Journal of Energy Storage 153 (2026) 121006

Contents lists available at ScienceDirect

Journal of Energy Storage

journal homepage: www.elsevier.com/locate/est

Review article

Generative data augmentation for improving state estimation and prognostics in lithium-ion batteries: Advances, Challenges, and Future directions ☆

Md. Sulyman Islam Sifat [a], Ummy Sawda [b], Md Alamgir Kabir [a],*, Mobyen Uddin Ahmed [c], M.S. Hossain Lipu [d], M.M. Manjurul Islam [e]

[a] Department of Computer Science and Engineering, Daffodil International University, Daffodil Smart City, Dhaka 1216, Bangladesh
[b] Department of Software Engineering, Daffodil International University, Daffodil Smart City, Dhaka 1216, Bangladesh
[c] School of Innovation, Design and Engineering, Mälardalen University, Västerås, 722 20, Sweden
[d] School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, NSW 2522, Australia
[e] Intelligent Systems Research Centre, Ulster University, Londonderry, BT48 7JL, United Kingdom

ARTICLE INFO

Keywords:
Generative Adversarial Networks
Lithium-ion batteries
Battery state estimation
Prognostics
Data augmentation
Systematic literature review
Deep learning
Battery health management

ABSTRACT

Lithium-ion battery state estimation and prognostics are crucial for the safe and reliable operation of electric vehicles (EVs), renewable energy systems, and portable devices. However, nonlinear battery behavior complicates State of Charge (SOC) estimation, limited and imbalanced data hinder State of Health (SOH), and scarce degradation trajectories with domain shift challenge Remaining Useful Life (RUL) prediction. Generative Adversarial Networks (GANs) offer an effective approach by generating realistic synthetic data that mitigates scarcity, improves diversity, and enhances model robustness. This study aims to systematically review and consolidate the current literature on GAN-based data augmentation for battery state estimation and prognostics. Specifically, it examines the effectiveness of different GAN architectures and techniques in improving SOC, SOH, and RUL estimation, assesses synthetic data quality and reliability, identifies technical challenges and limitations, and outlines evidence-based guidelines and future research directions. A systematic literature review was conducted following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines, incorporating relevant studies obtained from four prominent digital libraries. Analysis of 31 primary studies reveals that widely used public datasets (National Aeronautics and Space Administration (NASA), Center for Advanced Life Cycle Engineering (CALCE) and Oxford) dominate the field, enabling reproducibility and benchmarking. GAN-based augmentation achieves 17% to 90% error reductions across root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and mean squared error (MSE) metrics. Time-series GANs and Wasserstein GANs (WGANs) emerge as most effective, with Adam optimizer learning rate (LR) = 0.001 and gradient penalty (λ_{GP}=10) providing improved stability. Major challenges include data scarcity, synthetic data quality concerns, GAN instability, and domain shift. Six priority areas are identified for advancing the field: physics-informed constraints, domain adaptation, uncertainty quantification, real-time deployment, multimodal learning, and data efficiency. This review establishes GAN-based augmentation as significant for battery state estimation and prognostics. It provides evidence-based insights and a roadmap towards developing reliable, interpretable, and deployable battery management systems.

Contents

☆ Abbreviations are listed in Appendix.
* Corresponding author.
E-mail addresses: sifat15-4004@diu.edu.bd (M.S.I. Sifat), sawda241-35-520@diu.edu.bd (U. Sawda), kabir.cse@diu.edu.bd (M.A. Kabir), mobyen.uddin.ahmed@mdu.se (M.U. Ahmed), molla_lipu@uow.edu.au (M.S.H. Lipu), m.islam@ulster.ac.uk (M.M.M. Islam).

https://doi.org/10.1016/j.est.2026.121006
Received 23 September 2025; Received in revised form 14 January 2026; Accepted 6 February 2026
Available online 12 February 2026

1. Introduction

Lithium-ion batteries are a fundamental component of modern energy storage and are used in electric vehicles, renewable-energy systems, spacecraft, and portable electronic devices because of their high energy density, rechargeability, and long cycle life [1,2]. Ensuring their reliability and safety requires accurate estimation of critical health indicators, namely State of Charge (SOC), State of Health (SOH), and Remaining Useful Life (RUL), which are vital for stable operation across EVs, renewable grids, and portable applications, particularly given the complexity of battery aging and environmental influences [3–5]. SOC represents the available energy relative to the maximum capacity, SOH reflects the level of degradation compared to the initial condition, and RUL forecasts the remaining operational cycles until the end-of-life threshold is reached. Precise prediction of these indicators is essential not only for enhancing operational efficiency but also for preventing costly downtime and catastrophic failures, as demonstrated in recent works employing deep learning and hybrid approaches [6–11]. Comprehensive reviews provide further insights into SOC, SOH, and RUL estimation, summarizing algorithmic advances, challenges in data modeling, and the integration of prognostics and health management (PHM) frameworks [12–14]. Recent surveys and benchmark studies consistently highlight data scarcity, limited cross-domain generalization, and deployment constraints as key barriers to reliable SOC, SOH, and RUL estimation and to their integration within PHM frameworks [15–18].

The electrochemical processes governing lithium-ion battery behavior are inherently nonlinear and time-varying, influenced by thermal effects, degradation phenomena, and dynamic loading conditions, making accurate modeling particularly challenging [19–21]. Traditional approaches, such as equivalent circuit models (ECM) and electrochemical models [22], often fail to capture complex dynamics and degradation mechanisms under real-world operating conditions. Although

ECMs are computationally efficient, they neglect critical internal processes such as lithium plating, solid–electrolyte interphase growth, and temperature-induced degradation, particularly at low SOC or high current rates [23,24]. Electrochemical models, such as the Doyle–Fuller–Newman (DFN) framework and its derivatives, provide deeper physical insights but are computationally too intensive for real-time applications [25,26]. Moreover, ECMs typically assume linear superposition and constant parameters, thereby ignoring history-dependent effects such as hysteresis and path-dependent degradation [27]. Recent comparisons further reveal that distributed or physics-informed ECM variants outperform traditional ECMs (e.g., Thevenin, Randles, and resistor–capacitor (RC) ladder models), especially under real-world cycling and high C-rate [21,26]. Recent hybrid approaches, including physicochemical ECMs and differentiable DFN formulations, can achieve DFN-level fidelity at substantially reduced computational cost, supporting near-real-time estimation [26,28].

Temporal Fusion Transformer (TFT) is an attention-based sequence architecture that is increasingly used as a modern Transformer baseline for SOC/SOH/RUL estimation under limited or imbalanced data.

In recent years, deep learning methods have demonstrated substantial potential in battery health prediction, owing to their ability to capture complex temporal dependencies and degradation patterns directly from raw measurements. Hybrid architectures, such as convolutional neural network (CNN)–bidirectional long short-term memory (BiLSTM) [29], leverage convolutional layers for feature extraction and bidirectional LSTMs for sequence learning, thus improving the accuracy of RUL prediction. Attention-based CNNs with positional encoding further enhance temporal modeling, offering faster and more accurate estimation compared to conventional recurrent neural network (RNN)-based approaches [30,31]. More advanced combinations, including CNN – gated recurrent unit (GRU) – Transformer frameworks [32], have shown strong capability in SOC estimation, while Transformer-based models themselves exhibit promising results for RUL prediction, particularly on limited datasets [33]. Despite these advancements, deep learning models are heavily data-dependent. They require large and balanced datasets to avoid overfitting and often fail to generalize when operating conditions differ from training distributions [9]. Transformer-style models (e.g., TFT variants and multimodal encoders) deliver sizable gains under small or imbalanced data, indicating improved temporal fusion and generalization for SOC/SOH/RUL [8,34,35]. These findings support updating deep-learning baselines beyond recurrent architectures in data-constrained settings. The lack of diverse and publicly available datasets further restricts model scalability, forcing reliance on data augmentation strategies [36]. To address these challenges, Generative Adversarial Networks (GANs) have been shown to be effective for synthetic data generation in battery applications. Time-series WGANs [37] and deep convolutional GANs [38] create realistic SOC and RUL trajectories, preserve temporal correlations, and significantly enhance predictive performance [36,39]. Recent reviews further identify GANs as promising enablers for augmenting rare failure modes and boosting the robustness of deep learning-based prognostics and health management frameworks [9].

1.1. Problem and motivation

Despite the demonstrated potential of GAN-based approaches for lithium-ion battery state estimation (SOC and SOH) and prognostics (RUL) [36,40–42], several unresolved challenges remain critical bottlenecks (i.e., addressed in research questions (RQ1–RQ5)). The foremost limitation arises from the inherently nonlinear and time-varying electrochemical dynamics of lithium-ion cells, which make precise modeling extremely difficult (address in RQ3 and RQ4). Previous studies [43,44] show that the nonlinearity and non-Gaussian behavior of batteries under dynamic loads complicate SOC and SOH estimation, and advanced filters such as Gauss–Hermite particle filters outperform standard methods in such cases [45]. Furthermore, temperature fluctuations and operational variability significantly affect real-time SOC estimation, as demonstrated in electric vehicle applications [3].

The second major challenge lies in the scarcity and imbalance of available datasets [46,47] (address in RQ1 and RQ4). Capacity estimation models often degrade in accuracy when trained on limited or unbalanced data, with experiments on NASA battery datasets confirming poor generalization for underrepresented failure modes [48]. Similarly, the absence of degraded-class data in early cycles results in unreliable SOH prediction, highlighting the need for improved sampling and augmentation strategies [49].

The third pressing limitation is the lack of comprehensive degradation trajectories and the prevalence of domain shift across operating conditions [50] (i.e., emphasize in RQ4). Shifts in load profiles, ambient temperature, and usage patterns reduce the robustness of RUL models, with hybrid fusion frameworks combining Kalman filters and nonlinear autoregressive models proposed as partial remedies [50]. Long-term degradation patterns also remain underrepresented, as models trained on short-term data fail to generalize, and even probabilistic approaches leveraging accelerated degradation tests cannot compensate for the lack of real-world long-term datasets [51].

