning.

Thus, continual learning should be treated as a controlled life cycle management process rather than as automatic retraining. Its purpose is not simply to improve average prediction error over time, but to preserve calibrated, auditable, and rollback-safe SOH/RUL estimation as batteries move through changing operating conditions, pack states, and deployment environments.

6.5. Toward Trustworthy Deployment

The preceding discussion highlights that trustworthy battery prognostics cannot be achieved through prediction accuracy alone. Reliable deployment requires simultaneous consideration of observability, uncertainty quantification, transferability, robustness, validation, and lifecycle adaptation.

Trustworthiness begins with calibration. For BMS decisions, an estimator should provide not only a mean SOH or RUL value but also an uncertainty measure whose frequency properties remain acceptable under domain shift. A narrow but miscalibrated interval is more dangerous than a wider but honest one because maintenance and derating decisions will become overconfident. Calibration should therefore be assessed by coverage, interval width, and decision-conditioned loss rather than by confidence graphics alone [36–40]. In deployment-oriented settings, calibrated uncertainty should also support selective prediction and abstention: the estimator must be allowed to defer, widen uncertainty bounds, or fall back to conservative rules when the operating condition is outside its validated domain.

Accordingly, future battery prognostic frameworks are expected to integrate multimodal sensing, physics-informed learning, uncertainty-aware estimation, and digital-twin-based validation architectures. These components should operate within a closed-loop evidence framework where field observations continuously support model calibration, performance monitoring, and risk-aware decision-making.

The multi-fidelity validation framework proposed in this review provides one possible pathway toward this objective. By explicitly linking laboratory evidence, model validation, uncertainty assessment, and deployment monitoring, the framework aims to improve confidence in SOH and RUL estimation across different battery technologies and operational conditions.

Ultimately, trustworthy deployment should be regarded as a system-level property rather than a model-level attribute. Accurate prediction remains necessary, but long-term reliability depends equally on transparency, validation rigor, uncertainty awareness, and the ability to maintain performance under evolving real-world conditions.

Translating battery prognostic models from laboratory demonstrations to real-world deployment requires a broader validation perspective than prediction accuracy alone. Trustworthy deployment demands evidence that models remain reliable under domain variability, maintain calibrated uncertainty estimates, and preserve performance across diverse operating conditions. Based on the findings synthesized throughout this review, Table 14 proposes a practical validation checklist to assist researchers and practitioners in evaluating the readiness of SOH and RUL estimation approaches for deployment-oriented applications.

Table 14. Trustworthy validation checklist for deployment-oriented SOH/RUL claims.

Validation Dimension	Minimum Evidence for a Robust Claim	Common Failure Mode
Data partitioning	Blind splits by cell, batch, protocol, and time	Train/test leakage through adjacent cycles or same-batch duplication
Robustness	Cross-temperature, cross-rate, cross-SOC-window evaluation	Single-protocol accuracy only
Transferability	External chemistry/manufacturer/domain hold-out	In-domain random split only

Uncertainty	Coverage error, calibration curve, risk-conditioned decision thresholds	Point estimate without confidence or abstention
OOD/drift	Shift detector, drift dashboard, fallback or rollback logic	Silent extrapolation after deployment
Interpretability	Mechanistic plausibility, feature stability, counterfactual checks	Post-hoc saliency only
Pack realism	Sparse-sensor tests, imbalance, thermal gradients, latency, balancing events	Cell-only evaluation
Lifecycle evidence	Shadow deployment and closed-loop acceptance criteria	Benchmark-only claims of deployability

Random splits over cycles from the same cells, shared protocols across training and testing procedures, or preprocessing decisions tuned on the full dataset can produce visually impressive but scientifically weak results. The field now has reached sufficient maturity to demand more rigorous standards. Dataset splits should be agnostic to cell, batch, protocol, and time; external holdouts should be preferred whenever possible; and claims of deployment readiness should be reserved for studies that include at least some pack-level or fleet-shadow evidence [1,7,8,11,50].

OOD detection and drift awareness are equally important. Many product-level failures will not arise because the estimator is inaccurate on familiar data, but because the system enters a regime the model has never encountered: a new climate, an alternative charger type, a firmware change, sensor bias, or an atypical duty cycle. Trustworthy deployment therefore requires gating logic, including novelty detection, disagreement monitors across model families, residual trend alarms, and rollback-safe fallbacks to conservative rules [8–11,36,37,39,40].

Table 14 makes explicit that trustworthy SOH/RUL estimation is not established by predictive performance alone, but by the quality and completeness of the supporting validation evidence. The checklist highlights that a credible claim must specify the health target, sensing assumptions, operating domain, fidelity level, train–test separation logic, uncertainty treatment, and acceptance criteria relevant to the intended decision context. This is particularly critical because many published studies still report strong numerical performance without clarifying whether the model was tested under chemistry shift, sparse sensing, delayed labels, pack-level interaction, or OOD conditions. By organizing these requirements into a structured validation checklist, Table 14 reframes SOH/RUL assessment as an evidence-based acceptance process rather than a benchmark-reporting exercise. In doing so, it supports one of the central arguments of this review—an estimator becomes trustworthy not when it achieves the lowest average error, but when its assumptions, limitations, uncertainty behavior, and deployment boundaries are all made explicit and validated against the operating conditions in which it is expected to function.

The proposed checklist's primary purpose is to encourage more comprehensive reporting and validation practices within the battery prognostics community. Future standardization efforts may further formalize such criteria—systematic consideration of transferability, uncertainty, robustness, and validation evidence already represents an important step toward trustworthy deployment.

6.6. Section Synthesis

This section addresses domain-shift sources, observability constraints, failure modes, continual-learning governance, and trustworthy deployment. The discussion clarifies that the minimum validation evidence for a trustworthy SOH/RUL estimator cannot be defined by prediction accuracy alone. A credible deployment claim requires explicit evidence of robustness under domain shift, observability-aware inference, calibrated uncertainty, controlled lifecycle adaptation, failure-mode awareness, and BMS-facing validation gates.

7. Digital Twins and Multi-Fidelity Validation

7.1. Diagnostic Signals as Information Sources

The diagnostic modalities discussed in Section 5 provide complementary observations of battery degradation processes. However, diagnostic signals do not directly constitute prognostic knowledge. Instead, they represent information sources whose usefulness depends on their observability, information content, noise sensitivity, and compatibility with the chosen modeling framework.

Various sensing modalities capture different aspects of battery behavior. Voltage-relaxation analysis is sensitive to diffusion-related processes and equilibrium dynamics. Incremental capacity and differential voltage analyses reveal changes in electrochemical signatures associated with aging. EIS provides information about charge-transfer resistance, transport limitations, and degradation mechanisms. Thermal measurements capture heat generation patterns that may reflect internal losses and degradation heterogeneity.

Consequently, the effectiveness of a prognostic algorithm depends not only on its architecture but also on the physical information contained within the selected diagnostic signals. Trustworthy battery prognostics therefore require explicit consideration of the relationship between sensing modalities and modeling approaches.

Table 15 addresses a key conceptual paradigm by identifying the specific information that be extracted from the different sensing and feature spaces along with the tailored numerical methods required for analyzing these data.

Table 15. Diagnostic modalities and their preferred learning paradigms.

Diagnostic Modality	Primary Information Content	Preferred Learning Paradigm
Voltage relaxation	Diffusion and equilibrium dynamics	Hybrid/PINN
ICA/DVA	Aging-sensitive electrochemical signatures	ML/Hybrid
EIS/DRT	Resistance growth and mechanistic degradation	Physics-informed
Thermal sensing	Heat generation patterns	Deep learning
Multimodal sensing	Complementary degradation evidence	Hybrid fusion

7.2. Mapping Diagnostic Modalities to Learning Paradigms

No single learning paradigm is universally optimal across all diagnostic modalities. The suitability of a given modeling approach depends strongly on the nature of the available measurements and the physical processes they represent.

Physics-informed models are particularly attractive when diagnostic measurements maintain a clear relationship with known electrochemical mechanisms. EIS and diffusion-sensitive relaxation measurements often provide information that can be directly incorporated into mechanistic constraints. Conversely, thermal imaging and high-dimensional sensing modalities frequently benefit from ML and deep learning (DL) architectures capable of extracting complex nonlinear patterns.

Hybrid approaches increasingly represent an effective compromise between physical interpretability and predictive flexibility. By combining mechanistic constraints with data-driven adaptation, hybrid models may improve transferability while preserving physical consistency. This capability becomes especially critical when models are expected to operate across heterogeneous deployment conditions.

7.3. Multi-Fidelity Validation Pathway

The transition from diagnostic measurements to deployment-oriented prognostics requires a structured validation process. Raw measurements alone do not guarantee trustworthy predictions. Instead, evidence must progressively clear multiple validation stages, each addressing a different source of uncertainty.

Figure 5 presents the proposed multi-fidelity validation roadmap as a staged evidence-filtering process integrated with a deployment-readiness funnel. Rather than a single benchmark exercise, each fidelity level introduces additional realism and stricter evidence requirements. Concurrently, validation gates assess critical criteria, including worst-domain error, cross-fidelity degradation, uncertainty calibration, warning risk, online feasibility, safety governance, and auditability.

