



Inline quality grading of commercial lithium-ion battery manufacturing via data-efficient learning and transferable assessment

Downloaded from: <https://research.chalmers.se>, 2026-09-12 09:42 UTC

Citation for the original published paper (version of record):

Liang, C., Tao, S., Xia, C. et al (2026). Inline quality grading of commercial lithium-ion battery manufacturing via data-efficient learning and transferable assessment. *Energy and Environmental Sciences*, 19(16): 5340-5354. <http://dx.doi.org/10.1039/d6ee02209j>

N.B. When citing this work, cite the original published paper.

Energy & Environmental Science

Volume 19
Number 16
18 August 2026
Pages 5195–5554

rsc.li/ees



ISSN 1754-5706

PAPER

Shengyu Tao, Ziyang Lyu, Guangmin Zhou, Huadong Mo *et al.*
Inline quality grading of commercial lithium-ion battery
manufacturing *via* data-efficient learning and transferable
assessment

PAPER

Published on 17 June 2026
Licensed under CC-BY 4.0Cite this: *Energy Environ. Sci.*,
2026, 19, 5340Inline quality grading of commercial lithium-ion
battery manufacturing *via* data-efficient learning
and transferable assessmentChen Liang,^a Shengyu Tao,^b Chunqiu Xia,^c Xinghao Huang,^b Hang Hu,^d
Rui Wang,^e Daoyi Dong,^f Ziyang Lyu,^{*a} Guangmin Zhou^{*b} and
Huadong Mo^{*g}

Lithium-ion battery manufacturing involves a complex sequence of tightly coupled processes, making reliable quality grading essential for ensuring cell consistency, production efficiency, and product reliability. However, existing grading paradigms rely heavily on long-cycle testing and dense labeling, resulting in energy consumption, time cost, and limited scalability. Here, we propose the data-efficient learning and transferable assessment (DELTA) framework, which combines feature extraction and semi-supervised consistency classification for early quality grading in manufacturing, using cycle-life at end of life (EOL) as the quality evaluation metric. The framework evaluates data from 6 publicly available datasets, encompassing 421 cells with 3 battery chemistries, 6 charging rates, 6 temperatures, and 8 rated capacities. A linear mixed-effects model extracts static features to quantify material effects, while dynamic features from pre-cycling tests characterize performance stability. A semi-supervised classifier based on Gaussian mixture models with an entropy-driven mechanism simulates the absence of true manufacturing labels. Experimental results show that DELTA achieves over 83% classification accuracy of cycle-life at EOL with only 30% labeled data, outperforming state-of-the-art methods such as FixMatch and UDA, while reducing training time by 50%. It maintains more than 95% accuracy on unseen datasets, enabling fast, low-cost, and scalable battery screening in manufacturing, establishing missingness-aware learning as a practical solution for data-limited manufacturing environments. A preliminary scenario analysis suggests that reducing reliance on long-cycle screening could potentially lower testing-related costs and energy consumption in large-scale battery production.

Received 6th April 2026,
Accepted 16th June 2026

DOI: 10.1039/d6ee02209j

rsc.li/ees

Broader context

Lithium-ion battery manufacturing involves a complex sequence of tightly coupled processes, where ensuring product quality and consistency is essential for both economic competitiveness and sustainable production. Among these processes, activation and testing dominate both energy consumption and production time, as they rely on prolonged cycling to verify product quality and consistency. This existing long-term cycling-based quality grading incurs substantial energy use and associated carbon emissions, increasingly conflicting with the demand for sustainable, low-carbon battery manufacturing. In addition, manufacturing-induced cell-to-cell variability can lead to substantial differences in long-term degradation behavior, reducing pack-level performance, accelerating ageing, and potentially increasing safety risks. Here, we present a data-efficient framework that enables rapid and reliable inline battery quality grading in a large-scale battery manufacturing context using minimal cycle data. By explicitly accounting for label availability in manufacturing environments, the proposed approach effectively leverages abundant unlabeled samples to enhance the robustness and generalizability of quality grading. The framework reduces reliance on costly long-term cycling tests while utilizing routinely collected manufacturing data, enabling faster and more scalable quality assessment. The method demonstrates consistent performance across diverse materials and operating conditions, offering a scalable solution for energy-efficient and environmentally sustainable battery manufacturing and demonstrating a billion USD market assuming a terawatt battery manufacturing scale.

^a School of Mathematics and Statistics, University of New South Wales, Sydney, NSW 2052, Australia. E-mail: ziyang.lyu@unsw.edu.au^b Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, 518055, China. E-mail: shengyu.tao@chalmers.se, guangminzhou@sz.tsinghua.edu.cn^c Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, 41296, Sweden^d Department of Mechanical Engineering, Tsinghua University, Beijing, 100084, China^e Faculty of Mechanical Engineering and Mechanics, Ningbo University, Ningbo, 315211, China^f Australian Artificial Intelligence Institute, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia^g School of Systems and Computing, University of New South Wales, Canberra, ACT 2610, Australia. E-mail: huadong.mo@unsw.edu.au

† These authors contributed equally to this article.

Introduction

Lithium-ion batteries have become a core technology for transportation electrification and large-scale energy storage as the global transition toward low-carbon energy systems accelerates.^{1–4} In line with development roadmaps such as Battery 2030+, battery development is expected not only to improve energy density and cycle life, but also to ensure high consistency, safety, and a low carbon footprint under large-scale manufacturing.⁵ These requirements are driving the industry from empirical-based production toward Industry 4.0, characterized by data-driven and artificial intelligence.⁶

Lithium-ion batteries are highly complex systems whose performance is jointly influenced by material chemistry, cell architecture, manufacturing processes, and operating conditions.^{7,8} Lithium-ion battery manufacturing involves a sequence of tightly coupled processes, including slurry preparation, coating, drying, calendaring, slitting, cell assembly, electrolyte filling, formation, aging, and capacity grading. Variations introduced at any stage can propagate through subsequent processes and ultimately affect cell consistency, safety, and lifetime performance. Consequently, effective quality control throughout the manufacturing workflow is essential for ensuring product reliability and reducing production losses. To achieve this goal, advanced characterization techniques, including electron microscopy,⁹ X-ray,^{10–12} optical and infrared imaging,^{13–15} and ultrasonic scanning,^{16–18} as well as multi-physics modeling approaches,^{19–22} have been widely employed. Although advanced characterization techniques and multi-physics modeling have significantly improved the understanding of battery degradation mechanisms, these approaches often require specialized equipment, extensive calibration, or substantial computational resources, limiting their applicability to large-scale manufacturing. In practice, rapid quality assessment is more valuable than detailed mechanistic analysis because it enables early identification of abnormal cells and supports timely process optimization.

Among manufacturing stages, activation and testing are the most time- and energy-intensive and critically determine cell consistency and delivery quality.^{23,24} However, industrial quality evaluation during formation, aging, and capacity grading primarily relies on capacity, internal resistance, and self-discharge measurements. These indicators reflect the current state of a cell but provide limited information about its future degradation trajectory. As a result, cells that pass conventional grading criteria and exhibit similar initial performance may still experience markedly different lifetime evolution during operation, ultimately affecting pack-level consistency and reliability. This limitation motivates the development of rapid quality assessment methods that can leverage routine early-cycle testing data to identify latent quality differences before cells are deployed.

Although several studies have applied data-driven methods to early battery degradation classification,^{25–30} challenges remain before such methods can be effectively deployed in manufacturing environments. First, many existing approaches

adopt a two-stage strategy that predicts numerical lifetime metrics and subsequently converts them into degradation categories.³¹ While effective for lifetime forecasting, this approach introduces additional computational complexity and provides limited direct value for manufacturing-oriented decision-making, where rapid quality screening is often more important than precise lifetime estimation. Second, current methods rely heavily on scarce and costly labeled data, whereas large volumes of early-cycle data generated during manufacturing and testing typically remain unlabeled, fundamentally constraining scalability and economic feasibility. More critically, the diversity of battery chemistries and operating conditions gives rise to highly heterogeneous early degradation behaviors, causing models trained on fixed datasets to exhibit poor generalization across materials and usage scenarios.³² In addition, deep learning models often involve large parameter sizes and high inference costs, further hindering deployment in resource-constrained manufacturing environments.³³ Semi-supervised learning offers a promising pathway to leverage abundant unlabeled data and alleviate label scarcity.^{34–38} However, existing statistical models, such as Gaussian mixture models (GMMs), generally do not explicitly account for information associated with label availability.^{39,40} In practice, complete lifetime labels are available only for a small subset of cells that undergo costly long-term ageing tests, whereas the majority of production cells are released after routine testing. As a result, label availability is governed by operational, economic, and testing constraints rather than by a random sampling process. This systematic selection mechanism can introduce distribution mismatch between labeled and unlabeled populations,^{41,42} motivating explicit consideration of the label-generation process when developing manufacturing-oriented battery quality assessment models.

In this study, we propose data-efficient learning and transferable assessment (DELTA), a missingness-aware semi-supervised framework for manufacturing-oriented battery quality assessment under limited label availability. Rather than estimating exact lifetime values, DELTA utilizes early-cycle signatures to identify relative lifetime categories that serve as indicators of manufacturing-related quality variation. As shown in Fig. 1a, the dataset consists of a small set of labeled and abundant unlabeled early-cycle data. DELTA combines the linear mixed-effects (LME) model with a missing-label-aware GMM to achieve direct ternary EOL classification. Fig. 1b illustrates the complete flow of DELTA. Labels are defined using the $\mu \pm 1\sigma$ criteria to identify low-, medium-, and high-performance cells, supporting manufacturing-oriented screening in which extreme cells require further processing. The entropy term does not assume knowledge of the true label-generation mechanism. Instead, it serves as a proxy for classification uncertainty. Samples with higher entropy exhibit more ambiguous class membership and therefore contribute less confidently to the estimation of class distributions. By incorporating entropy into the missing-label mechanism, the model accounts for potential differences between labeled and unlabeled samples and reduces the bias introduced by limited label availability.