Together, these challenges severely constrain the reliability and robustness of current battery health management frameworks [52]. While extensive experimental datasets could theoretically alleviate these issues, their collection is prohibitively expensive, time-consuming, and incapable of covering the full diversity of chemistries, operating modes, and environmental factors [36]. Traditional data augmentation techniques, such as interpolation (mixup) [53], SMOTE [54], or Gaussian noise injection [55], are similarly inadequate. For example, Zhang et al. [56] demonstrated that Gaussian data augmentation (GDA) improves performance only under specific signal-to-noise conditions, and inappropriate GDA can actually degrade model accuracy—highlighting the fragility of such methods that fail to capture temporal dependencies and electrochemical characteristics. Against this backdrop, researchers increasingly recognize GANs as an effective approach to data augmentation for battery state estimation and prognostics [57]. By synthesizing high-fidelity, statistically consistent, and temporally coherent data [58], GANs can mitigate scarcity, enrich diversity, and enhance the generalizability of deep learning models—ultimately reducing dependence on costly experimental campaigns while enabling more reliable SOC, SOH, and RUL prediction [36,52]. Stabilized and conditional GAN variants (e.g., WGAN-GP and time-series/conditional models) can further improve robustness and cross-domain utility through gradient-penalty regularization, conditional controls, and few-shot or domain-adaptive synthesis [59–61]. These advances support the rationale for GAN-based augmentation under small, imbalanced, or shifted battery datasets.

1.2. Research gaps

Our analysis reveals seven distinct critical research gaps in the current research on GAN-based battery prognostics. For each gap, we provide a concise operational definition and evidence anchor (see Fig. 12, Tables 9–11, and Fig. 13). For complete per-study codings, counts, and definitions of research gap categories and mitigations, see the OSF data extraction table. These research gaps are presented in Table 1.

These seven gaps are derived from the 31-study corpus analyzed in this systematic literature review (SLR) and directly motivate our research questions (Table 3) and contributions. In particular, our benchmarking and guidance address Gaps 1–4 (reporting, validation, baselines, and integration), while our roadmap targets Gaps 5–7 (transferability, physics-informed constraints, and uncertainty).

For data availability, all detailed evidence presented in this gaps table, including per-study gap analysis, comprehensive analysis tables, dataset inventories, GAN architecture specifications, performance metrics, challenge categorizations, and future research direction mappings

Table 1
Research gaps identified from the 31 primary studies on Generative Adversarial Network (GAN)-based data augmentation for lithium-ion battery state estimation and prognostics. Each row defines a gap category, summarizes its meaning, and provides evidence anchors by citing representative primary studies (PS).

Gap#	Research Gaps	Description	Evidence
Gap 1	Systematic evaluation and reporting	Lack of standardized experimental protocols and inconsistent reporting of dataset splits, pre-processing steps, real-to-synthetic data ratios, and model hyperparameters prevents fair comparison and reproducibility across studies	Incomplete reporting of splits, preprocessing, and hyperparameters (PS6, PS7, PS10, PS11, PS25); heavy reliance on the NASA, CALCE, and Oxford datasets (PS1, PS3, PS4, PS5, PS8, PS9, PS11, PS13, PS22, PS26, PS29, PS30, PS31); small cell counts (2–30 cells); missing baselines in PS1, PS5, PS11, PS24, PS29.
Gap 2	Quality validation protocols	Absence of systematic validation frameworks for synthetic data quality, with inconsistent use of statistical metrics (MMD, DTW, PCA, and t-SNE) and limited downstream performance validation to ensure synthetic data reliability	Inconsistent quality metrics: KL divergence (PS1, PS11), MMD filtering (PS10), DTW similarity (PS12), and PCA and t-SNE (PS2, PS25); quality challenges in PS2, PS6, PS7, PS12, PS22, PS31.
Gap 3	Architecture and hyperparameter baselines	Proliferation of diverse GAN architectures without systematic comparison or ablation studies, making it difficult to determine optimal configurations for specific battery state estimation tasks	12 GAN families: TimeGAN (PS1, PS9, PS11, PS25, PS30), WGAN variants (PS2, PS7, PS8, PS10, PS17, PS18, PS20, PS22), CGAN (PS6, PS12, PS15, PS28, PS29), InfoGAN, VAE-GAN, CR-GAN, CTSGAN, CLSTM-CGAN, TS-DCGAN, GRU-LSGAN, GAN-CLS, RGAN; limited cross-architecture comparison (PS6: CGAN vs. DCGAN vs. WGAN; PS10: WTimeGAN vs. TimeGAN; PS29: RNN-CGAN vs. LSTM-CGAN vs. CNN-CGAN).
Gap 4	Integration strategies (Real–Synthetic fusion)	Insufficient investigation of optimal strategies for combining real and synthetic data, including mixing ratios, curriculum learning approaches, and replacement versus augmentation strategies	Limited ratio sensitivity across studies; PS8 uses discriminator-weighted fusion; PS20 evaluates replacement vs. augmentation; PS25 shows preprocessing impact; missing baselines in PS1, PS5, PS11, PS24, PS29.
Gap 5	Transferability and domain shift	Limited evaluation of GAN model generalization across different battery chemistries, operating temperatures, drive cycles, and datasets, hindering deployment in diverse real-world conditions	Domain shift in PS20, PS21, PS27, PS31; transfer learning: PS20 (W-DC-GAN-GP-TL), PS21 (GAN-LSTM-TL), PS27 (DeepCoral + CTSGAN), PS31 (feature mapping).
Gap 6	Physics-informed constraints	GAN generators lack integration of electrochemical knowledge and physical constraints, failing to ensure generated data adheres to battery physics principles (e.g., SOC-OCV relationships, capacity fade monotonicity, voltage–current feasibility)	Physics constraints proposed in PS1, PS3–PS6, PS9, PS12–PS13, PS18, PS20, PS23; current focus on data-driven objectives.
Gap 7	Uncertainty quantification and reliability	Absence of uncertainty quantification and reliability assessment in GAN-generated synthetic data, critical for safety-sensitive battery management applications where prediction confidence is essential	Uncertainty quantification (UQ) recommended in PS3, PS8, PS12, PS13, PS18, PS21, PS27; current focus on point metrics without calibration.

for all 31 primary studies, are available in the Open Science Framework (OSF) data extraction repository https://osf.io/rmj8b/. The evidence statements above are derived from a systematic analysis of the complete OSF data extraction tables.

These observations are consistent with broader related reviews [5, 62–66], which also note inconsistent reporting and limited cross-domain validation. Our work adds GAN-specific evidence and quantified prevalence tailored to augmentation for battery prognostics.

1.3. Contributions

We summarize our key contributions as follows:

1. We consolidate GAN based augmentation for SOC, SOH, and RUL and provide quantitative synthesis, standardized benchmarks, validation of synthetic data, and protocols to integrate real and synthetic data.
2. Following PRISMA 2020 guidelines, provides a systematic analysis of 31 primary studies on GAN-based data augmentation for lithium-ion battery state estimation and prognostics, covering SOC, SOH, and RUL with comprehensive task coverage analysis.
3. Releases a complete OSF reproducibility package containing the protocol, PRISMA 2020 checklist, screening ledgers with exclusion reasons, deduplication and search logs/strategies, and data-extraction and CASP QA tables to enable end-to-end replication and reuse.
4. Establishes quantitative performance benchmarks showing 17% –90% error reductions across RMSE, MAE, MAPE, and MSE metrics, with empirical evidence from 25 of 31 studies demonstrating consistent performance improvements.
5. Identifies effective GAN architecture families and their prevalence across the corpus, including time-series GAN variants, WGAN and WGAN-GP variants, and conditional GANs, along with optimal hyperparameter configurations for stable training.
6. Provides evidence-based implementation guidelines for practitioners, including dataset preprocessing pipelines, validation protocols, and practical guidance on real–synthetic data integration strategies (see Fig. 5; Tables 7, 12, 13; and OSF artifacts at https://osf.io/rmj8b/).
7. Establishes six priority research directions for advancing the field, including physics-informed constraints, domain adaptation, uncertainty quantification, real-time deployment, multimodal learning, and data efficiency improvements.

Collectively, these contributions address gaps in reporting, validation, baselines, integration, transferability, physics informed constraints, and uncertainty. Evidence for each contribution appears in Figs. 11 and 12 and Tables 7, 9, and 11.

1.4. Organization

The remainder of this paper is organized as follows: Section 2 provides background on lithium-ion battery fundamentals, the evolution of data-driven battery state estimation, and theoretical foundations of GANs and their variants for time-series data. Section 3 presents the systematic review methodology, including research questions, search strategy, inclusion/exclusion criteria, and quality assessment framework. Section 4 provides the comprehensive analysis of the primary studies, examining GAN architectures, performance metrics, and implementation details, along with the per-RQ findings. Section 4 discusses key findings regarding the effectiveness of GAN-based data augmentation, including performance improvements, validation strategies, and identified challenges. Finally, Section 5 concludes the review with a synthesis of key findings and recommendations for practitioners and researchers.

2. Background and related work

2.1. Background

2.1.1. Generative adversarial networks theories

GANs, introduced by Goodfellow et al. [67], represent a significant advance in generative modeling. GAN model is based on game theory, specifically the minimax game between two competing neural networks: the generator G and the discriminator D. The generator aims to create synthetic data that is indistinguishable from real data, while the discriminator learns to distinguish between real and generated samples. Subsequent formulations generalized the divergence being minimized and analyzed convergence and stability properties [68,69].

As shown in Fig. 1, the mathematical formulation of the GAN objective function is (see [67]):

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \tag{1}$$

where $p_{data}(x)$ represents the real data distribution, $p_z(z)$ is the prior noise distribution, and z is the random noise vector input to the generator. The training process involves alternating optimization of the generator and discriminator networks until they reach a Nash equilibrium.

The theoretical analysis of GAN convergence reveals several important properties. When the discriminator is optimal for a fixed generator, the objective function becomes:

$$V(G,D^*) = \mathbb{E}_{x\sim p_{data}(x)}[\log D^*(x)] + \mathbb{E}_{x\sim p_g(x)}[\log(1 - D^*(x))] \tag{2}$$

where $D^*(x) = \frac{p_{data}(x)}{p_{data}(x)+p_g(x)}$ is the optimal discriminator [67].

Substituting this optimal discriminator into the objective function yields [67]:

$$V(G,D^*) = 2\cdot \mathrm{JSD}(p_{data} \parallel p_g) - \log(4) \tag{3}$$

where JSD represents the Jensen–Shannon divergence between the real and generated data distributions.

The training dynamics of GANs can be analyzed using the gradient flow equations:

$$\frac{\partial \theta_D}{\partial t} = -\nabla_{\theta_D} V(D,G) \tag{4}$$

$$\frac{\partial \theta_G}{\partial t} = \nabla_{\theta_G} V(D,G) \tag{5}$$

where θ_D and θ_G are the parameters of the discriminator and generator, respectively.