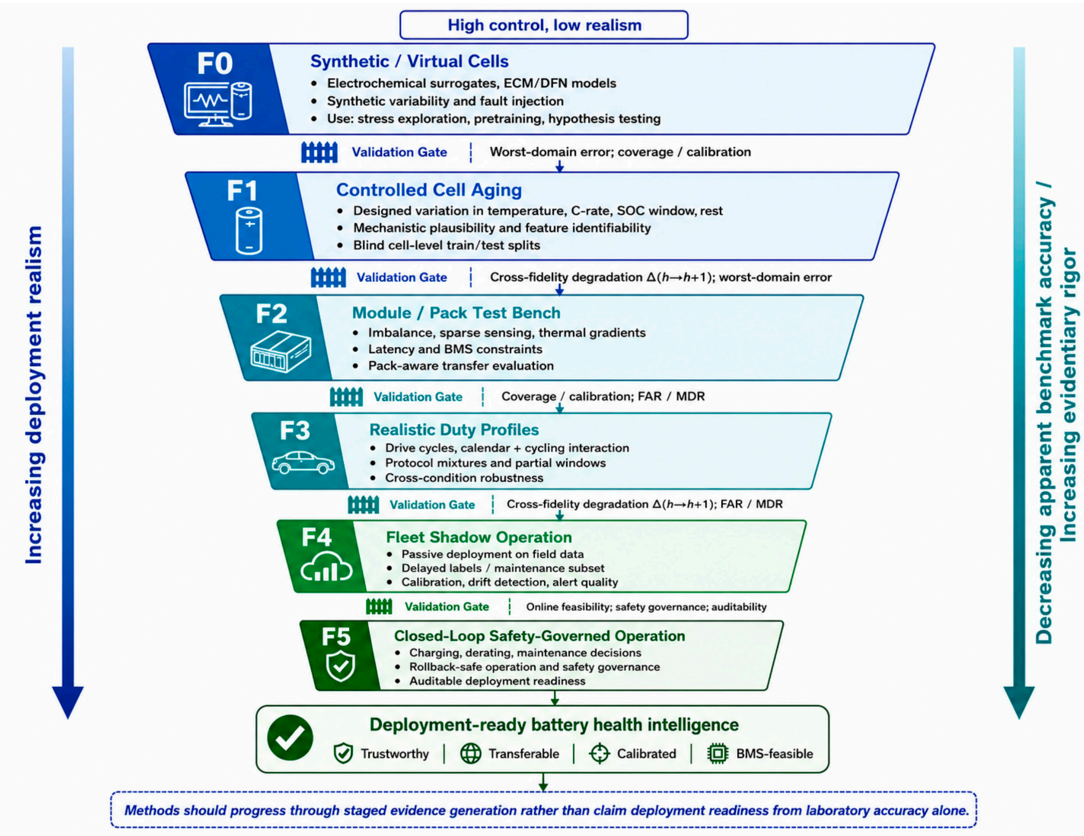

Figure 5. Integrated multi-fidelity validation funnel for trustworthy SOH/RUL estimators.

The multi-fidelity perspective proposed in this review considers validation a hierarchical process linking sensing, feature extraction, model development, uncertainty assessment, and deployment evaluation. Figure 5 summarizes this progression and illustrates how evidence is progressively refined before supporting operational decisions.

To avoid treating multi-fidelity validation and deployment attrition as isolated concepts, these two perspectives can be integrated into a single validation funnel. In this view, each fidelity level introduces additional sources of realism, uncertainty, and decision relevance, while simultaneously narrowing the pool of estimators that can credibly claim deployment readiness. The resulting framework emphasizes that laboratory accuracy represents merely the initial step in the validation process. Trustworthy SOH/RUL estimation requires staged evidence generation, explicit validation gates, cross-fidelity degradation analysis, calibrated uncertainty, and BMS-oriented feasibility assessment.

The framework highlights that validation should not be viewed as a single performanceevaluation step but rather as a continuous process linking measurement quality, model credibility, uncertainty quantification, and deployment readiness.

7.4. Digital Twins as Evidence Integration Frameworks

In the context of battery prognostics, a digital twin should be viewed as an evidenceintegration architecture that combines sensing, physical knowledge, machine learning, uncertainty estimation, and operational feedback.

The twin becomes especially powerful when estimator development is multi-staged based on fidelity, rather than attempted in a single leap from coin-cell or pouch-cell benchmarks to field deployment. This progression motivates the six-stage roadmap proposed in this review [8–10,64–67].

7.5. Toward Closed-Loop Trustworthy Deployment

This integrated view also clarifies why cell-level accuracy alone is insufficient for deployment claims. A model that performs well under controlled aging conditions may still fail when exposed to pack-level heterogeneity, field-data irregularity, or closed-loop BMS decision logic. The funnel therefore structures the central argument of this review: trustworthy and transferable battery health intelligence requires progressive validation across fidelity levels, with each transition acting as a go/no-go gate for robustness, calibration, and practical deployability.

Table 16 translates the digital twin and multi-fidelity concepts into a staged validation structure that is easier to audit, compare, and operationalize.

Table 16. Proposed multi-fidelity validation roadmap.

Fidelity	Test Stage	Primary Purpose	Required Decision Gate
F0	Synthetic and virtual cells	Fast stress exploration, mechanism priors, surrogate generation	Parameter posterior consistency, metadata traceability
F1	Single-cell controlled aging	DOE over temperature, rate, SOC window, rest, fast charge	Blind splits and cross-condition hold-outs
F2	Module/pack hardware-in-the-loop	Sensor sparsity, imbalance, thermal gradients, BMS latency	Pack realism without losing traceable labels
F3	Accelerated duty-cycle realism	Drive cycles, grid service duty, calendar + cycling mixtures	Robustness to mixed mechanisms and duty switching
F4	Fleet shadow deployment	Passive online inference with maintenance ground-truth subset	Calibration, drift monitoring, alert quality
F5	Closed-loop safety-governed operation	Estimator informs control, maintenance, and warranty decisions	Risk acceptance, rollback readiness, auditability

F0 corresponds to synthetic and virtual cells, including electrochemical surrogates, reduced-order Doyle–Fuller–Newman (DFN) and ECMs, manufacturing variability emulation, and synthetic fault injection. This stage is valuable for stress exploration, representation pre-training, and hypothesis testing, but should never be mistaken for deployment evidence. F1 introduces controlled cell aging with designed variation in temperature, rate, SOC window, rest periods, and fast charging. This is where feature identifiability, mechanistic plausibility, and blinded splits must first be established [12,21–23].

F2 and F3 add operational realism that most papers still under-represent. F2 introduces module or pack hardware-in-the-loop (HIL) conditions, such as cell imbalance, sparse sensing, thermal gradients, estimator latency, and BMS constraints. F3 introduces mixed duty cycles and realistic protocol combinations, including calendar-plus-cycling stress, drive cycles, and grid-service operation. Only after these stages does it become meaningful to assess whether an estimator is robust across interacting degradation mechanisms and imperfect observability [12,17,18,21,72].

F4 and F5 represent the decisive gates for trustworthy deployment. F4 involves fleet shadow operation, where predictions run passively on field data while a limited maintenance or reference subset provides delayed ground truth. This stage enables calibration assessment, drift detection, and alert-quality evaluation. F5 constitutes closed-loop safety-governed operation, where outputs influence charging, de-rating, maintenance, or warranty decisions under auditable rollback and safety constraints. The roadmap therefore reframes validation as a progression from model fitting to governable use [11,12,36,37,39,40].

By organizing the evidence ladder into explicit fidelity levels, validation goals, and transition requirements, the table clarifies that deployment readiness is not a binary property but the outcome of progressively accumulated evidence. It therefore supports the primary claim of this review: trustworthy SOH/RUL estimation requires not only methodological performance but also traceable progression from laboratory control to field-relevant operational realism.

To bridge the gap between validated estimators and operational battery management, Figure 6 introduces a deployment pipeline that links sensing inputs, SOH/RUL inference, uncertainty handling, validation gates, BMS decision logic, and feedback mechanisms within an auditable operational loop.

This pipeline highlights that SOH/RUL estimation becomes deployment-relevant only when these core components are comprehensively integrated. Consequently, estimator outputs should be interpreted through risk-aware thresholds and fallback logic rather than point estimates alone.

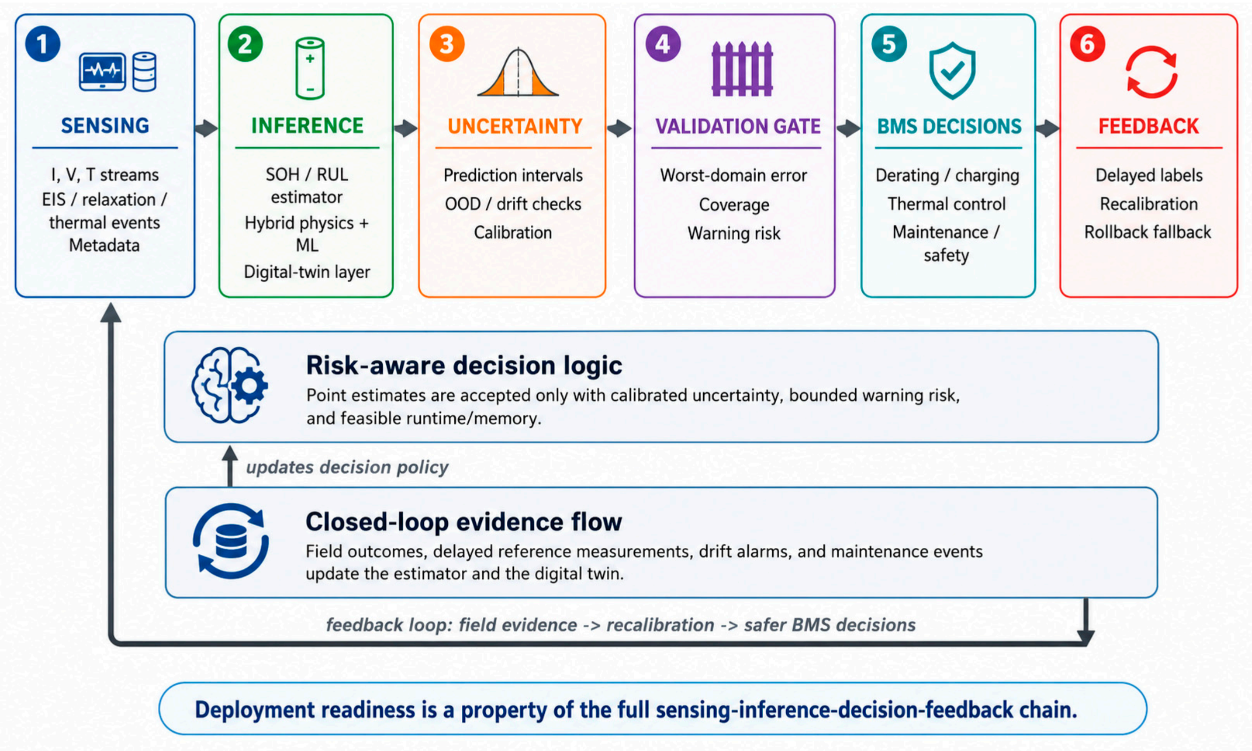

Figure 6. Deployment pipeline linking SOH/RUL estimates, uncertainty, validation gates, BMS decisions, and feedback loops.