Importantly, the contribution of this work is not the introduction of a new battery lifetime classification task. Instead,

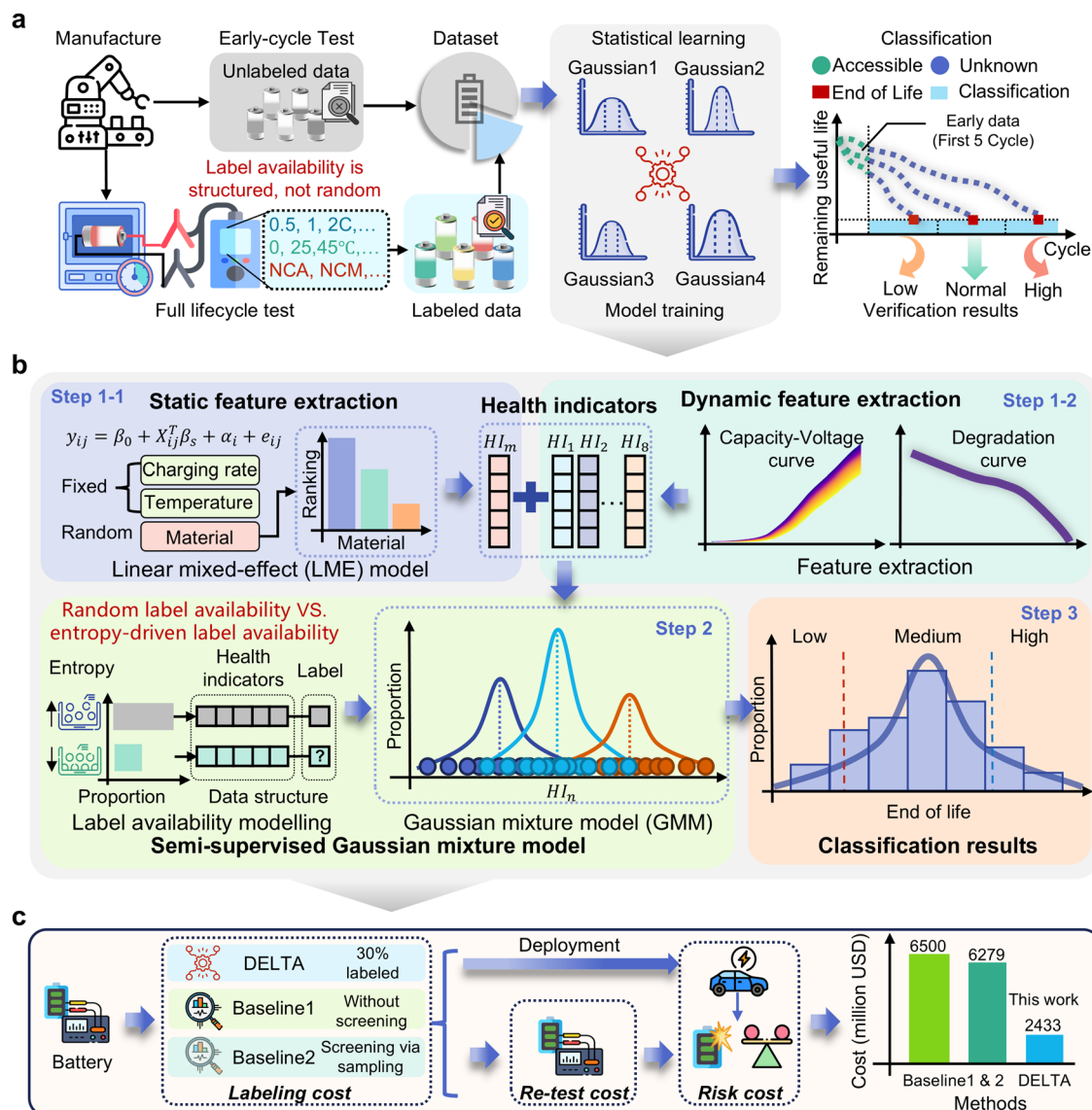


Fig. 1 Model motivation, model architecture, and model deployment. (a) Utilize a small amount of labeled and a large amount of unlabeled early data to achieve ternary classification of cycle life at EOL based on statistical learning. (b) The DELTA framework primarily consists of feature extraction and semi-supervised classification modules. In Step 1, the LME model is responsible for feature extraction. In Step 1-1, under fixed temperature and charge rate conditions, it estimates random effects at the material level to obtain static features. In Step 1-2, dynamic features are extracted from the charging data. These features are then fed into the semi-supervised classification model in Step 2, which explicitly accounts for label availability, and the GMM is used in Step 3 to achieve semi-supervised classification. (c) Economic analysis of the proposed deployment framework, comparing labeling cost, re-test cost, and potential risk cost among different strategies. The results highlight that DELTA can significantly reduce overall deployment cost while maintaining strong predictive performance.

the key contribution lies in demonstrating that reliable, manufacturing-oriented quality assessment can be achieved under realistic industrial conditions in which lifetime labels are scarce, but early-cycle testing data are abundant. By integrating physics-informed feature extraction, missingness-aware semi-supervised learning, and cross-material transferability, DELTA provides a practical framework for scalable battery quality grading. Because early-cycle electrochemical behavior preserves information established during manufacturing, DELTA leverages routine formation and capacity-grading data to identify cell-to-cell variability without requiring direct process

parameters. The framework enables early identification of quality-risk cells and provides a practical pathway toward intelligent battery manufacturing. As shown in Fig. 1c, large-scale deployment of DELTA has the potential to reduce testing costs, improve production efficiency, and enhance battery consistency management, highlighting its practical value for battery manufacturing. Overall, the framework demonstrates that incorporating a missing-label mechanism-aware semi-supervised learning approach into battery manufacturing enables early identification of extreme cells and reduces training time and energy consumption.

Results

Dataset curation

The input data used in this work originate from routine formation and capacity-grading procedures performed before cell deployment (see SI1 for detailed information). These early-cycle measurements are generated during standard end-of-line manufacturing tests and therefore represent manufacturing-end testing data rather than long-term field operation data. Consequently, the proposed framework links manufacturing-end testing signatures to subsequent lifetime outcomes. In this sense, the framework performs outcome-oriented quality assessment rather than direct monitoring of upstream manufacturing process parameters.

To evaluate its performance and cross-domain generalization capability, as shown in Table 1, experimental data from 6 publicly available battery datasets (CAS,⁴³ MICH,⁴⁴ RWTH,⁴⁵ Stanford,⁴⁶ TJU,⁴⁷ XJTU⁴⁸) were used, comprising a total of 421 cells. These data span covering 3 material types, 6 charging rates, 6 temperature settings, 8 rated capacities and 8 cut-off voltages (see SI2 for detailed information), covering a wide range of operating and design conditions. This diversity mimics real cell-to-cell variability and provides a realistic testbed for evaluating model robustness and transferability.

Battery degradation is a gradual and continuous process rather than a discrete event. The electrochemical state of a cell during its first few operational cycles remains highly consistent with the state established at the completion of formation, aging, and capacity grading. As a result, early-cycle performance can be viewed as a direct continuation of the manufacturing-end health state. Although direct manufacturing process parameters are unavailable, the early-cycle data analyzed in this study preserve the cumulative effects of manufacturing and end-of-line testing processes. Therefore, these data provide a practical basis for outcome-oriented assessment of manufacturing quality assessment by revealing latent quality differences that may not be captured by conventional grading metrics such as capacity and internal resistance alone.

Two classification criteria were established in this study: a batch-based criterion and a material-based criterion. The batch-based criterion groups cells originating from the same production batch and assigns corresponding batch labels (*b_label*), with the objective of identifying relative lifetime differences within a batch. The material-based criterion groups cells according to their material systems and assigns material labels (*m_label*), with the objective of evaluating cross-material lifetime classification and generalization across different battery chemistries.

The purpose of the L/N/H categorization is not to compare absolute lifetimes across different battery populations. Instead, it is designed to identify cells whose lifetime performance significantly deviates from the expected behaviour of a given reference population. Consequently, a cell classified as H in one population may exhibit a shorter absolute lifetime than a cell classified as N in another population. The classification reflects relative performance within the corresponding population rather than absolute lifetime magnitude.

Based on these two criteria, a batch-based dataset (*batch_dataset*) and a material-based dataset (*material_dataset*) were constructed, respectively. Through this dual-grouping strategy, the robustness of the DELTA framework was systematically evaluated from both the batch and material perspectives. Detailed cell specifications for each dataset are provided in Tables S1 and S2, with representative ageing trajectories shown in Fig. S1 and S2. In addition, two sub-datasets were constructed from the batch-based dataset for evaluation (see Table S3 for details): the base dataset and the extended dataset. The extended dataset simulates the incorporation of newly collected data during real-world deployment, enabling the assessment of the framework's generalization capability and scalability under practical conditions.