The convergence of GAN training is guaranteed under certain conditions, including the assumption that both networks have sufficient capacity and that the discriminator is trained to optimality at each step. However, in practice, these conditions are rarely met, leading to training instability and mode collapse.

2.1.2. Variants of GAN for time-series data

Standard GANs face challenges when applied to time-series data due to their inability to capture temporal dependencies and sequential patterns. Several GAN variants have been developed specifically for time-series generation, each addressing different aspects of the temporal modeling challenge; alternative objectives such as Least-Squares GAN further improve gradient signals and stability [70].

Conditional GANs (CGANs) extend the basic GAN framework by conditioning both the generator and discriminator on additional information, such as class labels or control variables. Fig. 2 depicts the conditioning of both G and D on y. Auxiliary-classifier designs and spectral normalization further stabilize training and improve conditional

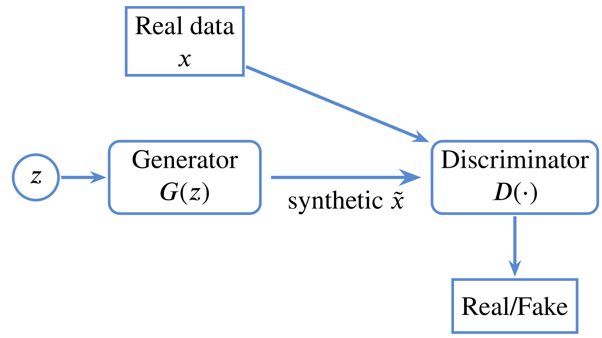

Fig. 1. Basic Generative Adversarial Network (GAN) architecture: generator G maps noise z to synthetic samples $\tilde{x}$, and discriminator D distinguishes real x from generated $\tilde{x}$.

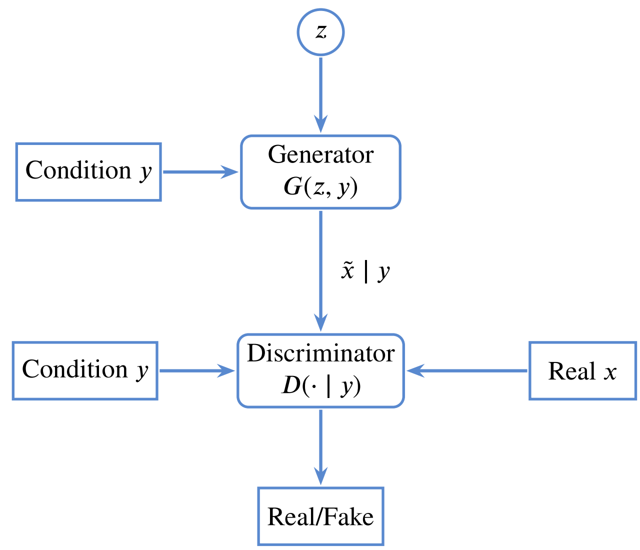

Fig. 2. Conditional Generative Adversarial Network (CGAN): both generator and discriminator are conditioned on y (e.g., temperature, drive cycle, and chemistry) for targeted time-series synthesis.

synthesis [71,72]. The objective function becomes [73]:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x,y\sim p_{data}(x,y)}[\log D(x|y)] + \mathbb{E}_{z\sim p_z(z), y\sim p_y(y)}[\log(1 - D(G(z|y)|y))] \tag{6}$$

The conditional generation process can be formulated as:

$$G : \mathcal{Z} \times \mathcal{Y} \to \mathcal{X} \tag{7}$$

$$D : \mathcal{X} \times \mathcal{Y} \to [0,1] \tag{8}$$

where $\mathcal{Z}$ is the noise space, $\mathcal{Y}$ is the condition space, and $\mathcal{X}$ is the data space.

In Eq. (6), conditioning on y injects operating context into both G and D, which naturally supports context-aware synthesis for battery trajectories. Conditional and time-series GANs enable context-conditioned sequence synthesis (e.g., conditioning on temperature, drive cycle, and chemistry) for battery monitoring, improving SOC, SOH, and RUL estimation by supporting temporally consistent generation in low-data settings [74,75].

Wasserstein GANs (WGANs) address training instability issues by using the Wasserstein distance instead of the Jensen–Shannon divergence. Fig. 3 illustrates the WGAN-GP critic enforcing the 1-Lipschitz

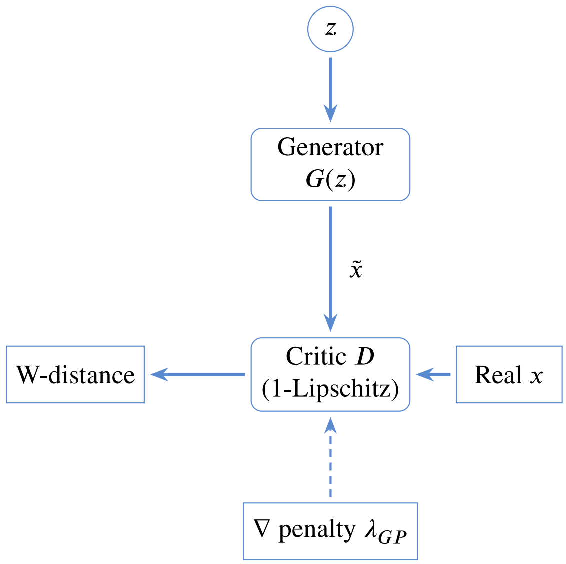

Fig. 3. Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP): the critic estimates the Earth Mover's distance; a gradient penalty with coefficient λ_{GP} enforces the 1-Lipschitz constraint.

constraint via a gradient penalty λ_{GP}. The objective function is [76]:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[D(x)] - \mathbb{E}_{\tilde{x}\sim p_g(\tilde{x})}[D(\tilde{x})] \tag{9}$$

The Wasserstein distance between two probability distributions P and Q is defined as [77,78]:

$$W(P,Q) = \inf_{\gamma\in\Pi(P,Q)} \mathbb{E}_{(x,y)\sim\gamma}[\|x-y\|] \tag{10}$$

where $\Pi(P,Q)$ is the set of all joint distributions with marginals P and Q.

The Kantorovich-Rubinstein duality provides an alternative formulation [77,78]:

$$W(P,Q) = \sup_{\|f\|_L\le 1} \mathbb{E}_{x\sim P}[f(x)] - \mathbb{E}_{x\sim Q}[f(x)] \tag{11}$$

where the supremum is over all 1-Lipschitz functions f.

Building on the Wasserstein objective in Eq. (9), practical time-series training typically relies on Lipschitz-enforcing stabilizers (e.g., gradient penalty) to keep the critic well-behaved. For time-dependent signals, gradient-penalty regularization and related stabilizers (e.g., synchronized noised gradient penalty and 1D WGAN-GP) can improve training stability and waveform fidelity relative to vanilla GANs, which is directly relevant for electrochemical time-series synthesis [79,80].

Time-series Generative Adversarial Network (TimeGAN), specifically designed for time-series data, incorporates temporal consistency through a supervised loss that encourages the generated sequences to follow realistic temporal patterns (see [81]). Alternatives include recurrent and convolutional recurrent GANs for sequential synthesis [82,83]. Fig. 4 summarizes the architecture: an embedding network E maps sequences to latent h, a recovery network R reconstructs $\hat{x}$, a supervisor S aligns temporal dynamics, and the generator G and discriminator D play the adversarial game.

The TimeGAN objective function combines adversarial, supervised, and reconstruction losses [81]:

$$\mathcal{L} = \mathcal{L}_R + \alpha\mathcal{L}_S + \beta\mathcal{L}_A \tag{12}$$

where $\mathcal{L}_R$ is the reconstruction loss, $\mathcal{L}_S$ is the supervised loss, and $\mathcal{L}_A$ is the adversarial loss.

The reconstruction loss ensures that the embedding and recovery networks can reconstruct the original time series [81]:

$$\mathcal{L}_R = \mathbb{E}_{x\sim p_{data}(x)}[\|x-\hat{x}\|_2^2] \tag{13}$$

where $\hat{x}$ is the reconstructed time series.

The supervised loss encourages the generator to produce sequences that follow realistic temporal dynamics [81]:

$$\mathcal{L}_S = \mathbb{E}_{x\sim p_{data}(x)}[\|h_{t+1}-\hat{h}_{t+1}\|_2^2] \tag{14}$$

where h_t and $\hat{h}_t$ are the real and generated hidden states, respectively.

Modern battery prognostics systems employ sophisticated deep learning architectures to process multi-modal sensor data and extract meaningful features for health prediction. These architectures must handle the unique challenges of battery data, including temporal dependencies, multiple sensor modalities, and varying time scales [84,85].

Long Short-Term Memory (LSTM) networks are particularly effective for battery prognostics due to their ability to capture long-term dependencies in sequential data. The LSTM cell structure includes input, forget, and output gates that control information flow [86,87]:

$$f_t = \sigma(W_f\cdot[h_{t-1},x_t]+b_f) \tag{15}$$

$$i_t = \sigma(W_i\cdot[h_{t-1},x_t]+b_i) \tag{16}$$

$$\tilde{C}_t = \tanh(W_C\cdot[h_{t-1},x_t]+b_C) \tag{17}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{18}$$

$$o_t = \sigma(W_o\cdot[h_{t-1},x_t]+b_o) \tag{19}$$

$$h_t = o_t * \tanh(C_t) \tag{20}$$

The LSTM architecture addresses the vanishing gradient problem through the cell state C_t, which acts as a highway for information flow. The forget gate f_t controls how much information from the previous cell state is retained, while the input gate i_t controls how much new information is stored in the cell state.

The gradient flow through the LSTM cell can be analyzed using the chain rule [86,88]:

$$\frac{\partial h_t}{\partial h_{t-1}} = \frac{\partial h_t}{\partial C_t}\frac{\partial C_t}{\partial C_{t-1}}\frac{\partial C_{t-1}}{\partial h_{t-1}} \tag{21}$$

The partial derivative of the cell state with respect to the previous cell state is [87]:

$$\frac{\partial C_t}{\partial C_{t-1}} = f_t + \text{additional terms} \tag{22}$$

where the additional terms involve gradients of the gate functions with respect to the previous hidden state.