8. Critical Gaps, Limitations, and Failure Modes

The review highlights six strategic gaps for the *Batteries* community. Firstly, public benchmarks still over-represent full-cycle laboratory data and under-represent partial windows, sparse sensors, delayed labels, and protocol shift. The growing interest in partial-charge and small-window diagnostics further highlights the need for realistic observability assessment. Although these approaches are among the most promising candidates for deployment-oriented SOH estimation, their effectiveness depends strongly on observation window length, SOC coverage, measurement quality, temperature variability, and the availability of contextual operating information. Future benchmark datasets should therefore report such conditions explicitly rather than treating all partial-window observations as equivalent sources of health information. Secondly, many learned representations do not sufficiently separate irreversible degradation from the reversible operating context, such as the temperature, current history, and SOC window. Thirdly, cross-fidelity transfer from cells to modules, packs, and fleets remains underdeveloped, even though pack imbalance, thermal gradients, balancing events, contact resistance, and sparse sensing can invalidate cell-level assumptions.

Fourthly, uncertainty is still too often reported as an auxiliary output rather than as a deployment acceptance criterion. Future studies should report interval coverage, calibration under domain shift, selective prediction, abstention policies, and decision-conditioned warning risk. Fifthly, cross-chemistry and cross-manufacturer generalization remains weakly validated; stronger studies should use blind external holdouts and report domain boundaries explicitly. Sixthly, rich intermittent diagnostics such as EIS, DRT, relaxation-voltage probes, and thermal maps should be treated as calibration anchors and twin-synchronization events rather than as either purely laboratory tools or continuously deployable signals.

The review protocol also has limitations. It was designed as a critical conceptual review with a structured research set classification rather than as a PRISMA-style systematic review or statistical meta-analysis. Consequently, the comparisons should be interpreted as evidence-maturity judgments rather than pooled quantitative rankings. The primary literature is heterogeneous in chemistry, format, protocol, sensing configuration, target definition, validation split, and reporting completeness, which limits direct numerical comparison across studies.

At the deployment level, the most important failure modes are observability failure, domain-shift failure, cross-fidelity over-claiming, uncertainty miscalibration, data or validation leakage, proxy-label failure, sensor/preprocessing failure, decision integration failure, computational infeasibility, and lifecycle drift. These risks can be mitigated only when validation explicitly links sensing assumptions, domain boundaries, calibration quality, warning risk, computational feasibility, and rollback-safe governance. Detailed evidence classification and traceability information is provided in Supplementary Table S1[103–165].

9. Conclusions

The literature now provides abundant evidence that SOH/RUL estimation can be highly accurate under controlled conditions and increasingly useful under realistic ones. For a journal focused on batteries, however, accuracy alone is no longer the right organizing metric. Robustness under realistic operating variability depends on matching the estimator structure to observability, integrating the mechanism with learning, quantifying uncertainty in deployment-relevant terms, and validating across cell, module, and pack fidelities before closed-loop use [8–11,50–52,166–170].

Rather than suggesting a single winning paradigm among model-based, machine learning, and physics-informed approaches, the reviewed evidence indicates that integrated battery health architectures may offer the greatest potential for improving deployment robustness. Such architectures can combine online electrical estimators with opportunistic diagnostics, mechanistic constraints, uncertainty-aware supervisory logic, and digital-twin-enabled lifecycle management. While the available evidence remains heterogeneous and further validation is required, these integrated approaches appear particularly promising for bridging the gap between benchmark performance and real-world battery health intelligence. However, the present literature does not yet provide sufficient evidence to conclude that any single architecture has emerged as a universally dominant solution across chemistries, applications, and deployment environments [8–11,168–176].

In practical terms, the proposed framework implies that future SOH/RUL studies should report not only mean prediction error but also the sensing assumptions, domain boundaries, uncertainty calibration, cross-fidelity degradation, and BMS-facing acceptance criteria under which the estimator remains valid. Such reporting would make battery-health estimators more comparable, auditable, and transferable across laboratory, pack, and field settings. Taken together, the reviewed literature suggests a gradual transition from model-centered benchmarking toward evidence-centered deployment science. Achieving this transition will require more rigorous reporting of sensing assumptions, domain boundaries, uncertainty calibration, cross-fidelity transferability, and deployment-oriented acceptance criteria. Such developments would improve the comparability, auditability, and practical usefulness of future battery-health estimation systems across laboratory, pack, and field environments.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/batteries12070255/s1, Table S1: Classified research set and evidence-audit workbook for the SOH/RUL review, including the search-protocol summary, set metadata, classification codebook, evidence-maturity rubric, diagnostic-modality categories, estimator-family categories, validation-fidelity labels, variability regimes, trustworthiness dimensions, DOI/URL audit, associated-record audit, and figure/table traceability. Supplementary Figure S1: Extended estimator-challenge matrix. Supplementary Figure S2: Digital-twin stack that links sensing, hybrid inference, validation, and operations.

Author Contributions: Conceptualization, A.S.; methodology, A.S.; investigation, S.R.; writing—original draft preparation, S.R.; writing—review and editing, G.M.; visualization, S.R.; supervision, G.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the research project № КП-06-Н57/2 15.11.2021 "Hybrid and Fusion Prediction of the Functionality of Energy Converting Systems" financed by the Bulgarian National Science Fund (BNSF).

Data Availability Statement: No new experimental battery data were generated in this review. The structured literature corpus and evidence classification workbook used to support the review synthesis are provided as Supplementary Table S1. The final deduplicated corpus contains 176 publication records; associated duplicate or code/software records are documented separately for traceability and are not counted as independent publications.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. 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. https://doi.org/10.1038/s41560-019-0356-8.

2. Berecibar, M.; Gandiaga, I.; Villarreal, I.; Omar, N.; Van Mierlo, J.; Van den Bossche, P. Critical review of state of health estimation methods of Li-ion batteries for real applications. *Renew. Sustain. Energy Rev.* **2016**, *56*, 572–587. https://doi.org/10.1016/j.rser.2015.11.042.
3. Vidal, C.; Malysz, P.; Kollmeyer, P.; Emadi, A. Machine learning applied to electrified vehicle battery state of charge and state of health estimation: State-of-the-art. *IEEE Access* **2020**, *8*, 52796–52814. https://doi.org/10.1109/ACCESS.2020.2980961.
4. Tian, H.; Qin, P.; Li, K.; Zhao, Z. A review of the state of health for lithium-ion batteries: Research status and suggestions. *J. Clean. Prod.* **2020**, *261*, 120813. https://doi.org/10.1016/j.jclepro.2020.120813.
5. Yao, L.; Xu, S.; Tang, A.; Zhou, F.; Hou, J.; Xiao, Y.; Fu, Z. A review of lithium-ion battery state of health estimation and prediction methods. *World Electr. Veh. J.* **2021**, *12*, 113. https://doi.org/10.3390/wevj12030113.
6. Meng, H.; Li, Y.-F. A review on prognostics and health management (PHM) methods of lithium-ion batteries. *Renew. Sustain. Energy Rev.* **2019**, *116*, 109405. https://doi.org/10.1016/j.rser.2019.109405.
7. Sulzer, V.; Mohtat, P.; Aitio, A.; Lee, S.; Yeh, Y.-T.; Steinbacher, F.; Khan, M.U.; Lee, J.W.; Siegel, J.B.; David, A.; et al. The challenge and opportunity of battery lifetime prediction from field data. *Joule* **2021**, *5*, 1934–1955. https://doi.org/10.1016/j.joule.2021.06.005.
8. Wang, Y.; Xiao, J.; Soo, Y.-Y.; Chen, Y.; Chen, Z. Digital twins for battery health prognosis: A comprehensive review of recent advances and challenges. *eTransportation* **2025**, *26*, 100489. https://doi.org/10.1016/j.etran.2025.100489.
9. Ramshankar, S.; Manimozhi, M. Integration of digital twin technologies for state estimation in electric vehicle batteries: A review. *Results Eng.* **2025**, *27*, 106858. https://doi.org/10.1016/j.rineng.2025.106858.
10. Madani, S.S.; Shabeer, Y.; Fowler, M.; Panchal, S.; Ziebert, C.; Chaoui, H.; Allard, F. Digital twin technologies for battery systems: Advancements, applications, and future directions. *Front. Batter. Electrochem.* **2026**, *5*, 1764210. https://doi.org/10.3389/fbael.2026.1764210.
11. Ahwiadi, M.; Wang, W. Battery health monitoring and remaining useful life prediction techniques: A review of technologies. *Batteries* **2025**, *11*, 31. https://doi.org/10.3390/batteries11010031.
12. Smith, K.; Gasper, P. BLAST-Lite (Battery Lifetime Analysis and Simulation Tool—Lite) [SWR-22-69] Related to: BLAST aka: BLAST-Py; National Renewable Energy Laboratory: Golden, CO, USA, 2023. https://doi.org/10.11578/dc.20230724.1.
13. Hassini, M.; Redondo-Iglesias, E.; Venet, P. Lithium-ion battery data: From production to prediction. *Batteries* **2023**, *9*, 385. https://doi.org/10.3390/batteries9070385.
14. Lucaferri, V.; Quercio, M.; Laudani, A.; Riganti Fulginei, F. A review on battery model-based and data-driven methods for battery management systems. *Energies* **2023**, *16*, 7807. https://doi.org/10.3390/en16237807.
15. Xiong, R.; Sun, F.; He, H. Model-based health condition monitoring method for a multi-cell series-connected battery pack. In *Proceedings of the* 2016 IEEE Transportation Electrification Conference and Expo (ITEC), Dearborn, MI, USA, 27–29 June 2016. https://doi.org/10.1109/ITEC.2016.7520202.
16. Ma, L.; Xu, Y.; Zhang, H.; Yang, F.; Wang, X.; Li, C. Co-estimation of state of charge and state of health for lithium-ion batteries based on fractional-order model with multi-innovations unscented Kalman filter method. *J. Energy Storage* **2022**, *52*, 104904. https://doi.org/10.1016/j.est.2022.104904.
17. Li, Q.; Song, R.; Wei, Y. A review of state-of-health estimation for lithium-ion battery packs. *J. Energy Storage* **2025**, *118*, 116078. https://doi.org/10.1016/j.est.2025.116078.
18. 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. https://doi.org/10.1016/j.rser.2025.116125.
19. Tang, K.; Luo, B.; Chen, D.; Wang, C.; Chen, L.; Li, F.; Cao, Y.; Wang, C. The state of health estimation of lithium-ion batteries: A review of health indicators, estimation methods, development trends and challenges. World Electr. *Veh. J.* **2025**, *16*, 429. https://doi.org/10.3390/wevj16080429.
20. United States Advanced Battery Consortium (USABC). Electric Vehicle Test Manual; USABC/USCAR: Southfield, MI, USA; Current Online Edition. Available online: https://uscar.org/usabc/manuals-models/ (accessed on 25 April 2026).
21. European Commission. Accelerated Multi-Physical and Virtual Testing for Battery Aging, Reliability, and Safety Evaluation (Batt4EU Partnership); Topic HORIZON-CL5-2026-01-D2-05; Horizon Europe Work Programme Topic Text; European Commission: Brussels, Belgium, 2025. Available online: https://cordis.europa.eu/programme/id/HORIZON_HORIZON-CL5-2026-01-D2-05 (accessed on 25 April 2026).