The resulting *batch_dataset* exhibits pronounced imbalance across different factor groups (Fig. 2a). Comparison of the EOL distributions across datasets reveals substantial inter-dataset discrepancies (Fig. 2b), indicating that battery ageing remains influenced by unquantified stochastic factors and environmental

Table 1 Description of datasets

Datasets	Material	Chemical formula	Q_n (Ah)	Cut-off voltage (V)	Temperature (°C)	Charge/discharge rate (C)	N
CAS	NCA	—	3.35	2.65–4.2	25	1/3, 1.5/3, 2/3	20
	NCA	—	3.35	2.65–4.2	25	1/3, 1.5/3, 2/3	30
	NCM	—	2.6	2.75–4.2	0, 25	1/3, 2/3	29
	LFP	—	1.5	2.0–4.0	25	1/3, 1.5/3, 2/3	60
MICH	NCM	$\text{LiNi}_{1/3}\text{Co}_{1/3}\text{Mn}_{1/3}\text{O}_2$	2.36	3.0–4.2	25, 45	1/1	40
RWTH	NCM	—	1.11	3.5–3.9	25	2/2	48
Stanford	NCM	$\text{LiNi}_{0.5}\text{Mn}_{0.3}\text{Co}_{0.2}\text{O}_2$	0.24	3–4.4	30	1/0.75	41
TJU	NCA	$\text{Li}_{0.86}\text{Ni}_{0.86}\text{Co}_{0.11}\text{Al}_{0.03}\text{O}_2$	3.5	2.65–4.2	25, 35, 45	0.25/1, 0.5/1, 1/1	66
	NCM	$\text{Li}_{0.86}\text{Ni}_{0.86}\text{Co}_{0.11}\text{Mn}_{0.07}\text{O}_2$	3.5	2.5–4.2	25, 35, 45	0.5/1	55
	NCA + NCM	—	3.5	—	25	0.5/1, 0.5/2, 0.5/4	9
XJTU	NCM	$\text{LiNi}_{0.5}\text{Co}_{0.2}\text{Mn}_{0.3}\text{O}_2$	2.0	2.5–4.2	20	2/1, 3/1	23

Note: “—” denotes that the battery chemistry is not specified. N represents the number of cells.

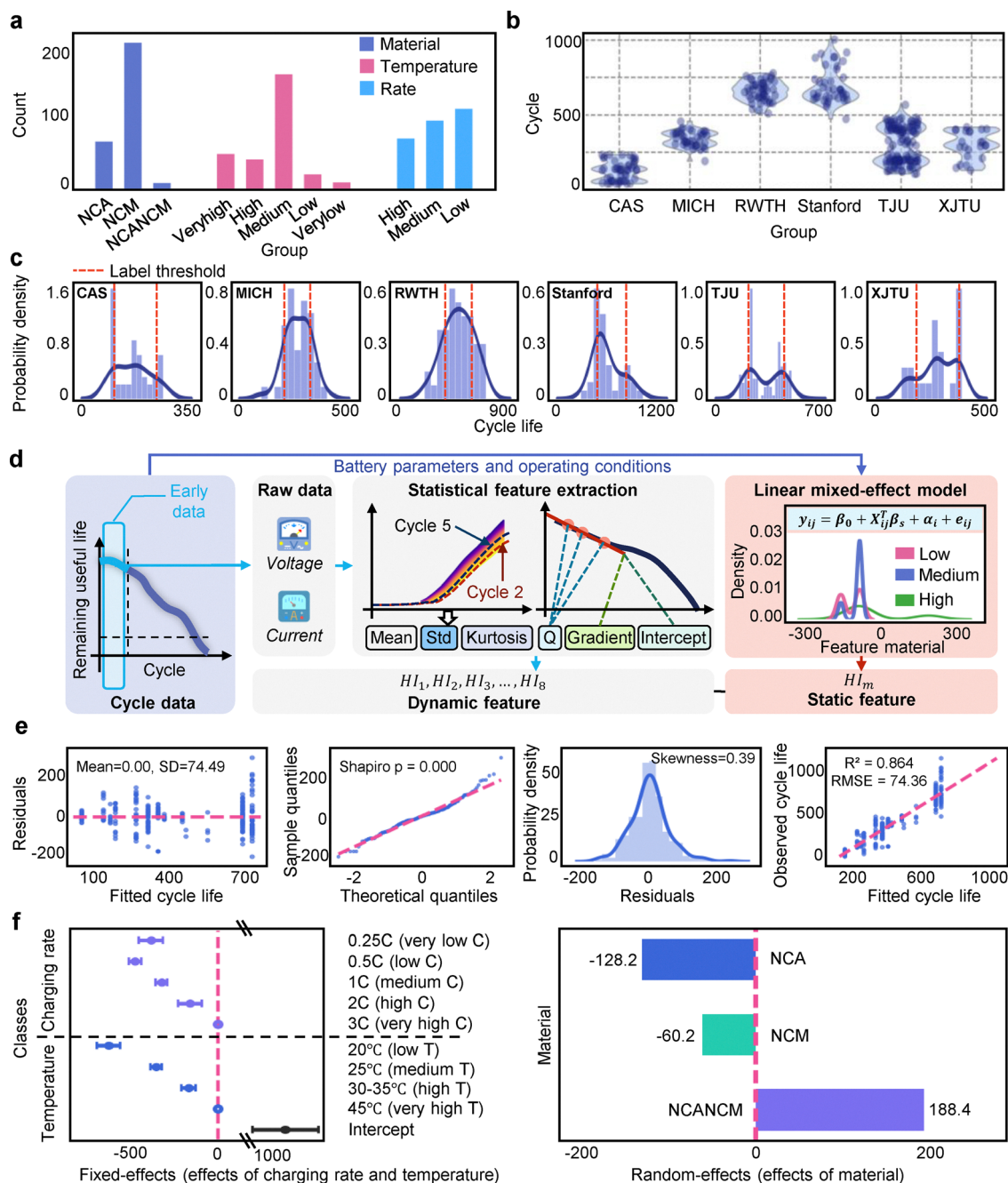


Fig. 2 Dataset characterization and feature construction. (a) Battery samples grouped by material type (NCA, NCANCM, NCM), operating temperature (very low, low, medium, high, very high), and charging rate (high, medium, low), showing the distribution of cells across experimental conditions. (b) Cycle-life distributions of batteries from six datasets (CAS, MICH, RWTH, Stanford, TJU, and XJTU), where each point represents an individual cell and the violin plots illustrate overall variability. (c) Cycle-life probability distributions within each dataset and the corresponding label thresholds used for classification. (d) Workflow for feature construction from early-cycle charging data. Raw voltage and current data are first collected, followed by statistical feature extraction (e.g., mean, standard deviation, kurtosis). Dynamic features are further derived from early-cycle trends, and a linear mixed-effect model is used to obtain both dynamic features $HI_1, \dots, HI_8$ and static material-related features. (e) Performance evaluation of the LME model through residual analysis and observed-fitted comparisons, indicating good agreement between predicted and observed cycle life. (f) The estimated model effects, including fixed effects representing the influence of charging rate and temperature, and random effects reflecting differences among battery materials.

variations even under nominally identical operating conditions. Cycle life labels were defined using a 3-class scheme based on the 1σ boundaries of a Gaussian distribution, reflecting the

approximately normal distribution of cycle life observed in practice (Fig. 2c). Further details are provided in SI3, and additional information on the *material_dataset* is presented in Fig. S3.

Feature extraction and construction

To enhance cross-material generalization, both dynamic and static features are extracted during the feature extraction (Fig. 2d). Dynamic features are extracted from charging data, including Q - V curves and capacity degradation trajectories (Fig. S1 and S2), from which health indicators (HIs) are computed. Sixteen candidate HIs (details in Table S4) are initially constructed, and the eight most strongly correlated with EOL cycle-life are retained. All dynamic HIs are normalized by rated capacity to eliminate the influence of nominal capacity differences across cells. This enables the model to generalize across materials with inherently different capacities.

The input features used in this work are extracted from early-cycle voltage, capacity, and Q - V signals. These signals correspond to routine formation/grading-stage testing data, which are typically collected before cells are assembled into battery packs. They are not upstream process parameters, but they provide observable end-of-line signatures of the final cell state after manufacturing.

Material, temperature, and charging rate are identified as the 3 dominant factors affecting battery ageing, with material being the most influential. Subsequently, in the LME model, material is treated as a random effect, while temperature and charge rate are considered fixed effects to quantify material-dependent ageing behavior. Since these factors are discrete variables, they were grouped accordingly, with detailed definitions provided for *batch_dataset* in Tables S5–S7 and *material_dataset* in Tables S8 and S9. Using LME, the material-dependent contribution to EOL cycle-life is quantified, and the estimated best linear unbiased predictors (EBLUPs) of the random effects are extracted as an additional feature HI_m . The distributions of the extracted HIs exhibit bimodal characteristics (see Fig. S4 and S5 for detailed information). To address this, K-means clustering is applied to GMM-based classification to separate the two modes.

The diagnostic results of the LME model (Fig. 2e and f) indicate that the model assumptions are reasonably satisfied and that the model effectively captures the overall variation in cycle life. The fixed-effects estimates (Fig. 2f and Tables S10, S11) show that both temperature and charging rate significantly influence cycle life, while the random effects quantify substantial material-dependent variability. After accounting for operating conditions, the NCANCM system exhibits a higher baseline cycle life, whereas NCA and NCM show negative deviations. These results confirm that material effects dominate variability in ageing behaviour, supporting the inclusion of material-derived features for robust EOL grouping. Additional results for the *material_dataset* are provided in Fig. S6, S7 and Tables S12, S13.

Performance of the DELTA framework

To evaluate the performance of the DELTA framework, we first conducted tenfold cross-validation on the base dataset. To investigate the influence of label availability, ten missingness gradients were designed, with label visibility ratios ranging

from 0.1 to 1.0. The overall results obtained on the base dataset are shown in Fig. 3.

The dependence of DELTA's overall performance on label availability is summarized in Fig. 3a, b and Tables S14, S15 for the *batch_dataset* and *material_dataset*, respectively. As label visibility increases from 0.1 to 1.0, classification accuracy improves from ~ 0.6 to near 1.0, with reduced variability (Fig. 3a and b). Notably, DELTA remains effective even when only 30% of the samples are labeled, maintaining accuracy above 0.8 once label visibility exceeds 0.5. The inset plots show that recall closely follows accuracy, indicating a strong positive correlation. Overall, increasing label availability consistently improves performance across both datasets.