Bidirectional LSTM (BiLSTM) processes sequences in forward and backward directions and captures dependencies from past and future time steps simultaneously. This architecture is particularly valuable for battery prognostics where both historical and future context are important [89,90].

The BiLSTM output at time t is computed as [91]:

$$h_t = [\vec{h}_t; \overleftarrow{h}_t] \tag{23}$$

where $\vec{h}_t$ and $\overleftarrow{h}_t$ are the forward and backward hidden states, respectively.

The forward hidden state is computed as:

$$\vec{h}_t = \text{LSTM}(x_t, \vec{h}_{t-1}) \tag{24}$$

and the backward hidden state is computed as:

$$\overleftarrow{h}_t = \text{LSTM}(x_t, \overleftarrow{h}_{t+1}) \tag{25}$$

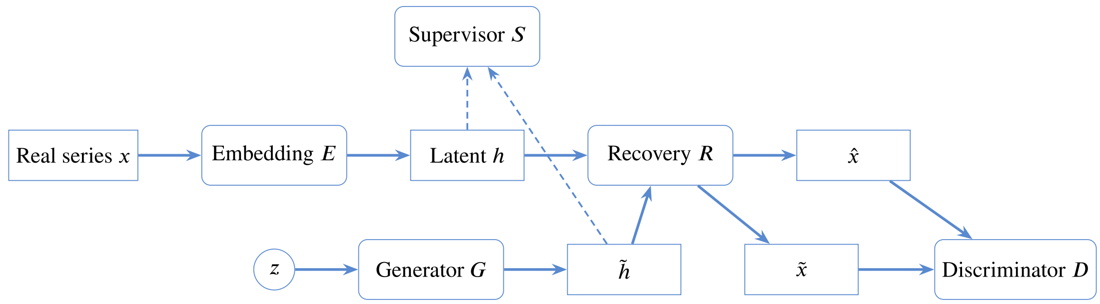

Fig. 4. Time-series Generative Adversarial Network (TimeGAN): embedding E and recovery R map between sequence space and latent space; generator G produces latent trajectories $\tilde{h}$ guided by supervisor S for temporal consistency; discriminator D distinguishes reconstructed real $\hat{x}$ from generated $\tilde{x}$.

Transformer Networks have significantly changed sequence modeling through their attention mechanism, which allows the model to focus on relevant parts of the input sequence. We adopt the standard encoder-style notation shown in Fig. 5 for downstream estimators. The multi-head self-attention mechanism is defined as [92]:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{26}$$

where Q, K, and V represent query, key, and value matrices, and d_k is the dimension of the key vectors.

The multi-head attention mechanism combines multiple attention heads [92]:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \tag{27}$$

where each head is computed as:

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{28}$$

The transformer architecture includes positional encoding to maintain sequence order information [92]:

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{model}}) \tag{29}$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{model}}) \tag{30}$$

where pos is the position in the sequence and i is the dimension.

With attention formalized in Eq. (26), Transformer-family models can fuse long-range dependencies across multivariate battery signals. Attention-based sequence models (e.g., TFT variants and seasonal/growth-aware Transformers) have been reported to improve SOC, SOH, and RUL estimation under limited or imbalanced data, complementing recurrent baselines [8,35].

2.1.3. Data augmentation and synthetic data generation

Data augmentation techniques aim to increase the diversity and quantity of training data without collecting additional real-world measurements. Traditional augmentation methods, such as noise injection, time shifting, and scaling, are limited in their ability to preserve the complex statistical properties of battery data [93,94].

The theory of data augmentation follows from data manifold learning. The real data distribution $p_{data}(x)$ lies on a low-dimensional manifold embedded in the high-dimensional data space. Traditional augmentation methods often fail to preserve the manifold structure, leading to unrealistic synthetic samples.

GAN-based data augmentation offers several advantages over traditional methods. First, GANs can learn the underlying data distribution and generate samples that preserve the statistical characteristics of real data. The generator learns a mapping from the noise space to the data manifold:

$$G : \mathcal{Z} \to \mathcal{M}_{data} \tag{31}$$

where $\mathcal{M}_{data}$ is the data manifold.

Second, GANs can generate data for scenarios that are difficult or expensive to obtain experimentally, such as extreme operating conditions or rare failure modes. This capability is particularly valuable for battery prognostics, where certain failure modes may be rare but critical for safety.

Third, GANs can help address class imbalance issues by generating synthetic samples for underrepresented conditions. An inverse-frequency weighted loss can be formulated as, following established strategies for learning from imbalanced data [95,96]:

$$\mathcal{L}_{CB} = \sum_{c=1}^{C} \frac{1}{n_c} \sum_{i=1}^{n_c} \mathcal{L}(y_i, \hat{y}_i) \tag{32}$$

where C is the number of classes, n_c is the number of samples in class c, and $\mathcal{L}$ is the base loss function.

The quality of synthetic data can be evaluated using various statistical metrics. The Fréchet Inception Distance (FID) measures the distance between real and synthetic data distributions [97]; complementary precision–recall diagnostics provide sample-quality vs. coverage trade-offs [98,99]:

$$\text{FID} = \|\mu_r - \mu_g\|_2^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2}) \tag{33}$$

where μ_r and μ_g are the mean feature vectors of real and generated data, respectively, and Σ_r and Σ_g are the corresponding covariance matrices.

The Inception Score (IS) measures both the quality and diversity of generated samples [100]:

$$\text{IS} = \exp(\mathbb{E}_{x\sim p_g(x)}[\text{KL}(p(y|x) \parallel p(y))]) \tag{34}$$

where $p(y|x)$ is the conditional class distribution and $p(y)$ is the marginal class distribution.

This motivates learned generative augmentation that models the data distribution more directly, rather than relying on heuristic perturbations. Compared with noise injection or SMOTE, generative augmentation using GANs and diffusion models has been reported to improve SOH and RUL accuracy while better preserving distributional fidelity (e.g., via FID or MMD), supporting the pipeline's utility under sparse or shifted data [52,101].

2.1.4. Evaluation metrics and validation strategies

The quality of GAN-generated synthetic data must be rigorously assessed to ensure its reliability for battery state estimation and prognostics. Common statistical metrics include:

Kullback–Leibler (KL) Divergence measures the difference between the probability distributions of real and synthetic data:

$$D_{KL}(P \parallel Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \tag{35}$$

The KL divergence is asymmetric and can be interpreted as the information loss when approximating distribution P with distribution Q. For continuous distributions, the KL divergence is defined as [102]:

$$D_{KL}(P \parallel Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx \tag{36}$$

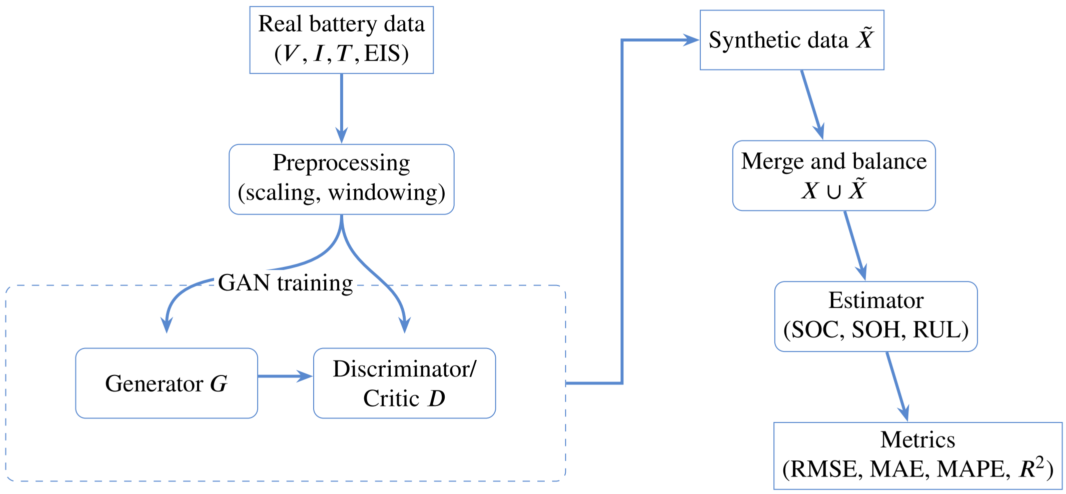

Fig. 5. Battery state estimation and prognostics pipeline for State of Charge (SOC), State of Health (SOH), and Remaining Useful Life (RUL) using Generative Adversarial Network (GAN)-based augmentation: real data are preprocessed and used to train G and D; generated $\tilde{X}$ are merged with real X to train downstream estimators, evaluated via standard metrics.

The Jensen–Shannon divergence provides a symmetric alternative [103]:

$$D_{JS}(P \parallel Q) = \frac{1}{2} D_{KL}(P \parallel M) + \frac{1}{2} D_{KL}(Q \parallel M) \tag{37}$$

where $M = \frac{1}{2}(P + Q)$ is the average distribution.

Principal Component Analysis (PCA) is used to compare the feature distributions and variance explained by principal components in real versus synthetic data. The PCA transformation is defined as [104]:

$$X_{PCA} = XW \tag{38}$$

where X is the data matrix and W is the matrix of eigenvectors of the covariance matrix $C = \frac{1}{n-1} X^T X$.

The explained variance ratio for each principal component is:

$$\text{Explained Variance Ratio}_i = \frac{\lambda_i}{\sum_{j=1}^{d} \lambda_j} \tag{39}$$

where λ_i is the eigenvalue corresponding to the ith principal component.

t-Distributed Stochastic Neighbor Embedding (t-SNE) provides visual comparison of data distributions in low-dimensional space, helping identify clustering patterns and potential mode collapse. The t-SNE algorithm minimizes the KL divergence between high-dimensional and low-dimensional neighbor probability distributions [105].

The conditional neighbor probability in high-dimensional space is [105]:

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)} \tag{40}$$

which is symmetrized as $p_{ij} = (p_{j|i} + p_{i|j})/(2n)$, and the joint probability in low-dimensional space is [105]:

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|y_k - y_l\|^2)^{-1}} \tag{41}$$

We evaluate performance for battery state estimation and prognostics using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{42}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{43}$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{44}$$

The RMSE penalizes large errors more heavily than small errors, making it sensitive to outliers. The MAE provides a more robust measure that is less sensitive to extreme values. The MAPE normalizes the error relative to the true value, making it useful for comparing performance across different scales.

Additional metrics specific to battery prognostics include the coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{45}$$

where $\bar{y}$ is the mean of the true values.