22. Smith, K.; Wood, E.; Santhanagopalan, S.; Kim, G.-H.; Shi, Y.; Pesaran, A. Predictive Models of Li-ion Battery Lifetime; National Renewable Energy Laboratory: Golden, CO, USA, 2014; NREL/PR-5400-62813. Available online: https://www.nrel.gov/docs/fy14osti/62813.pdf (accessed on 25 April 2026).
23. Grant, P.S.; Greenwood, D.; Pardikar, K.; Smith, R.; Entwistle, T.; Middlemiss, L.A.; Murray, G.; Cussen, S.A.; Lain, M.J.; Capener, M.J.; et al. Roadmap on Li-ion battery manufacturing research. *J. Phys. Energy* **2022**, *4*, 042006. https://doi.org/10.1088/2515-7655/ac8e30.
24. Shah, A.; Shah, K.; Shah, C.; Shah, M. State of charge, remaining useful life and knee point estimation based on artificial intelligence and Machine learning in lithium-ion EV batteries: A comprehensive review. *Renew. Energy Focus* **2022**, *42*, 146–164. https://doi.org/10.1016/j.ref.2022.06.001.
25. Murphy, K.P. Machine Learning: A Probabilistic Perspective; MIT Press: Cambridge, MA, USA, 2012. Available online: https://moodle2.units.it/pluginfile.php/241303/mod_resource/content/1/Murphy_Machine_Learning.pdf (accessed on 25 April 2026).
26. Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378. https://doi.org/10.1198/016214506000001437.
27. Wang, F.; Zhai, Z.; Zhao, Z.; Di, Y.; Chen, X. Physics-informed neural network for lithium-ion battery degradation stable modeling and prognosis. *Nat. Commun.* **2024**, *15*, 4332. https://doi.org/10.1038/s41467-024-48779-z.
28. Ye, J.; Xie, Q.; Lin, M.; Wu, J. A method for estimating the state of health of lithium-ion batteries based on physics-informed neural network. *Energy* **2024**, *294*, 130828. https://doi.org/10.1016/j.energy.2024.130828.
29. Wen, P.; Ye, Z.-S.; Li, Y.; Chen, S.; Xie, P.; Zhao, S. Physics-informed neural networks for prognostics and health management of lithium-ion batteries. *IEEE Trans. Intell. Veh.* **2024**, *9*, 2276–2289. https://doi.org/10.1109/TIV.2023.3315548.
30. Navidi, S.; Thelen, A.; Li, T.; Hu, C. Physics-informed neural networks for degradation diagnostics of lithium-ion batteries. In Proceedings of ASME Conference Proceedings, Boston, MA, USA, 20–23 August 2023; ASME: Houston, TX, USA, 2023; Volume 3A. https://doi.org/10.1115/DETC2023-116940.
31. Xu, Z.; Guo, Y.; Saleh, J.H. A physics-informed dynamic deep autoencoder for accurate state-of-health prediction of lithium-ion batteries. Neural Comput. *Appl.* **2022**, *34*, 15997–16017. https://doi.org/10.1007/s00521-022-07291-5.
32. Tian, H.; Xi, C.; Zhang, Q. A framework for estimating battery state of health using multi-source domain adaptation and real-vehicle data. *J. Energy Storage* **2025**, *136*, 118449. https://doi.org/10.1016/j.est.2025.118449.
33. Lanubile, A.; Bosoni, P.; Pozzato, G.; Allam, A.; Acquarone, M.; Onori, S. Domain knowledge-guided machine learning framework for state-of-health estimation in lithium-ion batteries. *Commun. Eng.* **2024**, *3*, 168. https://doi.org/10.1038/s44172-024-00304-2.
34. Navidi, S.; Thelen, A.; Li, T.; Hu, C. Physics-informed machine learning for battery degradation diagnostics: A comparison of state-of-the-art methods. *Energy Storage Mater.* **2024**, *68*, 103343. https://doi.org/10.1016/j.ensm.2024.103343.
35. Sagawa, S.; Koh, P.W.; Hashimoto, T.B.; Liang, P. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. *arXiv* 2020, arXiv:1911.08731. https://doi.org/10.48550/arXiv.1911.08731.
36. da Silva, S.F.; Ozkan, M.F.; Idrissi, F.E.; Canova, M. Conformalized transfer learning for Li-ion battery state-of-health forecasting under manufacturing and usage variability. *arXiv* 2026, arXiv:2603.24475. https://doi.org/10.48550/arXiv.2603.24475.
37. Angelopoulos, A.N.; Bates, S. Conformal Prediction: A Gentle Introduction. *Found. Trends Mach. Learn.* **2023**, *16*, 494–591. https://doi.org/10.1561/2200000101.
38. Pang, T.; Fan, G.; Zhou, B.; Wang, Y.; Wang, Y.; Zhang, X. Robust capacity estimation with uncertainty quantification for Li-ion batteries under temporal data-masking challenges: A progressive learning approach. *Appl. Energy* **2025**, *401*, 126648. https://doi.org/10.1016/j.apenergy.2025.126648.
39. Martiri, L.; Cristaldi, L. Evidential deep learning for quantification of uncertainty in Lithium-Ion Batteries remaining useful life estimation. *Energies* **2026**, *19*, 1513. https://doi.org/10.3390/en19061513.
40. Parimala, C.H.; Varma, P.S.; Kumar, R.S.; Singh, A.R.; Mohapatra, B.; Addis, H.K. Uncertainty aware hybrid learning framework for fast and safe charging of lithium-ion batteries using multi-fidelity observers. *Sci. Rep.* **2026**, *16*, 2452. https://doi.org/10.1038/s41598-025-31976-1.
41. You, H.; Wang, X.; Zhu, J.; Jiang, B.; Han, G.; Wei, X.; Dai, H. Investigation of lithium-ion battery nonlinear degradation by experiments and model-based simulation. *Energy Storage Mater.* **2024**, *65*, 103083. https://doi.org/10.1016/j.ensm.2023.103083.
42. Xu, Y.; Ge, X.; Guo, R.; Shen, W. Recent advances in model-based fault diagnosis for lithium-ion batteries: A comprehensive review. *Renew. Sustain. Energy Rev.* **2025**, *207*, 114922. https://doi.org/10.1016/j.rser.2024.114922.