To gain deeper insight into model behaviour under fewer labels, more detailed analyses are presented in Fig. 3c–e. Fig. 3c shows the confusion matrices obtained under different label availability ratios, where the upper row corresponds to the *batch_dataset* and the lower row to the *material_dataset*. As label visibility gradually increases, the diagonal elements become increasingly dominant while the off-diagonal misclassification rates decrease steadily, indicating progressively improved discriminative capability across classes. Notably, even with under 30% label visibility, about 80% of samples remain correctly classified, demonstrating the robustness of DELTA in highly missing-data scenarios.

Fig. 3d and e summarize the class-wise performance for both datasets. Accuracy and F1 scores for all three categories (L, N, and H) improve with increasing label availability. The H category remains highly stable, maintaining an accuracy of 0.71 even at a missingness ratio of 0.9. In contrast, the N category is the most sensitive to reduced label availability, with its F1 score decreasing from 0.69 to 0.48 as label availability declines from 50% to 10%. The L category shows intermediate behaviour, with its F1 score remaining around 0.56 under high missingness. This disparity arises because the N category corresponds to intermediate cycle life, where degradation patterns are less distinct and often lie near decision boundaries between short- and long-lifetime groups. In addition, these samples tend to exhibit greater variability due to multiple influencing factors, making them harder to model and classify under limited supervision.

The robustness of DELTA under noisy conditions is further evaluated in Fig. 3f, g and Tables S16, S17 for the *batch_dataset* and *material_dataset*, respectively. Specifically, we introduce feature noise by adding zero-mean Gaussian perturbations to the normalized input features (standard deviation = 0.03 or 0.05) and introduce label noise by randomly flipping a proportion of the available training labels (noise ratio = 0.03 or 0.05) to other classes. A mixed setting combining feature noise (0.03) and label noise (0.03) is also considered, while the clean setting serves as the baseline. Overall, the proposed framework maintains strong classification performance across different noise settings, and the F1 score consistently increases as label availability increases. Although noise slightly degrades performance at very low label availability, the model quickly recovers as more labeled data become available. These results further

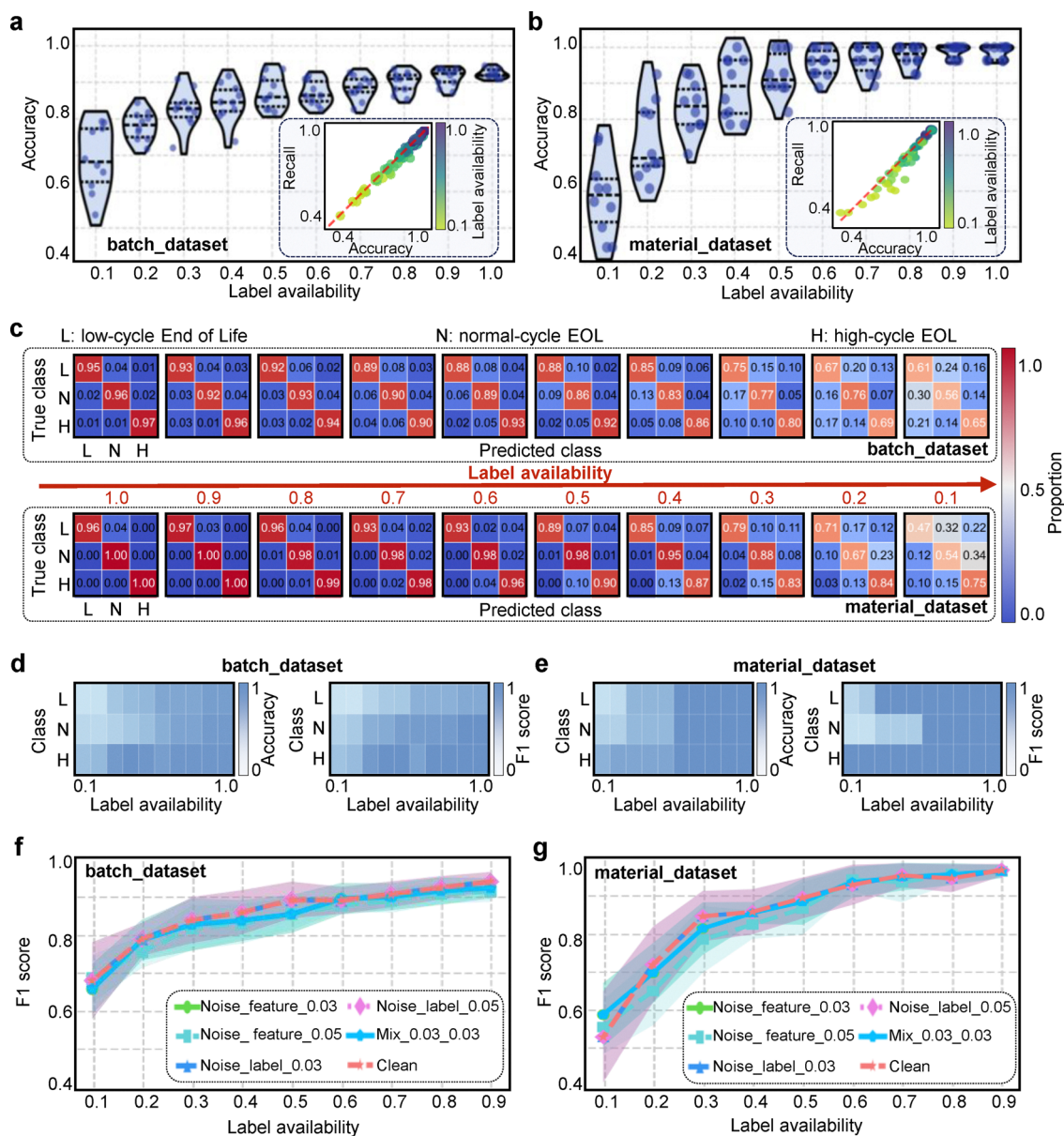


Fig. 3 Performance of DELTA under varying label availability on the base datasets. (a) and (b) Overall classification accuracy of DELTA on the *batch_dataset* (a) and *material_dataset* (b) under different label availability ratios. Violin plots illustrate the distribution of accuracy across multiple runs, while the inset scatter plots show the relationship between recall and accuracy, with color indicating the proportion of labeled data. As label availability increases, model performance improves and becomes more stable. (c) Confusion matrices under representative label availability ratios for both datasets. The upper row corresponds to *batch_dataset*, while the lower row corresponds to *material_dataset*. The results show that higher label availability leads to clearer class separation and improved prediction accuracy across low-cycle (L), normal-cycle (N), and high-cycle (H) EOL classes. (d) and (e) Class-wise accuracy and F1 scores as functions of label availability for *batch_dataset* (d) and *material_dataset* (e), demonstrating consistent performance gains across classes as more labeled data become available. (f) and (g) Robustness evaluation under different label-noise settings on *batch_dataset* (f) and *material_dataset* (g). The F1-score curves indicate that DELTA maintains stable performance even in the presence of feature noise and label noise, highlighting its robustness in low-label and noisy scenarios.

demonstrate the robustness and stability of DELTA under realistic conditions characterized by limited label availability and noisy observations.

Comparison with baseline methods

Fig. 4 presents a comprehensive comparison between the proposed DELTA framework and several mainstream semi-supervised

learning methods, including both modern deep semi-supervised learning approaches (FixMatch,⁴⁹ UDA,⁵⁰ SPRED⁵¹) and classical self-training paradigms (ST-SVM, ST-RF),⁵² providing a comprehensive benchmark across two datasets (*batch_dataset* and *material_dataset*). These baselines were selected to represent the major methodological families in semi-supervised learning. Specifically, FixMatch and UDA are widely adopted