However, low prediction error alone does not guarantee that generated samples are realistic or cover the target operating distribution. In practice, distributional-fidelity checks (e.g., FID, MMD, and soft dynamic time warping (soft-DTW)) are paired with task-level error metrics for SOC, SOH, and RUL to verify that improvements in synthetic data quality translate into downstream prognostic accuracy [101,106].

2.1.5. Challenges in GAN-based battery state estimation and prognostics

Despite their potential, GAN-based approaches face several theoretical and practical challenges that must be addressed for successful deployment in battery state estimation and prognostics systems.

Training instability, characterized by mode collapse and vanishing gradients, remains a significant issue. Mode collapse occurs when the generator produces limited varieties of samples, failing to capture the full diversity of the real data distribution. This phenomenon can be analyzed through the lens of game theory, where the generator and discriminator reach a suboptimal Nash equilibrium.

The mode collapse problem can be quantified using the effective sample size (ESS) [107]:

$$\text{ESS} = \frac{(\sum_{i=1}^{n} w_i)^2}{\sum_{i=1}^{n} w_i^2} \tag{46}$$

where w_i are the importance weights of the generated samples.

The vanishing gradient problem arises when the discriminator becomes too confident, providing insufficient gradient information for the generator. This occurs when the discriminator achieves near-perfect classification, making the gradient of the generator's loss function approach zero; practical stabilizers include gradient-penalty and local-noise regularizers that smooth the discriminator landscape [108,109]:

$$\nabla_{\theta_G} \mathcal{L}_G = \mathbb{E}_{z \sim p_z(z)} [\nabla_{\theta_G} \log(1 - D(G(z)))] \tag{47}$$

When $D(G(z)) \approx 0$ for all generated samples, the gradient becomes very small, slowing down generator training.

Domain shift and generalization issues are particularly relevant for battery state estimation and prognostics, where operating conditions and battery chemistries vary significantly. The domain adaptation (DA) problem can be formalized as [110]:

$$\mathcal{L}_{DA} = \mathcal{L}_S + \lambda \mathcal{L}_{MMD} \tag{48}$$

where $\mathcal{L}_S$ is the supervised loss on the source domain, $\mathcal{L}_{MMD}$ is the maximum mean discrepancy between source and target domains, and λ is a trade-off parameter.

The maximum mean discrepancy (MMD) measures the distance between two probability distributions:

$$\begin{aligned}\mathrm{MMD}^2(P,Q) = \mathbb{E}_{x,x'\sim P}[k(x,x')] + \mathbb{E}_{y,y'\sim Q}[k(y,y')] \\ -2\mathbb{E}_{x\sim P,y\sim Q}[k(x,y)]\end{aligned} \tag{49}$$

where $k(\cdot,\cdot)$ is a kernel function.

The computational complexity of GAN training poses challenges for real-time applications, where computational resources and response times are constrained. The computational complexity of the forward pass through a GAN is:

$$O(L_G N_G^2 + L_D N_D^2) \tag{50}$$

where L_G and L_D are the number of layers in the generator and discriminator, respectively, and N_G and N_D are the average number of neurons per layer.

The lack of interpretability in GAN-generated data presents another challenge. Unlike physics-based models, GANs operate as black boxes, making it difficult to understand the underlying mechanisms and validate the physical plausibility of generated samples. This limitation is particularly concerning in safety-critical applications such as battery management systems (BMS).

The interpretability challenge can be partially addressed through attention mechanisms and saliency maps [111]:

$$S(x) = \frac{\partial f(x)}{\partial x} \tag{51}$$

where $f(x)$ is the model output and $S(x)$ is the saliency map indicating the importance of each input feature.

Additionally, the evaluation of GAN performance in battery management systems requires careful consideration of the specific requirements of the application domain. Unlike image generation, where visual quality is the primary concern, battery data generation must preserve physical constraints and temporal consistency.

2.2. Related work

Recent reviews address lithium-ion battery state estimation, health management, degradation analysis, and prognostics. Collectively, they survey on SOC and SOH estimation methods across traditional model-based and modern data-driven techniques [5,62–64]. Several works critically examine the evolution of battery management systems and highlight how advanced machine learning and deep learning algorithms improve accuracy and reliability [65,112,113]. Others focus on prognostics and health management (PHM), and discuss challenges and opportunities in leveraging large-scale battery datasets and artificial intelligence for predictive maintenance [66,114]. Notably, multiple articles explore specialized neural networks and the role of data augmentation in enhancing battery health prognostics, while also surveying RUL and knee point estimation methodologies [115,116]. Table 2 summarizes representative reviews. Together, these studies establish a solid foundation for current practice and future directions in data-driven, AI-enabled battery research.

Despite these significant advances, several critical gaps remain unaddressed in the existing literature. Most importantly, there is a lack of systematic analysis and benchmarking of generative adversarial network (GAN)–based data augmentation techniques specifically tailored for lithium-ion battery state estimation and prognostics. While prior reviews acknowledge the potential of synthetic data and generative models, they often treat these topics superficially or as part of broader discussions on machine learning. Few, if any, provide a focused, comparative evaluation of GAN-driven approaches, their effectiveness in mitigating data scarcity, or their impact on downstream state estimation and health prediction tasks. Additionally, recent trends in the development and application of advanced generative models—such as conditional GANs, Wasserstein GANs, and time-series GANs—are underrepresented in the review landscape. By systematically addressing these gaps, the present SLR delivers a focused, quantitative assessment of GAN-based augmentation for SOC, SOH, and RUL, including validation protocols and real–synthetic integration strategies, and provides benchmarking results and practical recommendations for future research and application.

3. Methodology

3.1. Review protocol

The study was conducted in accordance with the PRISMA 2020 reporting guidance, and the protocol was registered on the Open Science Framework (OSF; [117], project: https://osf.io/yshnq/ , protocol: https://osf.io/d3hpb/) after the screening stage. Registration occurred post hoc; however, no deviations were made to the eligibility criteria, outcomes, or synthesis plan. The scope was defined a priori using Population, Intervention, Comparator, Outcomes, Study design (PICOS): lithium-ion battery studies as the population; GAN-based data augmentation as the intervention; non-augmented or alternative augmentation as comparators; outcomes including SOC, SOH, and RUL estimation errors; and empirical, peer-reviewed studies published in English between 2015 and 2025. Information sources comprised ScienceDirect, IEEE Xplore, Scopus, and Wiley Online Library, which were searched on 14 February 2025 using database-adapted strategies (archived on OSF). Bibliographic exports were consolidated and deduplicated in Mendeley Desktop, with logs and settings archived; counts reconcile with the PRISMA flow diagram (Fig. 6). Study selection used dual independent title/abstract and full-text screening after a brief calibration; disagreements were resolved by discussion. Data were extracted with a piloted template, and methodological quality was appraised using a CASP-based rubric by two reviewers with consensus. The synthesis was pre-specified as narrative with quantitative before/after comparisons where matched baselines were available; heterogeneity precluded meta-analysis. All materials (search strings, exports, deduplication logs, screening ledger, extraction sheets, and analysis scripts) are publicly available on OSF. Detailed operational procedures are provided in the following subsections.

3.2. Research questions

This systematic literature review identifies and addresses key research questions, listed in Table 3, to guide the analysis of GAN-based data augmentation for lithium-ion battery state estimation and prognostics. These questions focus on datasets, methods, effectiveness, challenges, and future directions. For operationalization: RQ1 extracts dataset provenance, chemistries, variables, and usage frequency; RQ2 captures GAN family, loss and stabilizers, training regimen, and downstream estimators; RQ3 compares base versus augmented metrics (prioritizing RMSE and MAE) and computes relative deltas; RQ4 codes challenge domains and mitigation strategies; RQ5 aggregates forward-looking themes and counts. The full extraction schema is available on OSF.

Table 2
Related review papers and their limitations with respect to Generative Adversarial Network (GAN)-based data augmentation for lithium-ion battery SOC, SOH, and RUL tasks. The table lists each review title and publication year, and summarizes the specific limitation motivating the present systematic review (e.g., lack of GAN-specific benchmarking, synthetic-data validation protocols, or real–synthetic integration guidance).

SL NO.	Title	Year	Limitation
1	A comprehensive review of state-of-charge and state-of-health estimation for lithium-ion battery energy storage systems [62]	2024	Limited treatment of GAN-based augmentation; no standardized augmentation benchmarks; minimal real–synthetic integration guidance.
2	A review of lithium-ion battery state of health and remaining useful life estimation methods based on bibliometric analysis [63]	2024	Bibliometric focus; lacks technical evaluation of augmentation or dataset fusion; no pipeline-level recommendations.
3	Review on Lithium-ion Battery PHM from the Perspective of Key PHM Steps [114]	2024	Process-focused perspective; augmentation methods not analyzed; no synthetic-data quality criteria.
4	Review of State Estimation and Remaining Useful Life Prediction Methods for Lithium–Ion Batteries [5]	2023	Broad state-estimation survey; does not benchmark GAN augmentation or synthetic-data validation.
5	Specialized deep neural networks for battery health prognostics: Opportunities and challenges [115]	2023	Model-centric; lacks augmentation experiments and uncertainty-aware synthetic-data validation.
6	A review of lithium-ion battery state of charge estimation based on deep learning: Directions for improvement and future trends [64]	2022	SOC-centric; augmentation discussed superficially; no GAN-specific comparisons or quality protocols.
7	Towards unified machine learning characterization of lithium-ion battery degradation across multiple levels: A critical review [65]	2022	General ML perspective; no targeted analysis of GAN augmentation; limited cross-dataset transfer evaluation.
8	Prognostics and health management of Lithium-ion battery using deep learning methods: A review [66]	2022	Deep-learning oriented; does not evaluate GAN augmentation; limited domain-shift/transfer discussion.
9	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 [116]	2022	EV-focused; augmentation methods not covered; no GAN or real–synthetic integration analysis.
10	State estimation for advanced battery management: Key challenges and future trends [113]	2019	High-level challenges; no augmentation benchmarking; lacks real–synthetic fusion strategies.
11	Towards a smarter battery management system: A critical review on battery state of health monitoring methods [112]	2018	Pre-GAN focus; minimal deep sequence coverage; no data-augmentation discussion.