43. Klass, V.; Behm, M.; Lindbergh, G. A support vector machine-based state-of-health estimation method for lithium-ion batteries under electric vehicle operation. *J. Power Sources* **2014**, *270*, 262–272. https://doi.org/10.1016/j.jpowsour.2014.07.116.
44. Ma, Z.; Yang, R.; Wang, Z. A novel data-model fusion state-of-health estimation approach for lithium-ion batteries. *Appl. Energy* **2019**, *237*, 836–847. https://doi.org/10.1016/j.apenergy.2018.12.071.
45. Bian, X.; Wei, Z.; Li, W.; Pou, J.; Sauer, D.U.; Liu, L. State-of-health estimation of lithium-ion batteries by fusing an open circuit voltage model and incremental capacity analysis. *IEEE Trans. Power Electron.* **2022**, *37*, 2226–2236. https://doi.org/10.1109/TPEL.2021.3104723.
46. Huang, P.; Gu, P.; Kang, Y.; Zhang, Y.; Duan, B.; Zhang, C. The state-of-health estimation of lithium-ion batteries based on data-driven and model-fusion method. *J. Clean. Prod.* **2022**, *366*, 132742. https://doi.org/10.1016/j.jclepro.2022.132742.
47. Chen, L.; Xie, S.; Lopes, A.M.; Li, H.; Bao, X.; Zhang, C.; Li, P. A new SOH estimation method for lithium-ion batteries based on model-data fusion. *Energy* **2024**, *286*, 129597. https://doi.org/10.1016/j.energy.2023.129597.
48. Wang, R.; Song, C.; Gao, M.; Zhao, J.; Xu, Z. Model-data-fusion domain adaptation for battery state-of-health estimation with fewer data and simplified feature extraction. *J. Energy Storage* **2023**, *60*, 106686. https://doi.org/10.1016/j.est.2023.106686.
49. Chen, L.; Lü, Z.; Lin, W.; Li, J.; Pan, H. A new state-of-health estimation method for lithium-ion batteries through the intrinsic relationship between ohmic internal resistance and capacity. *Measurement* **2018**, *116*, 586–595. https://doi.org/10.1016/j.measurement.2017.11.016.
50. Su, L.; Xu, Y.; Dong, Z. State-of-health estimation of lithium-ion batteries: A comprehensive literature review from cell to pack levels. *Energy Convers. Econ.* **2024**, *5*, 224–242. https://doi.org/10.1049/enc2.12125.
51. Liu, C.; Li, H.; Li, K.; Wu, Y.; Lv, B. Deep learning for state of health estimation of lithium-ion batteries in electric vehicles: A systematic review. *Energies* **2025**, *18*, 1463. https://doi.org/10.3390/en18061463.
52. Yang, K.; Wang, S.; Zhou, L.; Fernandez, C.; Blaabjerg, F. A critical review of AI-based battery remaining useful life prediction for energy storage systems. *Batteries* **2025**, *11*, 376. https://doi.org/10.3390/batteries11100376.
53. Patel, K.; Gosala, V.; Stender, M.; Braun, M.; Ehlers, S. Predicting remaining useful life of lithium-ion batteries: A review of degradation mechanisms and open-source data availability. *Future Batter.* **2025**, *8*, 100124. https://doi.org/10.1016/j.fub.2025.100124.
54. Shen, S.; Sadoughi, M.; Li, M.; Wang, Z.; Hu, C. Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries. *Appl. Energy* **2022**, *260*, 114296. https://doi.org/10.1016/j.apenergy.2019.114296.
55. Fan, Y.; Huang, Z.; Li, H.; Kaleem, M.B.; Wu, Y. State-of-health Estimation for Battery Packs of real-world electric vehicles using transfer learning. *Energy* **2025**, *336*, 138328. https://doi.org/10.1016/j.energy.2025.138328.
56. He, R.; Wang, C.; Yin, C.; Yang, S.; Wang, Y.; Fang, Y.; Chen, K.; Zhang, J. Instance-based transfer learning-improved battery state-of-health estimation with self-attention mechanism. *Energies* **2025**, *18*, 5672. https://doi.org/10.3390/en18215672.
57. Ma, G.; Xu, S.; Yang, T.; Du, Z.; Zhu, L.; Ding, H. A transfer-learning-based method for personalized state-of-health estimation of lithium-ion batteries. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 759–769. https://doi.org/10.1109/TNNLS.2022.3176925.
58. Ye, Z.; Yu, J. State-of-health estimation for lithium-ion batteries using domain-adversarial transfer learning. *IEEE Trans. Power Electron.* **2022**, *37*, 3528–3543. https://doi.org/10.1109/TPEL.2021.3117788.
59. Wang, Q.; Ye, M.; Cai, X.; Sauer, D.U.; Li, W. Transferable data-driven capacity estimation for lithium-ion batteries with deep learning: A case study from laboratory to field applications. *Appl. Energy* **2023**, *350*, 121747. https://doi.org/10.1016/j.apenergy.2023.121747.
60. Fu, S.; Tao, S.; Fan, H.; He, K.; Liu, X.; Tao, Y.; Zuo, J.; Zhang, X.; Wang, Y.; Sun, Y. Data-driven capacity estimation for lithium-ion batteries with a feature-matching-based transfer learning method. *Appl. Energy* **2024**, *353*, 121991. https://doi.org/10.1016/j.apenergy.2023.121991.
61. Deng, Z.; Xu, L.; Liu, H.; Hu, X.; Wang, B.; Zhou, J. Rapid health estimation of in-service battery packs based on limited labels and domain adaptation. *J. Energy Chem.* **2025**, *89*, 345–354. https://doi.org/10.1016/j.jechem.2023.10.056.
62. Tian, J.; Liu, X.; Li, S.; Wei, Z.; Zhang, X.; Xiao, G.; Wang, P. Lithium-ion battery health estimation with real-world data for electric vehicles. *Energy* **2023**, *270*, 126855. https://doi.org/10.1016/j.energy.2023.126855.
63. Wen, S.; Lin, N.; Huang, S.; Li, X.; Wang, Z.; Zhang, Z. Lithium battery state-of-health estimation using real-world vehicle data and an interpretable hybrid framework. *J. Energy Storage* **2024**, *96*, 112623. https://doi.org/10.1016/j.est.2024.112623.
64. Qasem, M.; Stubblefield, J.; Qandil, M.; Yassin, Y.; Haddadin, M.; Krishnamurthy, M. Real-time AI-enabled digital twin for battery health estimation and fast charging using partial-discharge data. *eTransportation* **2026**, *27*, 100528. https://doi.org/10.1016/j.etran.2025.100528.

65. Li, W.; Li, Y.; Garg, A.; Gao, L. Enhancing real-time degradation prediction of lithium-ion battery: A digital twin framework with CNN-LSTM-attention model. *Energy* **2024**, *286*, 129681. https://doi.org/10.1016/j.energy.2023.129681.
66. Li, W.; Rentemeister, M.; Badeda, J.; Jöst, D.; Schulte, D.; Sauer, D.U. Digital twin for battery systems: Cloud battery management system with online state-of-charge and state-of-health estimation. *J. Energy Storage* **2020**, *30*, 101557. https://doi.org/10.1016/j.est.2020.101557.
67. Naseri, F.; Gil, S.; Barbu, C.; Cetkin, E.; Yarimca, G.; Jensen, A.C.; Larsen, P.G.; Gomes, C. Digital twin of electric vehicle battery systems: Comprehensive review of the use cases, requirements, and platforms. *Renew. Sustain. Energy Rev.* **2023**, *179*, 113280. https://doi.org/10.1016/j.rser.2023.113280.
68. Jiang, B.; Zhu, J.; Wang, X.; Wei, X.; Shang, W.; Dai, H. A comparative study of different features extracted from electrochemical impedance spectroscopy in state of health estimation for lithium-ion batteries. *Appl. Energy* **2022**, *322*, 119502. https://doi.org/10.1016/j.apenergy.2022.119502.
69. Zhou, Z.; Duan, B.; Kang, Y.; Zhang, Q.; Shang, Y.; Zhang, C. Online state-of-health estimation for a series-connected LiFePO4 battery pack based on differential voltage and inconsistency analysis. *IEEE Trans. Transp. Electrif.* **2024**, *10*, 989–998. https://doi.org/10.1109/TTE.2023.3274819.
70. Li, Y.; Abdel-Monem, M.; Gopalakrishnan, R.; Berecibar, M.; Nanini-Maury, E.; Omar, N.; Bossche, P.; Van Mierlo, J. A quick on-line state of health estimation method for Li-ion battery with incremental capacity curves processed by Gaussian filter. *J. Power Sources* **2019**, *373*, 40–53. https://doi.org/10.1016/j.jpowsour.2017.10.092.
71. Bilfinger, P.; Rosner, P.; Schreiber, M.; Kröger, T.; Gamra, K.A.; Ank, M.; Wassiliadis, N.; Dietermann, B.; Lienkamp, M. Battery pack diagnostics for electric vehicles: Transfer of differential voltage and incremental capacity analysis from cell to vehicle level. *eTransportation* **2024**, *22*, 100356. https://doi.org/10.1016/j.etran.2024.100356.
72. Xia, F.; Wang, K.; Chen, J. State of health and remaining useful life prediction of lithium-ion batteries based on a disturbance-free incremental capacity and differential voltage analysis method. *J. Energy Storage* **2023**, *64*, 107161. https://doi.org/10.1016/j.est.2023.107161.
73. Beatty, M.; Strickland, D.; Ferreira, P. A review of methods of generating incremental capacity-differential voltage curves for battery health determination. *Energies* **2024**, *17*, 4309. https://doi.org/10.3390/en17174309.
74. Zhou, Q.; Vuylsteke, G.; Anderson, R.D.; Sun, J. Battery state-of-health estimation and incremental capacity analysis under a dynamic charging profile using neural networks. *arXiv* 2025, arXiv:2502.19586. https://doi.org/10.48550/arXiv.2502.19586.
75. Ko, C.-J.; Chen, K.-C. Using tens of seconds of relaxation voltage to estimate open-circuit voltage and state of health of lithium-ion batteries. *Appl. Energy* **2024**, *357*, 122488. https://doi.org/10.1016/j.apenergy.2023.122488.
76. Baghdadi, I.; Briat, O.; Gyan, P.; Vinassa, J.M. State-of-health assessment for lithium batteries based on voltage-time relaxation measure. *Electrochim. Acta* **2016**, *194*, 461–472. https://doi.org/10.1016/j.electacta.2016.02.109.
77. Ke, X.; Hong, H.; Zheng, P.; Zhang, S.; Zhu, L.; Li, Z.; Cai, J.; Fan, P.; Yang, J.; Wang, J.; et al. State-of-health estimation for lithium-ion batteries using relaxation voltage under dynamic conditions. *J. Energy Storage* **2024**, *100*, 113506. https://doi.org/10.1016/j.est.2024.113506.
78. Zhang, S.; Zhu, H.; Wu, J.; Chen, Z. Voltage relaxation-based state-of-health estimation of lithium-ion batteries using convolutional neural networks and transfer learning. *J. Energy Storage* **2023**, *73*, 108579. https://doi.org/10.1016/j.est.2023.108579.
79. Liu, Y.; Liu, Y.; Bao, X.; Shen, H. Estimation of battery state of health and open-circuit voltage at various depths of discharge based on deep learning and relaxation voltage. *Energy* **2025**, *328*, 136555. https://doi.org/10.1016/j.energy.2025.136555.
80. Meng, J.; Cai, L.; Stroe, D.-I.; Luo, G.; Sui, X.; Teodorescu, R. Lithium-ion battery state-of-health estimation in electric vehicle using optimized partial charging voltage profiles. *Energy* **2019**, *185*, 1054–1062. https://doi.org/10.1016/j.energy.2019.07.127.
81. Lin, C.; Xu, J.; Jiang, D.; Hou, J.; Liang, Y.; Zhang, X.; Li, E.; Mei, X. A comparative study of data-driven battery capacity estimation based on partial charging curves. *J. Energy Chem.* **2024**, *88*, 409–420. https://doi.org/10.1016/j.jechem.2023.09.025.
82. Schaltz, E.; Stroe, D.-I.; Nørregaard, K.; Johnsen, B.; Christensen, A. Partial charging method for lithium-ion battery state-of-health estimation. In *Proceedings of the* 2019 Fourteenth International Conference on Ecological Vehicles and Renewable Energies (EVER), Monte-Carlo, Monaco, 8–10 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–5. https://doi.org/10.1109/EVER.2019.8813645.
83. Tian, J.; Xiong, R.; Shen, W.; Lu, J.; Sun, F. Flexible battery state of health and state of charge estimation using partial charging data and deep learning. *Energy Storage Mater.* **2022**, *51*, 372–381. https://doi.org/10.1016/j.ensm.2022.06.053.
84. Jones, P.K.; Stimming, U.; Lee, A.A. Impedance-based forecasting of lithium-ion battery performance amid uneven usage. *Nat. Commun.* **2022**, *13*, 4806. https://doi.org/10.1038/s41467-022-32422-w.