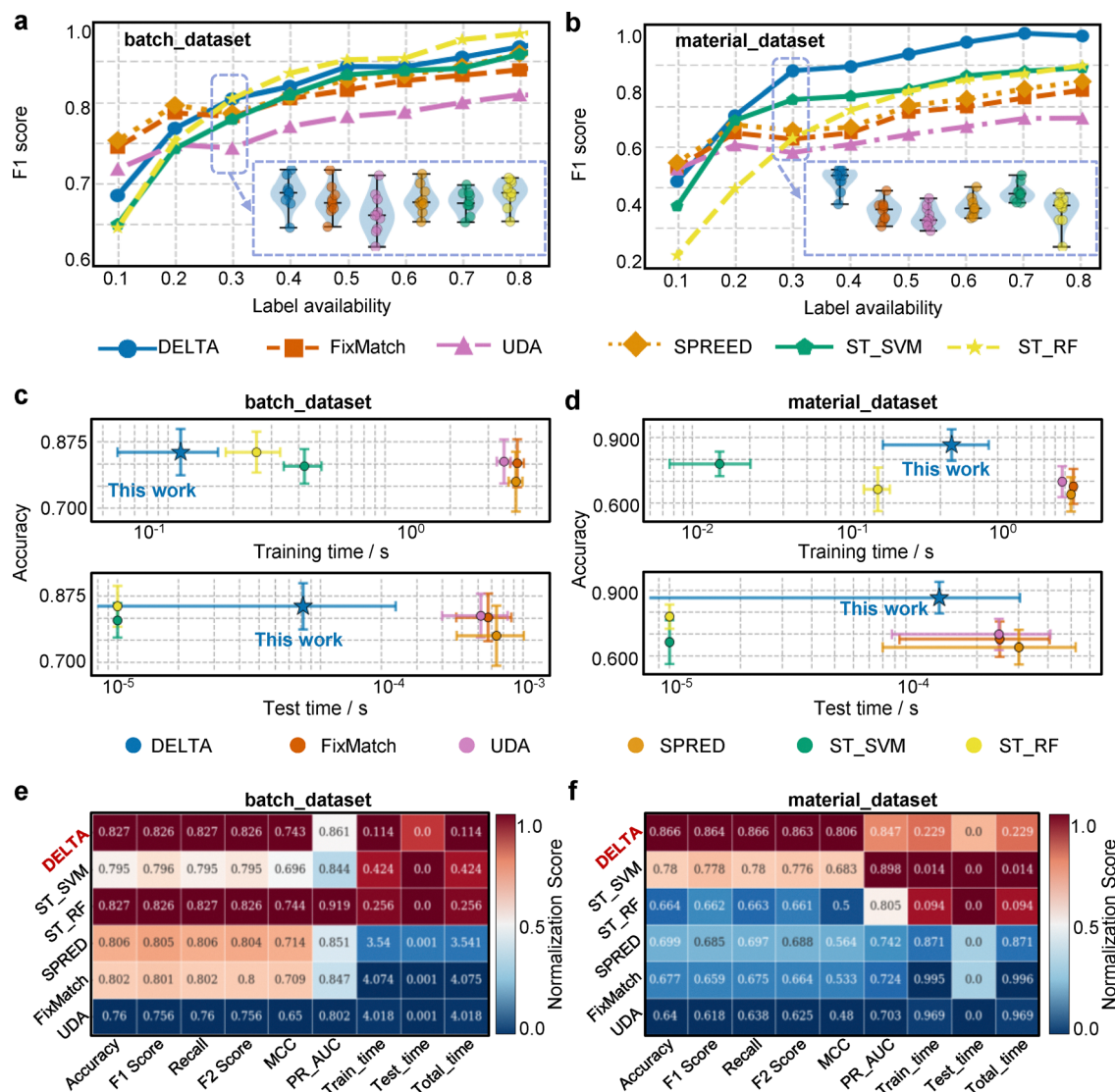


Fig. 4 Performance, stability, and computational efficiency comparison between DELTA and baseline semi-supervised methods. (a) and (b) F1-score comparison under different label availability ratios on the *batch_dataset* (a) and *material_dataset* (b). DELTA consistently achieves higher or comparable performance across most label availability settings, especially in low-label regimes. The inset violin plots illustrate the distribution and stability of results across repeated runs at 0.3 label availability, indicating that DELTA maintains a high F1 score compared with other semi-supervised baselines such as FixMatch, UDA, and SPRED. (c) and (d) Computational efficiency comparison during training and inference on the *batch_dataset* (c) and *material_dataset* (d). Each point represents the trade-off between predictive accuracy and computational time. DELTA demonstrates competitive accuracy while maintaining moderate training and testing time, highlighting its practical efficiency for real-world applications. (e) and (f) Normalized multi-metric comparison of all methods on the *batch_dataset* (e) and *material_dataset* (f), including accuracy, F1 score, recall, F2 score, MCC, PR-AUC, and computational cost metrics. The results show that DELTA achieves strong overall performance across multiple evaluation criteria, demonstrating balanced predictive capability and efficiency compared with other baseline approaches.

consistency-regularization and pseudo-labeling methods, SPRED represents recent prototype-based semi-supervised learning, while ST-SVM and ST-RF serve as representative self-training approaches commonly used in practical machine-learning applications. The comparison evaluates predictive performance, robustness under varying label availability, and computational efficiency across both *batch_dataset* and *material_dataset*.

Fig. 4a, b and Tables S18, S19 show the predictive performance under different label availability ratios. Overall, DELTA

maintains competitive or superior F1 scores across both datasets, particularly under limited-label conditions. The strong performance of DELTA can be attributed to the combination of LME-based feature extraction and probabilistic semi-supervised learning. Specifically, the LME model quantifies the effects of material systems and operating conditions on battery lifetime and converts this information into statistical descriptors that are subsequently used as inputs to the GMM classifier. This enables DELTA to incorporate material- and condition-dependent lifetime information into the classification process.

In addition, unlike conventional semi-supervised approaches that primarily rely on pseudo-label propagation, DELTA exploits the underlying statistical structure of both labeled and unlabeled samples through probabilistic modeling. As a result, the framework remains relatively robust as label availability decreases.

Fig. 4c, d and Tables S20, S21 compare the computational efficiency of different methods. Deep semi-supervised approaches generally require substantially longer training times due to their large parameter space and iterative optimization procedures. In contrast, DELTA relies on LME estimation and GMM inference, both of which are computationally lightweight. Consequently, DELTA achieves competitive predictive performance while maintaining training and inference times that are orders of magnitude lower than those of deep-learning-based alternatives. This efficiency is particularly important for manufacturing applications, where models may need to be retrained or updated frequently as new production data become available.

Fig. 4e and f summarize the overall trade-off between predictive performance and computational cost. Although some baseline methods achieve comparable performance on individual metrics, DELTA provides the most balanced overall performance across accuracy, robustness, computational efficiency, and scalability. This advantage does not originate from a single model component. Rather, it arises from the synergy of three design elements: (i) LME-based quantification of material- and condition-dependent lifetime effects, (ii) data-efficient probabilistic learning through GMM, and (iii) entropy-based modeling of label availability. Together, these components make DELTA particularly suitable for battery manufacturing scenarios characterized by heterogeneous operating conditions and limited lifetime labels.

Unlike conventional semi-supervised learning methods that focus primarily on improving predictive accuracy, DELTA is specifically designed for manufacturing-oriented battery assessment. Therefore, its advantage lies not only in competitive classification performance, but also in its ability to operate efficiently under limited-label conditions while maintaining cross-material transferability and low deployment cost.

Robustness analysis

First, we evaluate the effect of the label availability within the DELTA framework (Fig. 5a). Ignoring the missing-label mechanism leads to systematically biased classification, making reliable deployment infeasible under realistic manufacturing conditions. The proposed method explicitly models the missing-label process under semi-supervised learning, enabling effective utilization of both labeled and unlabeled samples. As the missing-label ratio increases, the performance of all models gradually declines due to reduced supervision. Nevertheless, DELTA consistently achieves a higher F1 score than baseline methods and exhibits a slower degradation trend under high missingness, indicating the benefit of explicitly accounting for label availability. The corresponding accuracy results are presented in Fig. S9.

Second, we conduct an ablation study on the random-effects-derived feature HI_m obtained from the LME model

(Fig. 5b). Specifically, we compare the full model incorporating HI_m with a variant in which this feature is removed. The results show that performance gradually deteriorates as label availability decreases; the model including HI_m gradually exhibits a more pronounced advantage in both accuracy and F1 score, while also displaying a substantially narrower performance fluctuation range than the model without this feature. This behaviour suggests that early-cycle data contain non-negligible inter-cell variability and latent heterogeneity. By introducing the LME model, DELTA explicitly captures random deviations across samples, effectively quantifying material-induced uncertainty and thereby suppressing performance degradation while improving overall classification stability and accuracy. The corresponding accuracy comparison is provided in Fig. S10.

Third, we investigate the influence of the number of early cycles used for dynamic feature extraction (Fig. 5c). The results show that an accuracy of approximately 0.8 can be achieved using data from only five cycles. As the number of early cycles increases, both prediction accuracy and stability exhibit moderate fluctuations. When 20 cycles are used, DELTA achieves its highest accuracy (above 0.85). However, compared to using only five early cycles, the performance gain is marginal while requiring substantially longer testing time, indicating diminishing returns from incorporating additional early-cycle data. Notably, incorporating excessive early-cycle data (for example, more than 35 cycles) introduces additional noise that is not directly relevant to the target task and does not yield further performance gains, but instead slightly degrades performance. To balance testing cost and predictive accuracy, five early cycles are therefore selected for dynamic feature extraction. These results demonstrate that DELTA is relatively insensitive to the amount of early data and can maintain reliable performance even under highly data-scarce conditions. Besides, early-cycle signals already encode sufficient information about long-term degradation trajectories, enabling reliable prediction without extended testing.

Model deployment in manufacturing scenarios

In practical factory environments, large-scale historical data are continuously accumulated across production batches, material systems, and operating conditions. Motivated by this setting, we investigate two strategies for incorporating extended datasets, as illustrated in Fig. 5d: direct training and training without labels from the extended dataset. Both strategies aim to exploit heterogeneous data sources to improve model robustness and generalization.

Under both strategies, the extended dataset can directly utilize the material-related features generated by the LME model trained on the base dataset. As long as the operating conditions represented in the extended dataset fall within the range covered by the base dataset, these features can be reused without retraining the LME model, thereby maintaining computational efficiency.

Direct training refers to joint training on the merged base and extended datasets. A prescribed proportion of samples from the combined dataset is assigned lifetime labels, allowing labeled information to be drawn from both datasets during