Table 3
Research questions (RQ1–RQ5) guiding the systematic literature review. Each row states a research question and defines the scope of evidence extracted from the 31 primary studies (datasets, Generative Adversarial Network (GAN) architectures and augmentation techniques, task-level effectiveness for State of Charge (SOC), State of Health (SOH), and Remaining Useful Life (RUL), technical challenges, and future directions).

#	Research question
RQ1	Which datasets have been used for training GAN-based models in lithium-ion battery state estimation and prognostics, and why?
RQ2	What GAN architectures and data-augmentation techniques have been applied, and what motivated their use?
RQ3	How effective is GAN-based data augmentation in improving state estimation and prognostics tasks (SOC, SOH, and RUL)?
RQ4	What technical challenges and limitations exist for GAN-based methods in this field?
RQ5	What future research directions are proposed for GAN-based data augmentation in battery state estimation and prognostics?

3.3. Search strategy

Relevant contributions on GAN-based data augmentation for lithium-ion battery state estimation were identified by searching ScienceDirect, IEEE Xplore, Scopus, and Wiley Online Library on February 14, 2025, with results limited to publications from 2015 to 2025 and to peer-reviewed journals and conference venues. Detailed record counts and the PRISMA flow are reported in the study selection process subsection.

A comprehensive Boolean search string was constructed using relevant keywords and synonyms for "Generative Adversarial Network", "data augmentation", and "lithium-ion battery state estimation" (e.g., SOC, SOH, RUL). The search string was adapted as needed for each database to maximize the retrieval of relevant literature.

Searches were restricted to English-language records at the database level for practicality, as the field is predominantly published in English. Gray literature (preprints, theses, technical reports) was not included; peer-reviewed venues were prioritized to ensure methodological rigor. Backward/ forward citation chasing was not performed systematically; key reference lists were scanned opportunistically. Search reporting follows PRISMA extension for Searching (PRISMA-S); detailed, database-specific strings and export logs are available on OSF.

3.4. Inclusion/exclusion criteria

To ensure rigor and relevance, inclusion and exclusion criteria were established a priori. Table 4 summarizes the criteria; when ambiguous cases arose (e.g., preprints, theses, simulation-only papers, non-Li-ion chemistries), the following rules were applied: preprints and theses were excluded; simulation-only studies without empirical validation were excluded; non-Li-ion chemistries were excluded unless explicitly used for cross-domain transfer experiments. Reasons for exclusion were recorded to align with PRISMA categories.

3.5. Study selection process

Study selection followed a structured, two-stage process conducted by two independent reviewers using the pre-specified eligibility criteria. First, titles and abstracts were screened to remove clearly ineligible records; second, full texts were assessed for compliance with all inclusion and exclusion parameters. Before screening commenced, a small-sample calibration was conducted to standardize interpretation and resolve ambiguities. Disagreements were resolved by discussion. The PRISMA flow diagram (Fig. 6) summarizes identification, screening, eligibility, and inclusion counts.

In total, 319 records were identified across the databases. After deduplication, 241 unique records remained for title/abstract screening, of which 196 were excluded. Forty-five articles proceeded to full-text review; 14 were excluded (typical reasons: out of scope, insufficient methodological detail, non-empirical). Thirty-one studies met all criteria and were included in the review.

This protocolized workflow—with independent dual review, piloted criteria, and documented reasons for exclusion—minimizes selection bias and enhances transparency and reproducibility.

Fig. 6 depicts the PRISMA 2020 flow: 319 records identified; 78 duplicates removed; 241 records screened (titles/abstracts) with 196 excluded; 45 full-text articles assessed with 14 excluded (typical reasons: out of scope, insufficient methodological detail, non-empirical); 31 studies included.

Table 4
Inclusion and exclusion criteria used to screen records for this systematic literature review. The table defines the eligibility rules applied during title/abstract and full-text screening for studies on lithium-ion batteries that use Generative Adversarial Network (GAN)-based (or closely related generative) augmentation and address SOC, SOH, or RUL outcomes.

Criteria	Description
Inclusion	Published 2015–2025, English; peer-reviewed journal articles or conference papers; focus on lithium-ion batteries; use of GANs or generative models for data augmentation; address battery prognostics (SOC, SOH, or RUL).
Exclusion	Not related to lithium-ion batteries; no use of GANs or generative data augmentation; not addressing SOC, SOH, or RUL in the context of battery prognostics; insufficient methodological detail or lack of experimental validation; non-English or non-peer-reviewed sources; duplicates.

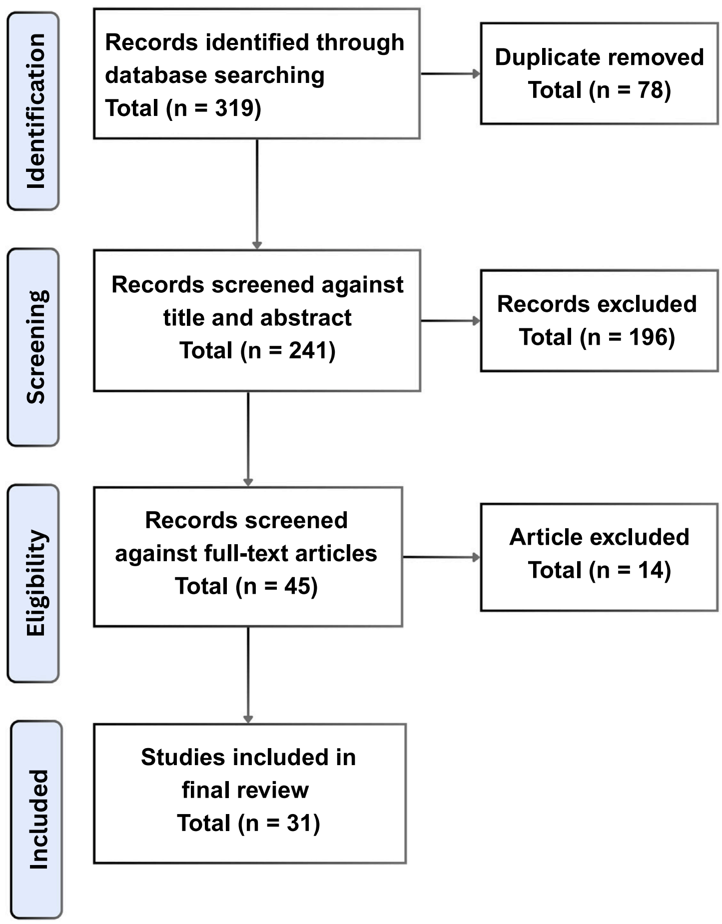

Fig. 6. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart of the study selection process. The diagram reports record counts (n) at each screening stage: records identified through database searching; duplicates removed; records screened by title and abstract; records excluded; full-text articles assessed for eligibility; full-text articles excluded; and studies included in the final review ($n = 31$).

3.6. Data extraction & quality assurance

Systematic data extraction was conducted using a structured approach. A comprehensive extraction form and supporting files were hosted on OSF (project hub: https://osf.io/yshnq/) to capture key methodological and bibliographic information from each included study. The extraction protocol encompassed study characteristics (author, publication year, publication venue), technical specifications (battery chemistry, estimation target), methodological details (GAN architecture, downstream learning model, dataset provenance and availability), performance metrics (evaluation criteria, baseline and augmented performance measures), and qualitative insights (reported challenges, limitations, and future research directions). A brief calibration round on a sample of studies was completed before full extraction to ensure consistency. This systematic data collection framework ensures comprehensive coverage of the evidence required to address the five research questions. The full data-extraction table for all 31 studies is hosted on OSF at https://osf.io/rmj8b/ (comma-separated values (CSV) and related artifacts, with version history).

Methodological quality was assessed in parallel. Each paper was scored against the ten-item Critical Appraisal Skills Programme (CASP) checklist (Table 5) using the 1/0.5/0 rubric in Table 6. Two reviewers worked independently; disagreements were resolved by discussion. Inter-rater reliability was not calculated; agreement was monitored, and differences were resolved by consensus. The resulting scores are visualized in Fig. 7.

It summarizes the total QA score (maximum = 10) obtained by each primary study. A complete per-criterion matrix for the 31 studies is available in OSF: https://osf.io/rmj8b/.

Furthermore, analysis of quality assessment reveals that three studies achieved perfect scores (10/10), demonstrating high methodological rigor across all criteria. These high-quality studies include PS2, PS20, and PS26. Three additional studies achieved near-perfect scores of 9.5/10: PS19, PS21, and PS22. Eleven studies scored 9/10, indicating very high quality in most criteria. These top-performing studies consistently demonstrated comprehensive GAN model descriptions, transparent implementation details, robust synthetic data validation techniques,

Table 5
Quality assessment (QA) criteria applied to every primary study (PS1–PS31). Each criterion specifies a reporting or methodological element used to judge study rigor and reproducibility (e.g., clarity of aims, dataset documentation, synthetic-data validation, and baseline comparisons).

#	Criterion
1	Aim clarity: Are the aims of the study clearly stated and aligned with battery state estimation and prognostics using GANs?
2	Scope definition: Is the scope of the study clearly defined, including the specific estimation task (e.g., SOC, SOH, and RUL)?
3	Research focus: Is the research question centered on GAN-based augmentation for lithium-ion batteries?
4	GAN Model description: Is the GAN architecture (e.g., DCGAN, CGAN, and CycleGAN) described in sufficient technical detail?
5	Implementation transparency: Are training details (hyper-parameters, loss functions, hardware) reported for reproducibility?
6	Dataset documentation: Is the battery dataset well documented, publicly available, and appropriate for the task?
7	Synthetic-data validation: Are the generated datasets validated using robust statistical or visual techniques (e.g., PCA, t-SNE, KL divergence, and SSIM)?
8	Performance metrics: Are suitable metrics (e.g., RMSE, MAE, and R^2) reported to quantify the impact of augmentation?
9	Baseline comparison: Does the study compare the GAN-based approach with simpler or non-augmented baselines?
10	Conclusion alignment: Are the conclusions consistent with the stated aims and evidence presented?

Table 6
Interpretation rubric for quality assessment (QA) scoring. Scores of 1/0.5/0 indicate sufficient/partial/insufficient evidence that a primary study satisfies a given QA criterion (Table 5).

Score	Meaning	Definition
1	Sufficient	Criterion is fully satisfied: information is clear, detailed, and directly relevant.
0.5	Neutral	Criterion is partially satisfied: information is vague, incomplete, or only indirectly relevant.
0	Insufficient	Criterion is not satisfied: information is missing, unclear, or irrelevant.

and thorough baseline comparisons. Their high-quality scores provide strong evidence for the reliability of findings related to the effectiveness of GAN-based data augmentation in estimating the state of lithium-ion batteries.