85. Iurilli, P.; Brivio, C.; Wood, V. On the use of electrochemical impedance spectroscopy to characterize and model the aging phenomena of lithium-ion batteries: A critical review. *J. Power Sources* **2021**, *505*, 229860. https://doi.org/10.1016/j.jpowsour.2021.229860.
86. Sun, X.; Zhang, Y.; Zhang, Y.; Wang, L.; Wang, K. Summary of health-state estimation of lithium-ion batteries based on electrochemical impedance spectroscopy. *Energies* **2023**, *16*, 5682. https://doi.org/10.3390/en16155682.
87. Wu, J.; Meng, J.; Lin, M.; Wang, W.; Wu, J.; Stroe, D.I. Lithium-ion battery state of health estimation using a hybrid model with electrochemical impedance spectroscopy. *Reliab. Eng. Syst. Saf.* **2024**, *252*, 110450. https://doi.org/10.1016/j.ress.2024.110450.
88. Li, D.; Yang, D.; Li, L.; Wang, L.; Wang, K. Electrochemical impedance spectroscopy based on the state of health estimation for Lithium-Ion Batteries. *Energies* **2022**, *15*, 6665. https://doi.org/10.3390/en15186665.
89. Zhang, W.; Ahmed, R.; Habibi, S. State-of-health estimation for lithium-ion batteries based on electrochemical impedance spectroscopy measurements combined with an unscented Kalman filter. *J. Power Sources* **2025**, *625*, 235450. https://doi.org/10.1016/j.jpowsour.2024.235450.
90. Xing, Q.; Zhang, M.; Fu, Y.; Wang, K. Transfer learning to estimate lithium-ion battery state of health with electrochemical impedance spectroscopy. *J. Energy Storage* **2025**, *110*, 115345. https://doi.org/10.1016/j.est.2025.115345.
91. Vendrame, F.; Damay, N.; Rabab, H.; Forgez, C.; Sayegh, M. Non-invasive diagnosis of solid electrolyte interphase and charge transfer resistances for degraded lithium-ion batteries. *J. Power Sources* **2024**, *595*, 234049. https://doi.org/10.1016/j.jpowsour.2023.234049.
92. Shao, B.; Zhong, J.; Tian, J.; Li, Y.; Chen, X.; Dou, W.; Liao, Q.; Lai, C.; Lu, T.; Xie, J. State-of-health estimation of lithium-ion batteries based on electrochemical impedance spectroscopy features and fusion interpretable deep learning framework. *Energies* **2025**, *18*, 1385. https://doi.org/10.3390/en18061385.
93. Peng, J.; Gao, Y.; Cai, L.; Zhang, M.; Sun, C.; Liu, H. State-of-health estimation for lithium-ion batteries using electrochemical impedance spectroscopy and a multi-scale kernel extreme learning machine. *World Electr. Veh. J.* **2025**, *16*, 224. https://doi.org/10.3390/wevj16040224.
94. Xia, B.; Qin, Z.; Fu, H. Rapid estimation of battery state of health using partial electrochemical impedance spectra and interpretable machine learning. *J. Power Sources* **2024**, *603*, 234413. https://doi.org/10.1016/j.jpowsour.2024.234413.
95. Al-Smadi, M.K.; Abu Qahouq, J.A. State of health estimation for lithium-ion batteries based on transition frequency's impedance and other impedance features with correlation analysis. *Batteries* **2025**, *11*, 133. https://doi.org/10.3390/batteries11040133.
96. Zhang, C.; Luo, L.; Yang, Z.; Du, B.; Zhou, Z.; Wu, J.; Chen, L. Flexible method for estimating the state of health of lithium-ion batteries using partial charging segments. *Energy* **2024**, *295*, 131009. https://doi.org/10.1016/j.energy.2024.131009.
97. Deng, Z.; Hu, X.; Li, P.; Lin, X.; Bian, X. Data-driven battery state of health estimation based on random partial charging data. *IEEE Trans. Power Electron.* **2022**, *37*, 5021–5031. https://doi.org/10.1109/TPEL.2021.3134701.
98. Zhou, Z.; Duan, B.; Kang, Y.; Shang, Y.; Zhang, Q.; Zhang, C. Practical state of health estimation for LiFePO4 batteries based on Gaussian mixture regression and incremental capacity analysis. *IEEE Trans. Ind. Electron.* **2023**, *70*, 2576–2585. https://doi.org/10.1109/TIE.2022.3167142.
99. Wei, Z.; Ruan, H.; Li, Y.; Li, J.; Zhang, C.; He, H. Multistage state of health estimation of lithium-ion battery with high tolerance to heavily partial charging. *IEEE Trans. Power Electron.* **2022**, *37*, 7432–7442. https://doi.org/10.1109/TPEL.2022.3144504.
100. Liu, X.; Tang, Q.; Feng, Y.; Lin, M.; Meng, J.; Wu, J. Fast sorting method of retired batteries based on multi-feature extraction from partial charging segment. *Appl. Energy* **2023**, *351*, 121930. https://doi.org/10.1016/j.apenergy.2023.121930.
101. Hoke, A.; Brissette, A.; Maksimović, D.; Kelly, D.; Pratt, A. Maximizing lithium ion vehicle battery life through optimized partial charging. In *Proceedings of the* 2013 IEEE PES Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 24–27 February 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 1–5. https://doi.org/10.1109/ISGT.2013.6497818.
102. Chen, K.; Zheng, F.; Jiang, J.; Zhang, W.; Jiang, Y.; Chen, K. Practical failure recognition model of lithium-ion batteries based on partial charging process. *Energy* **2017**, *138*, 1199–1208. https://doi.org/10.1016/j.energy.2017.08.017.
103. Zhang, S.; Wu, S.; Cao, G.; Zhang, X. Capacity estimation for lithium-ion battery via a novel health indicator extracted from partial constant voltage charging curve. *J. Clean. Prod.* **2023**, *409*, 137220. https://doi.org/10.1016/j.jclepro.2023.137220.
104. Wang, Z.; Zeng, S.; Guo, J.; Qin, T. State of health estimation of lithium-ion batteries based on the constant voltage charging curve. *Energy* **2019**, *167*, 661–669. https://doi.org/10.1016/j.energy.2018.11.008.
105. Sun, J.; Liu, X.; Li, X.; Chen, S.; Xing, S.; Guo, Y. State of health estimation of lithium-ion battery based on constant current charging time feature extraction and internal resistance compensation. *Energy* 2025, 134472. https://doi.org/10.1016/j.energy.2025.134472.