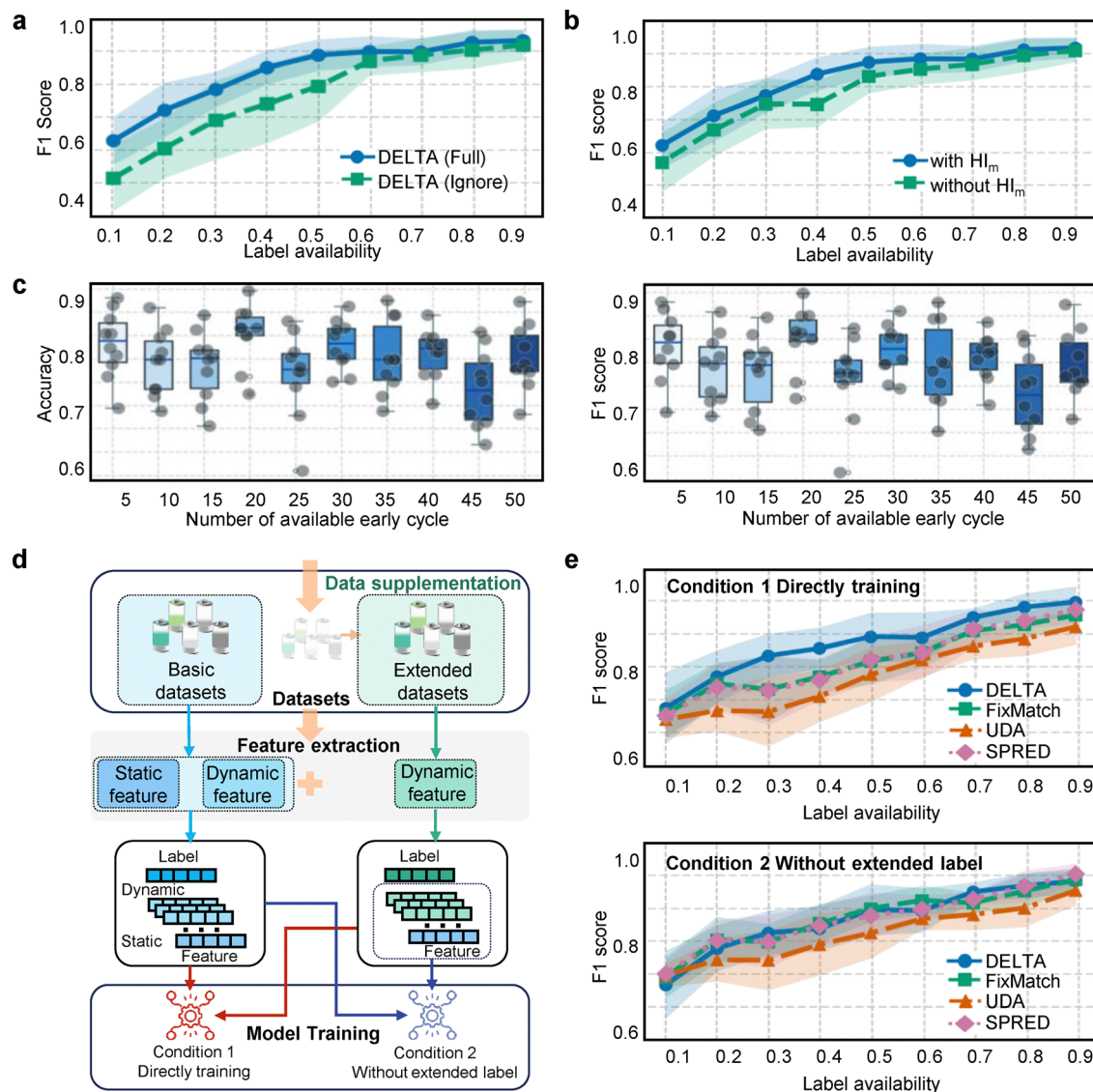


Fig. 5 Uncertainty analysis and model deployment. (a) Ablation study evaluating the impact of explicitly accounting for label availability under different label availability ratios. The results show that the full DELTA framework consistently outperforms the variant that ignores missing-label handling, particularly in low-label scenarios. (b) Ablation analysis of the LME-derived material feature HI_m . Performance comparisons between models with and without this feature demonstrate that incorporating random-effects modeling improves prediction stability and overall F1 performance across label availability settings. (c) Influence of the number of early cycles used for feature extraction on model performance. Boxplots show the distributions of accuracy and F1 score as the number of available early cycles increases, indicating that reliable predictions can be achieved using relatively few early-cycle measurements. (d) Illustration of two deployment conditions integrating extended datasets. The first condition directly trains the model using both dynamic and static features, while the second leverages extended datasets to extract additional feature representations without requiring additional labels. (e) Performance comparison of the two deployment strategies across different label availability ratios, showing that the proposed DELTA framework consistently achieves higher accuracy and F1 scores than baseline semi-supervised approaches.

model training. Model performance is evaluated using tenfold cross-validation. This setting represents a scenario in which at least a subset of cells in the extended dataset has completed long-term ageing tests and the corresponding lifetime labels have become available.

Training without extended-dataset labels, by contrast, is designed to emulate a more realistic manufacturing scenario. In this setting, all labels from the extended dataset are discarded, and the corresponding samples are incorporated into the training process solely as unlabeled data, while labeled

samples originate exclusively from the base dataset. The key difference between the two strategies, therefore, lies not only in the amount of training data, but also in the availability of lifetime labels within the extended dataset. In the direct-training setting, part of the extended dataset contributes supervised information, whereas in the unlabeled-fusion setting the extended dataset contributes only through its feature distribution.

The latter setting more closely reflects practical battery manufacturing conditions, where newly generated production data are continuously accumulated, but their corresponding

end-of-life labels are generally unavailable because long-term ageing tests have not yet been completed. In industrial practice, newly collected monitoring data can therefore be continuously appended to the existing database and incorporated into model training without waiting for costly and time-consuming lifetime testing. By leveraging such unlabeled data, DELTA can exploit their latent information content while substantially improving model robustness and generalization.

The performance of the two strategies is shown in Fig. 5e. Under the direct training condition, as the missing-label ratio increases, the accuracy and F1 score of all models decline due to reduced effective information. When the missing-label ratio is below 0.3, performance differences among models are relatively small, with F1 scores ranging from approximately 0.85 to 0.95. Across the entire missingness range, DELTA consistently outperforms all baseline methods (FixMatch, UDA and SPRED). In particular, under high missingness (0.6–0.9), DELTA exhibits the slowest performance degradation. For example, at a missing-label ratio of 0.7, DELTA maintains performance around 0.8, whereas other models drop to 0.75 or lower. Under the direct-mixing strategy, DELTA benefits from the strong distribution-modelling capability of the Gaussian mixture model and a robust semi-supervised learning mechanism, enabling more effective utilization of mixed data and superior generalization and stability under severe label scarcity. Without an extended label condition, it simulates scenarios in which newly acquired production data lack labels. Overall, performance without an extended label condition is slightly lower than under the direct training condition, particularly at high missing-label ratios, due to the loss of label information from the extended dataset. The overall degradation trend is similar to that of the direct training condition. Despite this reduction in available supervision, DELTA remains the best-performing method at 30% label availability. These results demonstrate that the GMM-based framework can efficiently extract useful information from unlabeled extended data, substantially improving generalization and mitigating label sparsity. Under the unlabeled fusion strategy, DELTA again demonstrates strong semi-supervised learning capability, effectively converting continuously accumulated unlabeled production data into gains in model generalization, highlighting the scalability of the proposed framework.

Under the end-of-line deployment scenario and the economic assumptions defined in SI4, we further evaluate the annualized economic cost of different screening strategies at a production scale of 1.3 TWh per year. The comparison considers three major components: labeling cost, re-grading cost, and potential risk-related losses caused by escaped defective cells. Under the adopted parameter settings, with $k = 5$ and a hide ratio of 0.7, the proposed DELTA framework achieves the lowest expected annual economic cost among all compared strategies. Specifically, the proposed method yields an average annual cost of approximately 2433.67M USD, substantially lower than the no-screening baseline, which reaches about 6500.16M USD per year, corresponding to an average annual saving of 4066.49M USD, or 62.57%. By contrast, the random sampling strategy (with $s = 0.2$) results in an annual cost of

approximately 6279.15M USD, providing only limited improvement over no screening. These results indicate that the proposed method can more effectively exploit limited labeled data together with unlabeled data to reduce the combined burden of missed-risk losses and unnecessary grading, thereby delivering markedly superior economic efficiency in practical battery manufacturing scenarios. Importantly, the framework requires no additional sensors and operates solely on existing formation-stage data, enabling seamless integration into current manufacturing pipelines without additional hardware overhead.

Discussion

Existing research on early battery quality assessment has largely been conducted using carefully curated datasets with complete lifetime labels.^{25–29} In practical manufacturing, obtaining such labels requires long-term ageing tests that are costly and incompatible with high-throughput production. Consequently, large volumes of early-cycle data generated during formation and grading remain unlabeled, creating a major bottleneck for scalable data-driven quality assessment.

The proposed DELTA framework addresses this challenge by enabling early battery quality classification using only the first few cycles of charging data together with material and operating-condition information. After offline training, DELTA can directly utilize early-cycle data generated during the formation stage to perform quality classification without requiring full ageing tests or extensive *post hoc* labeling. The success of DELTA stems from combining interpretable statistical modelling with semi-supervised learning: material-dependent variability is quantified through a linear mixed-effects model, while a Gaussian mixture model incorporating an entropy-based missing-label mechanism enables efficient exploitation of large volumes of unlabeled data commonly generated during battery manufacturing. This shifts battery quality control from offline, *post hoc* evaluation to early-stage, data-driven decision-making.

Through comprehensive evaluations across six publicly available battery datasets comprising 421 cells, 3 material systems, and multiple operating conditions, the proposed framework demonstrates strong performance under realistic label-scarce scenarios. DELTA achieves classification accuracy exceeding 83% when only about 30% of samples are labeled, while maintaining robust performance across batch-based and material-based classification criteria. Compared with several representative semi-supervised learning approaches, including FixMatch, UDA, SPRED, and self-training-based models, DELTA provides competitive predictive performance while significantly improving computational efficiency. In particular, the training time of DELTA is typically on the order of 10^{-1} seconds, which is several times faster than deep semi-supervised models, while inference requires only about 10^{-4} seconds per sample. Unlike deep learning approaches that typically require large labeled datasets and substantial computational resources, DELTA achieves competitive performance with significantly lower data and

computational requirements, making it particularly suitable for real-world deployment. The framework also demonstrates strong robustness to feature and label noise, as well as stable performance when integrating newly collected data *via* unlabeled data fusion. These results indicate that statistically structured semi-supervised models can achieve reliable early classification performance while maintaining the computational efficiency required for industrial deployment.

The novelty of DELTA does not lie in introducing a new battery lifetime classification task itself. Battery lifetime classification has been explored in previous studies, and predicting lifetime categories rather than exact EOL values is not, by itself, a new concept. Instead, the contribution of this work lies in addressing a practical manufacturing quality assessment problem under limited label availability. In realistic production environments, the key challenge is not the classification task itself, but how to perform reliable and transferable quality assessment using a small number of labeled cells together with a large amount of routinely generated early-cycle data.