Moreover, the quality assessment process ensured methodological rigor and transparency in our systematic review, allowing us to evaluate the reliability and validity of the included studies. This comprehensive evaluation framework provides a solid foundation for synthesizing evidence and drawing meaningful conclusions about the effectiveness of GAN-based data enhancement approaches in the prognostics of lithium-ion batteries.

4. Results and discussion

4.1. Bibliometric analysis

This section analyzes the bibliometric characteristics of the 31 primary studies through three key visualizations: temporal distribution, publication venues, and citation patterns. Citation counts were retrieved from Scopus on 14 February 2025 and venue metadata was taken from publisher records. The comma-separated values (CSV) files and plotting scripts are available on OSF (project hub: https://osf.io/yshnq/).

Fig. 8 shows the distribution of the publications from 2020 to 2025. No eligible studies were identified in 2015–2019 (0 papers). Output rose from 2 papers in 2020 to a peak of 13 in 2023 (41.9% of all studies), with 4 in 2021 and 2 in 2022, and then stabilized at 5 in both 2024 and 2025. The citation curve shows that papers from 2021–2022 have accumulated the most citations, with the highest citation counts occurring for papers published in these years rather than in 2023, despite 2023 having the highest publication volume. To mitigate recency effects, the citations per year are additionally calculated and reported in the OSF CSVs; the prevalence of zero citations in recent years is also evident in Fig. 10. The compound annual growth rate (CAGR) of publications over 2020–2025 is $\left(\frac{5}{2}\right)^{1/5} - 1 \approx 20.1\%$ per year (CAGR $= (V_{\text{end}}/V_{\text{start}})^{1/\text{years}} - 1$).

Fig. 9 displays the distribution across different publication venues. The analysis of the 31 studies shows that 25 publications (80.6%) appeared in journals, while 6 publications (19.4%) were presented at conferences. The most frequent publication venues include *IEEE Transactions on Transportation Electrification* (3 papers), *Journal of Energy Storage* (3 papers), and *Applied Energy* (2 papers).

Fig. 10 presents the citation distribution across the 31 studies. The data shows a wide range of citation counts, from 0 to 91 citations. The most highly cited papers are PS26 (91 citations), PS14 (46 citations), PS11 (36 citations), and PS15 (33 citations). The citation analysis reveals that 9 out of 31 papers (29.0%) have zero citations, all of which are recent publications from 2023–2025. Papers published in 2021–2022 show the highest citation counts, indicating their important role in establishing the research direction for subsequent studies. Additionally, the median citations-per-year is 1.0 (interquartile range (IQR) 0.0–5.5), computed from Scopus counts normalized by years since publication (snapshot: 14 Feb 2025); the CSV is available on OSF. Citations-per-year were computed as $\text{cpy} = \frac{\text{citations}}{2025-\text{publication\_year}+1}$.

4.2. RQ1: Which datasets have been used for training GAN-based models in lithium-ion battery state estimation and prognostics, and why?

A detailed analysis of dataset usage in GAN-based lithium-ion battery state estimation and prognostics research demonstrates a clear dominance of a few public datasets (Fig. 11), with the NASA battery datasets (run-to-failure benchmark) being the most prevalent [118]. Specifically, 13 primary studies (PS1, PS5, PS8, PS9, PS11, PS13, PS21, PS22, PS25, PS26, PS29, PS30, and PS31) use the NASA datasets, making them the de facto standard for benchmarking and model development in this field. Their widespread adoption is driven by their comprehensive run-to-failure cycling data, which includes voltage, current, temperature, and capacity measurements across multiple cells and test conditions. This richness enables robust training, validation, and benchmarking of GAN models, and supports a variety of tasks such as SOH and SOC estimation, RUL prediction, and synthetic data generation. The repeated use of these datasets is further justified by their public accessibility, which ensures reproducibility and facilitates direct comparison across studies. Nevertheless, NASA subsets are typically lab-conditioned with limited temperatures and standardized constant-current constant-voltage (CCCV) / constant-current (CC) protocols and chemistries, which introduces domain shift relative to real EV profiles and other chemistries (e.g., PS21, PS24, PS31).

We next discuss CALCE and Oxford usage patterns and their implications for generalization and transfer.

The University of Maryland CALCE datasets [119] (diverse chemistries, transfer) are the second most frequently used, appearing in 6 studies (PS4, PS7, PS12, PS21, PS29, and PS31). These datasets are highly valued for their diversity in battery chemistries (including $LiCoO_2$, nickel manganese cobalt (NMC), and lithium iron phosphate (LFP)) and cycling protocols, which allow researchers to rigorously test the generalization and transfer learning capabilities of GAN models. For example, studies such as PS7 and PS12 combine

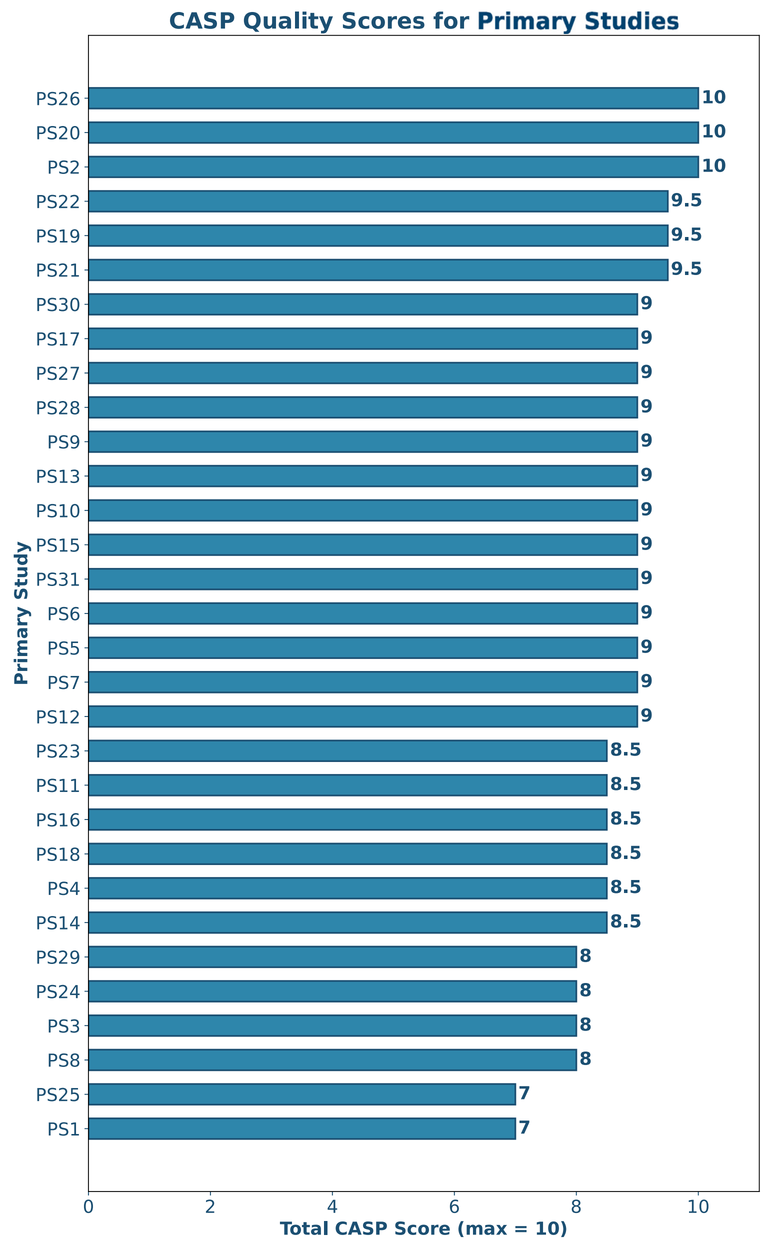

Fig. 7. Critical Appraisal Skills Programme (CASP) quality assessment (QA) scores for the 31 primary studies (PS). Each bar reports the total QA score for a primary study, computed by summing the ten CASP criteria scored using the 1/0.5/0 rubric (Tables 5–6; maximum total = 10). Studies are ordered by total score to summarize overall methodological reporting quality in the included corpus.

CALCE data with proprietary datasets to validate model performance across different chemistries and operational scenarios, while PS21 and PS31 leverage CALCE datasets for cross-domain transfer learning and feature mapping. However, CALCE subsets commonly have small sample counts and are often room-temperature and lithium cobalt oxide (LCO)-centric, limiting degradation diversity and cross-chemistry transferability (PS21, PS12, PS31).

Five studies (i.e., PS3, PS6, PS11, PS20, and PS31) incorporate Oxford battery datasets [120] (small-N, full life-cycle) into their analyses. These datasets are particularly important for research focused on data scarcity and augmentation, as their full life-cycle data and relatively small sample sizes make them ideal for testing the effectiveness of GAN-based data augmentation techniques. Studies such as PS3 and PS6 demonstrate how GANs can improve model accuracy and robustness when training data is limited, while PS20 and PS31 use Oxford datasets for transfer learning and cross-dataset validation. Still, Oxford sets feature few cells and narrow protocols, yielding limited degradation-mode coverage and exacerbating cross-dataset shift (PS3, PS6, PS20, PS31).

Other notable public datasets include the LG 18650HG2 Li-ion Battery Dataset [121] (high-frequency EV-like; PS2, PS18), IEEE DataPort Automotive Li-ion Cell Usage Dataset (IEEE-hosted repository; real EV drive cycles; PS24) [122], PowerStream LiR2032 Coin Cell Battery Dataset (electrochemical impedance spectroscopy (EIS)-rich, SOH; PS16) [123], and the Sandia National Laboratories LFP Battery Dataset (domain adaptation; PS27) [124]. Each of these datasets is selected for its unique characteristics, such as high-frequency real-world cycling data (LG 18650HG2), realistic electric vehicle drive cycles (IEEE DataPort), rich impedance spectra (PowerStream), and domain adaptation opportunities (Sandia).