106. Chen, S.-Z.; Liang, Z.; Yuan, H.; Yang, L.; Xu, F.; Fan, Y. A novel state of health estimation method for lithium-ion batteries based on constant-voltage charging partial data and convolutional neural network. *Energy* **2023**, *283*, 129103. https://doi.org/10.1016/j.energy.2023.129103.
107. Li, H.; Wang, S.; Yang, L.; Zhou, Z.; Meng, Y.; Zhang, W.; Shuan, Z.; Li, Y.; Lv, F. SOH estimation method for lithium-ion battery packs under real-world operating conditions based on a new attenuated model without additional experiments. *Energy* 2025, 136802. https://doi.org/10.1016/j.energy.2025.136802.
108. Lin, M.; Ke, L.; Meng, J.; Wang, W.; Wu, J.; Stroe, D.-I. Health status estimation of lithium-ion battery under arbitrary charging voltage information using ensemble learning framework. *Reliab. Eng. Syst. Saf.* **2025**, *256*, 110782. https://doi.org/10.1016/j.ress.2024.110782.
109. Yao, Q.; Han, X. Data-driven lithium-ion batteries capacity estimation based on deep transfer learning using partial segment of charging/discharging data. *Energy* **2023**, *271*, 127033. https://doi.org/10.1016/j.energy.2023.127033.
110. Bockrath, S.; Lorentz, V.; Pruckner, M. State of health estimation of lithium-ion batteries with a temporal convolutional neural network using partial load profiles. *Appl. Energy* **2023**, *329*, 120307. https://doi.org/10.1016/j.apenergy.2022.120307.
111. Chacón, X.C.A.; Laureti, S.; Ricci, M.; Cappuccino, G. A review of non-destructive techniques for lithium-ion battery performance analysis. *World Electr. Veh. J.* **2023**, *14*, 305. https://doi.org/10.3390/wevj14110305.
112. Wang, Z.-J.; Li, Z.-Q.; Liu, Q. Infrared thermography non-destructive evaluation of lithium-ion battery. *Proc. SPIE* **2011**, *8193*, 81934I. https://doi.org/10.1117/12.901592.
113. Kunz, A.; Kirst, C.; Durdel, A.; Singer, J.P.; Jossen, A. Assessing degradation modes in cylindrical lithium-ion batteries: A non-invasive method using the entropic heat flow or surface temperature. *J. Power Sources* **2025**, *627*, 235732. https://doi.org/10.1016/j.jpowsour.2024.235732.
114. Ahuja, N.; Moon, J.; Bhaskar, K.; Rahn, C.D. Lithium-ion battery State of Health estimation based solely on temperature sensing. *J. Power Sources* **2026**, *666*, 239068. https://doi.org/10.1016/j.jpowsour.2025.239068.
115. Jang, J.; Jo, J.; Kim, J.; Lee, S.; Lee, T.; Yoo, J. State of health estimation of lithium-ion battery cell based on optical thermometry with physics-informed machine learning. *Eng. Appl. Artif. Intell.* **2025**, *140*, 109704. https://doi.org/10.1016/j.engappai.2024.109704.
116. Zhang, Y.; Zhang, D.; Wu, T. Method for Estimating the State of Health of Lithium-ion Batteries Based on Differential Thermal Voltammetry and Sparrow Search Algorithm-Elman Neural Network. *Energy Eng.* **2025**, *122*, 203–220. https://doi.org/10.32604/ee.2024.056244.
117. Lv, Z.; Ou, S.; Jing, H.; Dong, H.; Shi, D.; Zhao, J. Battery state-of-health estimation via thermoelectric graph learning and interpretable temporal decomposition for real-world electric vehicles. *J. Energy Storage* **2026**, *153*, 120956. https://doi.org/10.1016/j.est.2026.120956.
118. Li, M.; Xu, W.; Zhang, S.; Liu, L.; Hussain, A.; Hu, E.; Zhang, J.; Mao, Z.; Chen, Z. State of Health Estimation and Battery Management: A Review of Health Indicators, Models and Machine Learning. *Materials* **2025**, *18*, 145. https://doi.org/10.3390/ma18010145.
119. Tian, L.; Dong, C.; Qi, H.; Fu, G.; Jia, H. Live-stream infrared perception for lithium-ion battery thermal diagnosis. *iScience* 2025, 113774. https://doi.org/10.1016/j.isci.2025.113774.
120. Bruj, O.; Calborean, A. Electrochemical Impedance Spectroscopy Investigation on the Charge–Discharge Cycle Life Performance of Lithium-Ion Batteries. *Energies* **2025**, *18*, 1324. https://doi.org/10.3390/en18061324.
121. Li, W.; Zhang, H.; van Vlijmen, B.; Dechent, P.A.; Sauer, D.U. Forecasting battery capacity and power degradation with multi-task learning. *Energy Storage Mater.* **2022**, *53*, 453–466. https://doi.org/10.1016/j.ensm.2022.09.013.
122. Li, W.; Sengupta, N.; Dechent, P.A.; Howey, D.; Annaswamy, A.; Sauer, D.U. One-shot battery degradation trajectory prediction with deep learning. *J. Power Sources* **2021**, *506*, 230024. https://doi.org/10.1016/j.jpowsour.2021.230024.
123. Thelen, A.; Lui, Y.H.; Shen, S.; Laflamme, S.; Hu, S.; Ye, H.; Hu, C. Integrating physics-based modeling and machine learning for degradation diagnostics of lithium-ion batteries. *Energy Storage Mater.* **2022**, *50*, 668–695. https://doi.org/10.1016/j.ensm.2022.05.047.
124. Lin, M.; You, Y.; Meng, J.; Wang, W.; Wu, J.; Stroe, D.-I. Lithium-ion battery degradation trajectory early prediction with synthetic dataset and deep learning. *J. Energy Chem.* **2023**, *85*, 534–546. https://doi.org/10.1016/j.jechem.2023.06.036.
125. Cai, L.; Li, J.; Xu, X.; Jin, H.; Meng, J.; Wang, B.; Wu, C.; Yang, S. Automatically constructing a health indicator for lithium-ion battery state-of-health estimation via adversarial and compound stacked autoencoder. *J. Energy Storage* **2024**, *84*, 110711. https://doi.org/10.1016/j.est.2024.110711.

126. Zhou, K.Q.; Qin, Y.; Yuen, C. Lithium-ion Battery State of Health Estimation by Matrix Profile Empowered Online Knee Onset Identification. *IEEE Trans. Transp. Electrif.* **2024**, *10*, 1935–1946. https://doi.org/10.1109/TTE.2023.3265981.
127. Liu, W.; Xu, Y. Data-driven online health estimation of Li-ion batteries using a novel energy-based health indicator. *IEEE Trans. Energy Convers.* **2020**, *35*, 1715–1718. https://doi.org/10.1109/TEC.2020.2994913.
128. Li, P.; Zhang, Z.; Grosu, R.; Deng, Z.; Hou, J.; Rong, Y.; Wu, R. An end-to-end neural network framework for state-of-health estimation and remaining useful life prediction of electric vehicle lithium batteries. *Renew. Sustain. Energy Rev.* **2022**, *156*, 111843. https://doi.org/10.1016/j.rser.2021.111843.
129. Park, K.; Choi, Y.; Choi, W.J.; Ryu, H.-Y.; Kim, H. LSTM-based battery remaining useful life prediction with multi-channel charging profiles. *IEEE Access* **2020**, *8*, 20786–20798. https://doi.org/10.1109/ACCESS.2020.2968939.
130. Wei, M.; Gu, H.; Ye, M.; Wang, Q.; Xu, X.; Wu, C. Remaining useful life prediction of lithium-ion batteries based on Monte Carlo dropout and gated recurrent unit. *Energy Rep.* **2021**, *7*, 2862–2871. https://doi.org/10.1016/j.egyr.2021.05.019.
131. Rouhi Ardeshiri, R.; Ma, C. Multivariate gated recurrent unit for battery remaining useful life prediction: A deep learning approach. *Int. J. Energy Res.* **2021**, *45*, 16633–16648. https://doi.org/10.1002/er.6910.
132. Zhang, L.; Xi, Z.; Ziebert, C. Remaining Useful Life Prediction for Lithium-Ion Batteries Based on Iterative Transfer Learning and Mogrifier LSTM. *Batteries* **2023**, *9*, 448. https://doi.org/10.3390/batteries9090448.
133. Tang, T.; Yuan, H. An indirect remaining useful life prognosis for Li-ion batteries based on health indicator and novel artificial neural network. *J. Energy Storage* **2022**, *52*, 104701. https://doi.org/10.1016/j.est.2022.104701.
134. Pugalenthi, K.; Park, H.; Hussain, S.; Raghavan, N. Remaining useful life prediction of lithium-ion batteries using neural networks with adaptive Bayesian learning. *Sensors* **2022**, *22*, 3803. https://doi.org/10.3390/s22103803.
135. Chen, L.; Ding, Y.; Liu, B.; Wu, S.; Wang, Y.; Pan, H. Remaining useful life prediction of lithium-ion battery using a novel particle filter framework with grey neural network. *Energy* **2022**, *244*, 122581. https://doi.org/10.1016/j.energy.2021.122581.
136. Qiu, J.-S.; Fan, Y.-C.; Wang, S.-L.; Yang, X.; Qiao, J.-L.; Liu, D.-L. Research on the remaining useful life prediction method of lithium-ion batteries based on aging feature extraction and multi-kernel relevance vector machine optimization model. *Int. J. Energy Res.* **2022**, *46*, 13931–13946. https://doi.org/10.1002/er.8110.
137. Li, S.; Fang, H.; Shi, B. Remaining useful life estimation of lithium-ion battery based on interacting multiple model particle filter and support vector regression. *Reliab. Eng. Syst. Saf.* **2021**, *210*, 107542. https://doi.org/10.1016/j.ress.2021.107542.
138. Li, X.; Ma, Y.; Zhu, J. An online dual filters RUL prediction method of lithium-ion battery based on unscented particle filter and least squares support vector machine. *Measurement* **2021**, *184*, 109935. https://doi.org/10.1016/j.measurement.2021.109935.
139. Yao, F.; He, W.; Wu, Y.; Ding, F.; Meng, D. Remaining useful life prediction of lithium-ion batteries using a hybrid model. *Energy* **2022**, *248*, 123622. https://doi.org/10.1016/j.energy.2022.123622.
140. Chen, Z.; Shi, N.; Ji, Y.; Niu, M.; Wang, Y. Lithium-ion batteries remaining useful life prediction based on BLS-RVM. *Energy* **2021**, *234*, 121269. https://doi.org/10.1016/j.energy.2021.121269.
141. Guo, W.; He, M. An optimal relevance vector machine with a modified degradation model for remaining useful lifetime prediction of lithium-ion batteries. *Appl. Soft Comput.* **2022**, *124*, 108967. https://doi.org/10.1016/j.asoc.2022.108967.
142. Zhou, Y.; Wang, S.; Xie, Y.; Shen, X.; Fernandez, C. Remaining useful life prediction and state of health diagnosis for lithium-ion batteries based on improved grey wolf optimization algorithm-deep extreme learning machine algorithm. *Energy* 2023, 128761. https://doi.org/10.1016/j.energy.2023.128761.
143. Zhou, Y.; Wang, S.; Xie, Y.; Zeng, J.; Fernandez, C. Remaining useful life prediction and state of health diagnosis of lithium-ion batteries with multiscale health features based on optimized CatBoost algorithm. *Energy* 2024, 131575. https://doi.org/10.1016/j.energy.2024.131575.
144. Xu, X.; Tang, S.; Yu, C.; Xie, J.; Han, X.; Ouyang, M. Remaining useful life prediction of lithium-ion batteries based on Wiener process under time-varying temperature condition. *Reliab. Eng. Syst. Saf.* **2021**, *214*, 107675. https://doi.org/10.1016/j.ress.2021.107675.
145. Ren, Y.; Tang, T.; Xia, Q.; Zhang, K.; Tian, J.; Hu, D.; Yang, D.; Sun, B.; Feng, Q.; Qian, C. A data and physical model joint driven method for lithium-ion battery remaining useful life prediction under complex dynamic conditions. *J. Energy Storage* 2024, 110065. https://doi.org/10.1016/j.est.2023.110065.
146. Zhang, J.; Huang, C.; Chow, M.-Y.; Li, X.; Tian, J.; Luo, H.; Yin, S. A data-model interactive remaining useful life prediction approach of lithium-ion batteries based on PF-BiGRU-TSAM. *IEEE Trans. Ind. Inform.* **2024**, *20*, 1144–1154. https://doi.org/10.1109/TII.2023.3266403.