Beyond predictive accuracy, the practical significance of DELTA lies in its ability to support manufacturing decision-making across multiple stages of the battery production workflow. During capacity grading, manufacturers traditionally rely on capacity, internal resistance, and self-discharge measurements to evaluate cell quality. However, these indicators primarily characterize the current state of a cell and provide limited insight into its future degradation trajectory. By incorporating the predicted lifetime category as an additional screening criterion, DELTA enables the identification of cells that exhibit similar initial performance yet are likely to experience substantially different long-term ageing behaviours. The value of such information extends beyond grading and is particularly relevant during pack assembly. Conventional cell matching strategies are largely based on voltage, capacity, and resistance measurements. Nevertheless, cells with comparable initial characteristics may still diverge significantly during long-term operation due to differences in degradation kinetics. Incorporating DELTA-derived lifetime categories into cell-matching strategies allows cells with more consistent ageing trajectories to be grouped together, thereby improving pack-level consistency and reducing the risk of premature performance degradation due to lifetime outliers.

Importantly, DELTA operates entirely on data routinely collected during formation and capacity-grading procedures and therefore requires neither additional sensing hardware nor supplementary testing protocols. As a result, the framework can be integrated into existing manufacturing execution systems with minimal deployment cost and disruption to current production workflows. It should also be emphasized that the abnormal cells identified in this study do not necessarily correspond to defective or unsafe products. Rather, abnormality is defined relative to the lifetime distribution of a given cell population. Cells assigned to the low-lifetime category are expected to degrade substantially faster than the population average, whereas cells belonging to extreme high-lifetime groups may exhibit ageing behaviours that differ markedly

from those of the majority population. Although such cells may satisfy conventional quality metrics at the time of manufacture, their deviation from the dominant ageing trend can introduce long-term inconsistency during pack operation. Early identification of these outliers therefore provides valuable information for cell sorting, pack assembly, and quality management in large-scale battery manufacturing.

We acknowledge that the present study does not include direct upstream manufacturing parameters. Therefore, the proposed framework does not aim to identify specific root causes in electrode fabrication or cell assembly. Instead, it provides an early quality assessment tool by linking end-of-line testing signatures to subsequent lifetime outcomes. The proposed entropy-based mechanism should not be interpreted as an exact representation of the real-world label-generation process. Rather, it provides a probabilistic approximation that allows the model to account for potential differences between labeled and unlabeled samples. This formulation is particularly relevant in battery manufacturing, where complete lifetime labels are available only for a subset of cells due to the cost and duration of long-term ageing tests. Although the entropy-based mechanism improves robustness under limited label availability, the true process governing label availability in industrial settings may depend on additional operational, economic, and safety-related factors that are not explicitly modeled in the present study.

Future research may also explore integrating richer sources of information to enhance early quality assessment. In addition to early-cycle voltage-capacity features, other non-invasive sensing signals, such as acoustic emissions,^{10–12} thermal imaging,^{13–15} ultrasonic grading,^{16–18} or impedance-related features,^{53,54} may provide complementary insights into internal battery states during manufacturing. Combining such multimodal signals with statistical learning frameworks could enable more comprehensive monitoring of cell quality and degradation behaviour. Furthermore, integrating physics-based battery models with data-driven learning approaches³² may help establish hybrid “physics–statistics” frameworks that provide both predictive accuracy and stronger physical interpretability for battery health assessment.

Methods

Data preparation

To enable early-stage EOL classification, a set of HIs is constructed from early-cycle capacity degradation and Q - V curve evolution. The features fall into three categories: normalized capacity features, Q - V curve-based features, and degradation trend features.

The normalized capacity is defined as

$$\tilde{Q}_k = Q_k / Q_{\text{rated}} \quad (1)$$

To capture early-cycle degradation behaviour, a differential Q - V function between selected cycles is defined as

$$\Delta Q(V) = Q_3(V) - Q_5(V) \quad (2)$$

Statistical descriptors derived from these signals, together with degradation trend features, are used as candidate HIs. All features are extracted exclusively from early-cycle data and selected based on their correlation with EOL cycle life. Eight informative features are retained (see SI5 for details).

Model architecture

The DELTA framework consists of two main components: a feature extraction module and a semi-supervised classification module. Within the feature extraction module, an LME model is employed to estimate the material-related random-effect feature HI_m , while a GMM is used to perform the semi-supervised classification task directly.

Linear mixed-effects model for material random effects. To account for material-dependent variability in battery ageing, an LME model is employed, in which temperature and charge rate are treated as fixed effects and material is modelled as a random effect. The model is formulated as

$$\text{Cycle}_{ij} = \beta_0 + \beta_T T_{ij} + \beta_C C_{ij} + \alpha_i + e_{ij} \quad (3)$$

where α_i represents the material-specific random effect. The estimated random effects are extracted as the feature HI_m and incorporated into the subsequent classification model (see SI6 for details).

Semi-supervised Gaussian mixture modelling. Battery health states are modelled using a finite Gaussian mixture model (GMM). Let $y_j \in \mathbb{R}^p$ denote the feature vector of the j -th cell, which is assumed to follow a mixture of g Gaussian components with parameters (μ_i, Σ_i, π_i) . Under the semi-supervised setting, class labels are available for a subset of samples, while the remaining class memberships are treated as latent variables and inferred *via* posterior probabilities (see SI7 for details).

Modelling the label availability. In practical battery testing, missing health-state labels are often correlated with the degree of battery ageing, making the missingness mechanism non-ignorable. To account for this limited label availability, a label indicator variable m_j is introduced, where $m_j = 1$ indicates that the label is unavailable and $m_j = 0$ indicates that the label is available. The probability of label missingness is modelled as

$$\Pr(m_j = 1 | y_j) = q(y_j; \theta, \xi) \quad (4)$$

where $q(\cdot)$ is a logistic function of the Shannon entropy computed from posterior class probabilities:

$$e_j = - \sum_{i=1}^g \tau_{ij} \log \tau_{ij} \quad (5)$$

Incorporating this mechanism, the complete partially classified log-likelihood is decomposed as

$$\log L^{\text{full}} = \log L^{\text{ig}} + \log L^{\text{miss}} \quad (6)$$

where the second term captures the contribution of the missing-label process. By explicitly modelling label missingness, the proposed framework exploits informative missingness to improve both classification accuracy and generalization.

Evaluation metrics

To comprehensively assess model performance in battery EOL cycle-life classification, multiple complementary evaluation metrics are adopted to examine overall accuracy, class discriminability, and the ability to identify minority-class samples from different perspectives. For a binary classification problem, the prediction outcomes are defined in terms of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) (see SI8 for details).

Author contributions

Chen Liang: writing – review & editing, writing – original draft, visualization, validation, methodology, formal analysis, conceptualization. Shengyu Tao: writing – review & editing, writing – original draft, visualization, validation, methodology, formal analysis, conceptualization, supervision, project administration, funding acquisition. Chunqiu Xia: writing – review & editing, visualization. Xinghao Huang: writing – review & editing, visualization. Hang Hu: writing – review & editing, visualization. Rui Wang: writing – review & editing, methodology, formal analysis. Daoyi Dong: resources, supervision, project administration, funding acquisition, conceptualization. Ziyang Lyu: project administration, writing – review & editing, supervision, conceptualization, funding acquisition. Guangmin Zhou: project administration, writing – review & editing, supervision, conceptualization. Huadong Mo: project administration, writing – review & editing, supervision, conceptualization, funding acquisition.

Conflicts of interest

There are no conflicts of interest to declare.

Data availability

Data from six publicly available battery datasets (CAS,⁴³ MICH,⁴⁴ RWTH,⁴⁵ Stanford,⁴⁶ TJU,⁴⁷ and XJTU⁴⁸) were used in this work. Data and code used in this work are available at <https://github.com/CheneyLc/DELTA>.

Supplementary information (SI), which contains additional methodological details, figures, tables, and supporting results, is available. See DOI: <https://doi.org/10.1039/d6ee02209j>.

Acknowledgements

This work was supported by the Marie Skłodowska-Curie Program HORIZON-MSCA-2025-PF, 101283078 BLESS [S. T.], Australian Economic Accelerator Ignite Grant ‘Health status estimation and resilient closed-loop supply chain for retired electric vehicle batteries’ (IG240100338) [H. M.] and the ACT Government under the Energy Innovation Fund ‘Online health monitoring and anomaly detection in Li-ion batteries *via* trustworthy AI’ [H. M.].

References

- 1 P. M. Attia, E. Moch and P. K. Herring, Challenges and opportunities for high-quality battery production at scale, *Nat. Commun.*, 2025, **16**, 611.
- 2 S. Tao, *et al.*, Non-destructive degradation pattern decoupling for early battery trajectory prediction via physics-informed learning, *Energy Environ. Sci.*, 2025, **18**, 1544–1559.
- 3 S. Tao, *et al.*, Generative learning assisted state-of-health estimation for sustainable battery recycling with random retirement conditions, *Nat. Commun.*, 2024, **15**, 10154.
- 4 S. Tao, X. Zhang and C. Zou, The role of machine-learning-enabled diagnostics in a circular battery economy, *Chem Circularity*, 2026, 100005, DOI: [10.1016/j.checir.2026.100005](https://doi.org/10.1016/j.checir.2026.100005).
- 5 J. Amici, *et al.*, A Roadmap for Transforming Research to Invent the Batteries of the Future Designed within the European Large Scale Research Initiative BATTERY 2030+, *Adv. Energy Mater.*, 2022, **12**, 2102785.
- 6 E. Ayerbe, M. Berecibar, S. Clark, A. A. Franco and J. Ruhland, Digitalization of Battery Manufacturing: Current Status, Challenges, and Opportunities, *Adv. Energy Mater.*, 2021, **12**, 2102696.
- 7 G. Qian, *et al.*, From in-situ experimentation to in-line metrology: Advanced imaging characterization for battery research and manufacturing, *Energy Storage Mater.*, 2024, **73**, 103819.
- 8 X. Huang, IC2ML: Unified battery state-of-health, degradation trajectory and remaining useful life prediction via intra-cycle and inter-cycle enhanced machine learning, *J. Power Sources*, 2026, **666**, 239148.
- 9 J. Wu, M. Fenech, R. F. Webster, R. D. Tilley and N. Sharma, Electron microscopy and its role in advanced lithium-ion battery research, *Sustainable Energy Fuels*, 2019, **3**, 1623–1646.
- 10 J. Scharf, *et al.*, Bridging nano- and microscale X-ray tomography for battery research by leveraging artificial intelligence, *Nat. Nanotechnol.*, 2022, **17**, 446–459.
- 11 T. M. M. Heenan, *et al.*, Theoretical transmissions for X-ray computed tomography studies of lithium-ion battery cathodes, *Mater. Des.*, 2020, **191**, 108585.
- 12 Y. Jiang, *et al.*, X-Ray Computed Tomography (CT) Technology for Detecting Battery Defects and Revealing Failure Mechanisms, *J. Electron. Mater.*, 2024, **53**, 5776–5787.
- 13 J. Yang, X. Lu, H. Du and J. Huang, A Miniaturized, Low-Cost Interrogator via Intrapulse Wavelength Demodulation for Operando Fiber-Optic Sensing in Batteries, *IEEE Trans. Instrum. Meas.*, 2025, **74**, 1–13.
- 14 T. Zheng, *et al.*, Operando monitoring of gassing dynamics in lithium-ion batteries with optical fiber photothermal spectroscopy, *Energy Environ. Sci.*, 2025, **18**, 8499–8514.
- 15 Z. Li, *et al.*, Data-driven assessment of lithium-ion battery degradation using thermal patterns from computer vision, *J. Energy Chem.*, 2025, **105**, 852–859.
- 16 Y. Shen, *et al.*, In situ detection of lithium-ion batteries by ultrasonic technologies, *Energy Storage Mater.*, 2023, **62**, 102915.
- 17 Y. Wang, *et al.*, Progress and challenges in ultrasonic technology for state estimation and defect detection of lithium-ion batteries, *Energy Storage Mater.*, 2024, **69**, 103430.
- 18 J. Zhang, *et al.*, Ultrasonic-assisted enhancement of lithium-oxygen battery, *Nano Energy*, 2022, **102**, 107655.
- 19 M. Fichtner, *et al.*, Rechargeable Batteries of the Future—The State of the Art from a BATTERY 2030+ Perspective, *Adv. Energy Mater.*, 2021, **12**, 2102904.
- 20 S. Tao, Rapid and sustainable battery health diagnosis for recycling pretreatment using fast pulse test and random forest machine learning, *J. Power Sources*, 2024, **597**, 234156.
- 21 S. Tao, Battery Cross-Operation-Condition Lifetime Prediction via Interpretable Feature Engineering Assisted Adaptive Machine Learning, *ACS Energy Lett.*, 2023, **8**, 3269–3279.
- 22 S. Tao, Immediate remaining capacity estimation of heterogeneous second-life lithium-ion batteries via deep generative transfer learning, *Energy Environ. Sci.*, 2025, **18**, 7413–7426.
- 23 D. L. Wood, J. Li and S. J. An, Formation Challenges of Lithium-Ion Battery Manufacturing, *Joule*, 2019, **3**, 2884–2888.
- 24 A. Weng, *et al.*, Predicting the impact of formation protocols on battery lifetime immediately after manufacturing, *Joule*, 2021, **5**, 2971–2992.
- 25 M. G. M. Abdolrasol, *et al.*, Advanced data-driven fault diagnosis in lithium-ion battery management systems for electric vehicles: Progress, challenges, and future perspectives, *eTransportation*, 2024, **22**, 100374.
- 26 Y. Zhang and M. Zhao, Cloud-based in-situ battery life prediction and classification using machine learning, *Energy Storage Mater.*, 2023, **57**, 346–359.
- 27 W. Guo, L. Yang, Z. Deng, B. Xiao and X. Bian, Early Diagnosis of Battery Faults Through an Unsupervised Health Scoring Method for Real-World Applications, *IEEE Trans. Transp. Electrification*, 2024, **10**, 2521–2532.
- 28 P. Wang, J. Chen, F. Lan, Y. Li and Y. Feng, Multiscale feature fusion approach to early fault diagnosis in EV power battery using operational data, *J. Energy Storage*, 2024, **98**, 112812.
- 29 V. Yamaçlı, State-of-health estimation and classification of series-connected batteries by using deep learning based hybrid decision approach, *Heliyon*, 2024, **10**, e39121.
- 30 S. Tao, Collaborative and privacy-preserving retired battery sorting for profitable direct recycling via federated machine learning, *Nat. Commun.*, 2023, **14**, 8032.
- 31 N. Guo, Semi-supervised learning for explainable few-shot battery lifetime prediction, *Joule*, 2024, **8**, 1820–1836.
- 32 X. Huang, *et al.*, iMOE: prediction of second-life battery degradation trajectory using interpretable mixture of experts, *Nat. Commun.*, 2026, **17**, 2549.
- 33 L. Su, *et al.*, Data sufficiency for transferable lithium-ion battery periodical SOH estimation under resource constraints, *Cell Rep. Phys. Sci.*, 2025, 102901, DOI: [10.1016/j.xcrp.2025.102901](https://doi.org/10.1016/j.xcrp.2025.102901).
- 34 X. Li, M. Lyv, X. Gao, K. Li and Y. Zhu, A co-estimation framework of state of health and remaining useful life for lithium-ion batteries using the semi-supervised learning algorithm, *Energy AI*, 2025, **19**, 100458.

- 35 M. Ye, *et al.*, Enhanced robust capacity estimation of lithium-ion batteries with unlabeled dataset and semi-supervised machine learning, *Exp. Syst. Appl.*, 2024, **238**, 121892.
- 36 J. Yao, Z. Chang, T. Han and J. Tian, Semi-supervised adversarial deep learning for capacity estimation of battery energy storage systems, *Energy*, 2024, **294**, 130882.
- 37 T. Liu, Y. Yang, G.-B. Huang, Y. K. Yeo and Z. Lin, Driver Distraction Detection Using Semi-Supervised Machine Learning, *IEEE Trans. Intell. Transport. Syst.*, 2016, **17**, 1108–1120.
- 38 H. Huo, *et al.*, Semi-supervised machine-learning classification of materials synthesis procedures, *npj Comput. Mater.*, 2019, **5**, 62.
- 39 Z. Lyu, D. Ahfock, R. Thompson and G. J. McLachlan, Semi-supervised Gaussian mixture modelling with a missing-data mechanism in R, *Aust. N. Z. J. Stat.*, 2024, **66**, 146–162.
- 40 Z. Lyu, Analysis of estimating the Bayes rule for Gaussian mixture models with a specified missing-data mechanism, *Comput. Stat.*, 2024, **39**, 3727–3751.
- 41 G. J. McLachlan, Estimating the Linear Discriminant Function from Initial Samples Containing a Small Number of Unclassified Observations, *J. Am. Stat. Assoc.*, 1977, **72**, 403–406.
- 42 N. V. Chawla and G. Karakoulas, Learning From Labeled And Unlabeled Data: An Empirical Study Across Techniques And Domains, *J. Artif. Intell. Res.*, 2005, **23**, 331–366.
- 43 C. Chen, *et al.*, The Operation Dependence of $C - N$ Fatigue for Lithium-Ion Batteries, *Adv. Energy Mater.*, 2023, **13**, 2300942.
- 44 A. Weng, *et al.*, Predicting the impact of formation protocols on battery lifetime immediately after manufacturing, *Joule*, 2021, **5**, 2971–2992.
- 45 W. Li, *et al.*, One-shot battery degradation trajectory prediction with deep learning, *J. Power Sources*, 2021, **506**, 230024.
- 46 X. Cui, *et al.*, Data-driven analysis of battery formation reveals the role of electrode utilization in extending cycle life, *Joule*, 2024, **8**, 3072–3087.
- 47 J. Zhu, *et al.*, Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation, *Nat. Commun.*, 2022, **13**, 2261.
- 48 F. Wang, Z. Zhai, Z. Zhao, Y. Di and X. Chen, Physics-informed neural network for lithium-ion battery degradation stable modeling and prognosis, *Nat. Commun.*, 2024, **15**, 4332.
- 49 K. Sohn, *et al.*, FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence, in *Advances in Neural Information Processing Systems*, ed. H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan and H. Lin, Curran Associates, Inc., 2020, vol. 33, pp. 596–608.
- 50 Q. Xie, Z. Dai, E. Hovy, T. Luong and Q. Le, Unsupervised Data Augmentation for Consistency Training, in *Advances in Neural Information Processing Systems*, ed. H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan and H. Lin, Curran Associates, Inc., 2020, vol. 33, pp. 6256–6268.
- 51 K. Xu, F. Zhuo, J. Li, X. Zou and J. Zhou, Self-reinforcing prototype evolution with dual-knowledge cooperation for semi-supervised lifelong person re-identification, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 3564–3574.
- 52 J. E. Van Engelen and H. H. Hoos, A survey on semi-supervised learning, *Mach. Learn.*, 2020, **109**, 373–440.
- 53 X. Du, *et al.*, Revolutionizing Battery Safety: Real-Time Insights with Dynamic Electrochemical Impedance Spectroscopy, *ACS Energy Lett.*, 2025, **10**, 2292–2304.
- 54 J. D. Huang and W. G. Zeier, Joint-Domain Impedance Spectroscopy for Solid-State Batteries: Enabling Accelerated Characterization and Data-Driven Insights, *ACS Energy Lett.*, 2026, 5c03055, DOI: [10.1021/acsenenergylett.5c03055](https://doi.org/10.1021/acsenenergylett.5c03055).