147. Wang, S.; Wu, F.; Takyi-Aninakwa, P.; Fernandez, C.; Stroe, D.-I.; Huang, Q. Improved singular filtering-Gaussian process regression-long-short-term memory model for whole-life-cycle remaining capacity estimation of lithium-ion batteries adaptive to fast aging and multi-current variations. *Energy* **2023**, *284*, 128677. https://doi.org/10.1016/j.energy.2023.128677.
148. Duan, W.; Song, S.; Xiao, F.; Chen, Y.; Peng, S.; Song, C. Battery SOH estimation and RUL prediction framework based on variable forgetting factor online sequential extreme learning machine and particle filter. *J. Energy Storage* **2023**, 107322. https://doi.org/10.1016/j.est.2023.107322.
149. He, N.; Yang, Z.; Qian, C.; Li, R.; Gao, F.; Cheng, F. Remaining useful life prediction of lithium-ion battery based on fusion model considering capacity regeneration phenomenon. *J. Energy Storage* **2024**, 111068. https://doi.org/10.1016/j.est.2024.111068.
150. Li, K.; Xie, N.; Li, H. A hybrid grey approach for battery remaining useful life prediction considering capacity regeneration. *Expert Syst. Appl.* **2025**, 126905. https://doi.org/10.1016/j.eswa.2025.126905.
151. Pang, H.; Chen, K.; Geng, Y.; Wu, L.; Wang, F.; Liu, J. Accurate capacity and remaining useful life prediction of lithium-ion batteries based on improved particle swarm optimization and particle filter. *Energy* **2024**, *293*, 130555. https://doi.org/10.1016/j.energy.2024.130555.
152. Wu, J.; Kong, L.; Cheng, Z.; Yang, Y.; Zuo, H. RUL prediction for lithium batteries using a novel ensemble learning method. *Energy Rep.* **2022**, *8*, 313–326. https://doi.org/10.1016/j.egyr.2022.10.298.
153. Jafari, S.; Kim, J.; Choi, W.; Byun, Y.-C. Integrating multilayer perceptron and support vector regression for enhanced state of health estimation in lithium-ion batteries. *IEEE Access* **2025**, *13*, 11463–11478. https://doi.org/10.1109/ACCESS.2024.3497656.
154. Feng, J.; Cai, F.; Zhao, Y.; Zhang, X.; Zhan, X.; Wang, S. A novel feature optimization and ensemble learning method for state-of-health prediction of mining lithium-ion batteries. *Energy* **2024**, *299*, 131474. https://doi.org/10.1016/j.energy.2024.131474.
155. Fan, Q.; He, G.; Ruan, D.; Gühmann, C. Multi-expert fusion for state-of-health estimation of lithium-ion batteries. *Sci. Rep.* **2025**, *15*, 42058. https://doi.org/10.1038/s41598-025-26163-1.
156. Hong, J.; Li, K.; Liang, F.; Yang, H.; Zhang, C.; Yang, Q.; Wang, J. A novel state of health prediction method for battery system in real-world vehicles based on gated recurrent unit neural networks. *Energy* **2024**, *289*, 129918. https://doi.org/10.1016/j.energy.2023.129918.
157. Wang, X.; Hu, B.; Su, X.; Xu, L.; Zhu, D. State of health estimation for lithium-ion batteries using random forest and gated recurrent unit. *J. Energy Storage* **2024**, *76*, 109796. https://doi.org/10.1016/j.est.2023.109796.
158. Luo, K.; Zheng, H.; Shi, Z. A simple feature extraction method for estimating the whole life cycle state of health of lithium-ion batteries using transformer-based neural network. *J. Power Sources* **2023**, *576*, 233139. https://doi.org/10.1016/j.jpowsour.2023.233139.
159. Soo, Y.-Y.; Wang, Y.; Xiang, H.; Chen, Z. Machine learning based battery pack health prediction using real-world data. *Energy* 2024, 132856. https://doi.org/10.1016/j.energy.2024.132856.
160. Che, Y.; Deng, Z.; Li, P.; Tang, X.; Khosravinia, K.; Lin, X.; Hu, X. State of health prognostics for series battery packs: a universal deep learning method. *Energy* **2022**, *238*, 121857. https://doi.org/10.1016/j.energy.2021.121857.
161. Hu, X.; Che, Y.; Lin, X.; Deng, Z. Health prognosis for electric vehicle battery packs: a data-driven approach. *IEEE/ASME Trans. Mechatron.* **2020**, *25*, 2622–2632. https://doi.org/10.1109/TMECH.2020.2986364.
162. Che, Y.; Deng, Z.; Tang, X.; Lin, X.; Nie, X.; Hu, X. Lifetime and aging degradation prognostics for lithium-ion battery packs based on a cell to pack method. *Chin. J. Mech. Eng.* **2022**, *35*, 4. https://doi.org/10.1186/s10033-021-00668-y.
163. Wang, J.; Deng, Z.; Peng, K.; Deng, X.; Xu, L.; Guan, G.; Abudula, A. Early prognostics of lithium-ion battery pack health. *Sustainability* **2022**, *14*, 2313. https://doi.org/10.3390/su14042313.
164. Qiang, X.; Liu, W.; Lyu, Z.; Ruan, H.; Li, X. A data-fusion-model method for state of health estimation of Li-ion battery packs based on partial charging curve. *Green Energy Intell. Transp.* **2024**, *3*, 100169. https://doi.org/10.1016/j.geits.2024.100169.
165. Dos Reis, G.; Strange, C.; Yadav, M.; Li, S. Lithium-ion battery data and where to find it. *Energy AI* **2021**, *5*, 100081. https://doi.org/10.1016/j.egyai.2021.100081.
166. Wu, B.; Widanage, W.D.; Yang, S.; Liu, X. Battery digital twins: Perspectives on the fusion of models, data and artificial intelligence for smart battery management systems. *Energy AI* **2020**, *1*, 100016. https://doi.org/10.1016/j.egyai.2020.100016.
167. Ren, J.; Ma, J.; Wang, H.; Yu, T.; Wang, K. A comprehensive review on research methods for lithium-ion battery of state of health estimation and end of life prediction: Methods, properties, and prospects. *Prot. Control Mod. Power Syst.* **2025**, *10*, 146–165. https://doi.org/10.23919/PCMP.2024.000211.
168. Christophersen, J.P. Battery Test Manual for Electric Vehicles, Revision 3; Idaho National Laboratory: Idaho Falls, ID, USA, 2015; INL/EXT-15-34184. https://doi.org/10.2172/1186745.

169. IEC 62660-1:2018; Secondary Lithium-Ion Cells for the Propulsion of Electric Road Vehicles—Part 1: Performance Testing. International Electrotechnical Commission (IEC): Geneva, Switzerland, 2018. Available online: https://webstore.iec.ch/en/publication/28965 (accessed on 05 November 2025).
170. IEC 62660-3:2022; Secondary Lithium-Ion Cells for the Propulsion of Electric Road Vehicles—Part 3: Safety Requirements. International Electrotechnical Commission (IEC): Geneva, Switzerland, 2022. Available online: https://webstore.iec.ch/en/publication/65084 (accessed on 12 February 2026).
171. Sui, X.; He, S.; Vilsen, S.B.; Meng, J.; Teodorescu, R.; Stroe, D.I. A review of non-probabilistic machine learning-based state of health estimation techniques for Lithium-ion battery. *Appl. Energy* **2021**, *300*, 117346. https://doi.org/10.1016/j.apenergy.2021.117346.
172. Urquizo, J.; Singh, P. A review of health estimation methods for Lithium-ion batteries in Electric Vehicles and their relevance for Battery Energy Storage Systems. *J. Energy Storage* **2023**, *73*, 109194. https://doi.org/10.1016/j.est.2023.109194.
173. von Bülow, F.; Meisen, T. A review on methods for state of health forecasting of lithium-ion batteries applicable in real-world operational conditions. *J. Energy Storage* **2023**, *57*, 105978. https://doi.org/10.1016/j.est.2022.105978.
174. Nazim, M.S.; Chakma, A.; Joha, M.I.; Alam, S.S.; Rahman, M.M.; Umam, M.K.S.; Jang, Y.M. Artificial intelligence for estimating State of Health and Remaining Useful Life of EV batteries: A systematic review. *ICT Express* **2025**, *11*, 769–789. https://doi.org/10.1016/j.icte.2025.05.013.
175. Mbagaya, L.; Reddy, K.; Botes, A. Machine learning techniques for battery state of health prediction: A comparative review. *World Electr. Veh. J.* **2025**, *16*, 594. https://doi.org/10.3390/wevj16110594.
176. Xiang, Y.; Chauhan, D.; Srinivasan, D. Battery state-of-health estimation with embedded impedance spectrum features under multiple battery chemistry and temperature conditions. *Batteries* **2026**, *12*, 77. https://doi.org/10.3390/batteries12020077.