Despite their value, NASA, CALCE, and Oxford corpora exhibit constraints that induce domain shift and limit degradation diversity. NASA Prognostics Center of Excellence (PCoE) subsets (e.g., B0005–B0018;

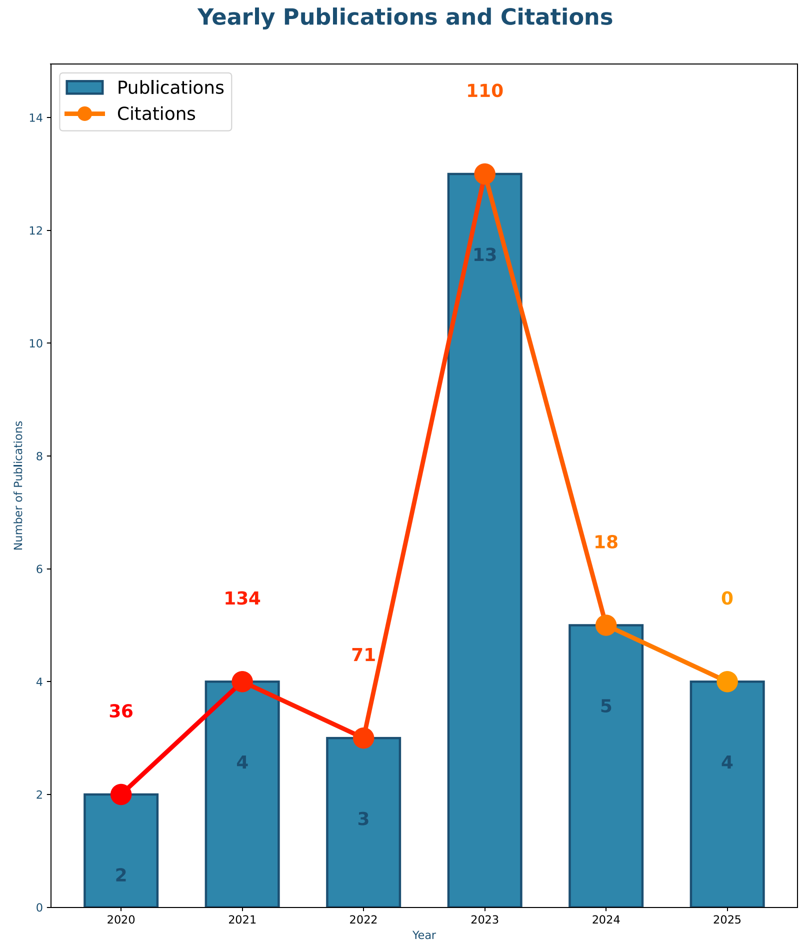

Fig. 8. Temporal distribution of publications and citations from 2020 to 2025. The bars represent the number of publications per year, while the curve shows the cumulative citations for each year.
Data source: Scopus; snapshot: 14 Feb 2025; artifacts on the Open Science Framework (OSF).

randomized vs. fixed protocols) are lab-conditioned (single/limited temperatures, CCCV/CC procedures) and chemistry-specific, leading to distribution mismatch with real EV profiles and other chemistries (see PS21, PS24, PS31). CALCE sets (e.g., CS2; CS35–CS37) are small-N, LCO-centric, and mostly room-temperature, constraining variability in aging trajectories and transfer across chemistries and duty cycles (PS21, PS12, PS31). Oxford datasets provide full life-cycle curves but few cells and narrow protocols, yielding limited degradation-mode coverage and magnifying cross-dataset shift (PS3, PS6, PS20, PS31). These constraints frequently motivate transfer learning, feature mapping, and augmentation to broaden condition coverage and improve generalization (PS20, PS21, PS31).

In summary, the NASA, CALCE, and Oxford datasets are the backbone of GAN-based lithium-ion battery state estimation and prognostics research, with their frequent use across numerous primary studies providing a strong foundation for methodological development and progress in the field. This predominance reflects their public availability, rich instrumentation, and longitudinal (run-to-failure) coverage, alignment with SOC/SOH/RUL tasks, and support for reproducible benchmarking across studies. The explicit referencing of PSs in this analysis highlights the central role these datasets play in advancing both the science and practice of battery state estimation and prognostics, while noting domain shift and limited degradation diversity.

4.3. RQ2: What GAN architectures and data-augmentation techniques have been applied, and what motivated their use?

Analysis of the primary studies shows that Time-series GANs, especially TimeGAN (embedding–recovery–supervised components) and its variants, are most frequently adopted in battery state estimation and prognostics for their ability to capture temporal dynamics and generate high-fidelity, heterogeneous sequences. These models are chosen for state-of-charge (SOC) and state-of-health (SOH) estimation because they can effectively model sequential battery data, address data scarcity, and provide realistic synthetic samples that support robust machine learning estimators. Their methodological appeal lies in their capacity to preserve autoregressive dynamics and temporal dependencies, which are essential for accurate battery state modeling.

As illustrated in Fig. 12, systematic analysis reveals distinct patterns in GAN family adoption: WGAN variants emerge as the most frequently used family with 9 studies (29.0%), followed by TimeGAN variants in 7 studies (22.6%), basic GAN architectures in 6 studies (19.4%), and conditional GANs (CGANs) in 5 studies (16.1%). Specialized architectures, including DCGAN, LSGAN, InfoGAN, and VAE-GAN, each appear in 1 study (3.2% each). The flow diagram clearly shows that the WGAN variants dominate the landscape, with strong connections to the SOH and SOC estimation tasks. The most prevalent specific variants are WGAN-GP implementations (6 studies, 19.4%), TimeGAN variants (7 studies, 22.6%), and conditional GAN variants (5 studies, 16.1%). RUL-focused works (PS3, PS8, PS12, PS13, PS24, PS26) often adopt recurrent or

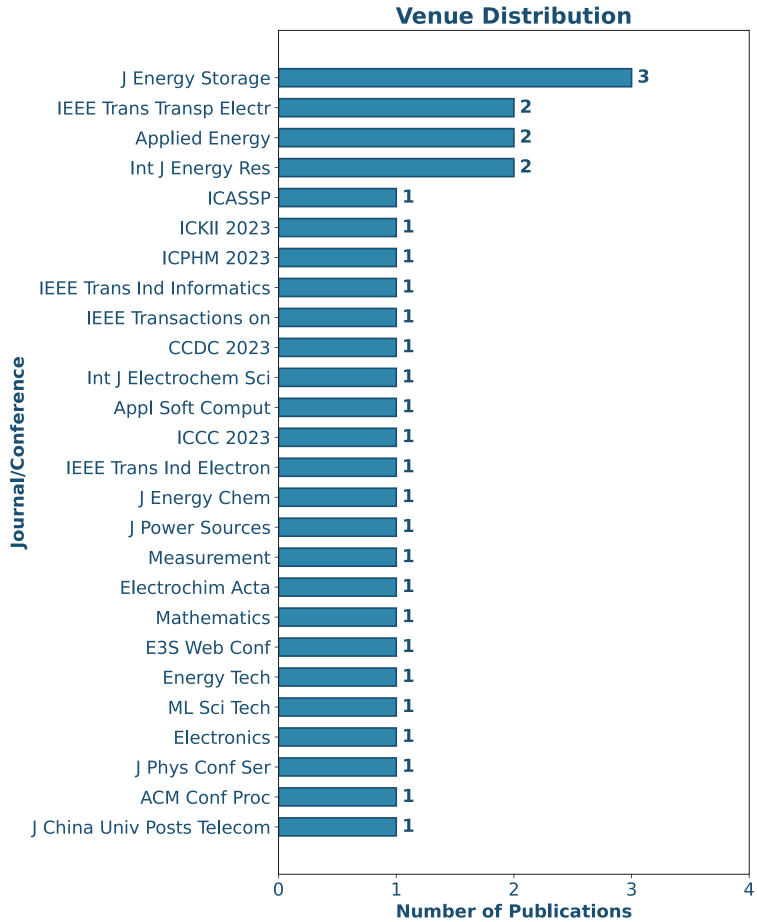

Fig. 9. Venue distribution of the 31 included primary studies. Horizontal bars report the number of publications appearing in each journal or conference venue (as recorded by the publishers), with counts annotated at the bar ends and venues ordered by frequency. The figure highlights that the evidence base is predominantly journal-published, while conference proceedings contribute a smaller share.
Data source: Publisher records; snapshot: 14 Feb 2025; artifacts on the Open Science Framework (OSF).

convolutional-recurrent generators for long-sequence stability, whereas EIS-centric SOH studies (e.g., PS14, PS16) prefer InfoGAN/VAE-GAN to learn robust latent structure.

Wasserstein GANs (WGAN, WGAN-GP, and hybrids) are widely used due to their stable training dynamics and ability to generate diverse, realistic data. The rationale for selecting these architectures is their robustness against mode collapse and their suitability for modeling long-sequence temporal features, which is crucial for SOC, SOH, and remaining useful life (RUL) tasks. The use of gradient penalty and conditional extensions further enhances their stability and adaptability to different battery prognostic scenarios. These models are often integrated with other deep learning components, such as LSTM or CNN modules, to further improve their ability to capture complex battery behaviors and ensure the generated data is both realistic and useful for downstream tasks.

The flow diagram in Fig. 12 provides a comprehensive visualization of these architectural choices and their application-specific patterns. The diagram reveals that SOH estimation (15 studies, 48.4%) dominates the application space, with WGAN variants (6 studies) and CGAN variants (4 studies) being the most frequently employed architectures for this task. SOC estimation (8 studies, 25.8%) shows a preference for TimeGAN variants (4 studies) and WGAN variants (3 studies), reflecting the importance of temporal modeling for charge state prediction. RUL prediction (6 studies, 19.4%) demonstrates more diverse architectural choices, with basic GAN variants (2 studies), WGAN variants (2 studies), and specialized architectures (LSGAN, TimeGAN, CR-GAN) each contributing to the field. Mixed applications (2 studies, 6.5%) primarily utilize WGAN variants, indicating their versatility across multiple prognostic tasks. These choices align with dataset characteristics summarized in RQ1: NASA/CALCE time-series corpora favor TimeGAN/WGAN hybrids for SOC/SOH, while EIS-oriented datasets motivate InfoGAN/VAE-GAN for SOH capacity mapping (see Fig. 11 and the OSF dataset inventory). The distribution reflects the varying complexity and requirements of different prognostic tasks, with SOH estimation benefiting from the stability of WGAN variants and the temporal modeling capabilities of TimeGAN, while RUL prediction requires more specialized architectures to handle long-sequence forecasting challenges. For the full per-study inventory of GAN families and specific variants, see the OSF data extraction table (snapshot details therein): https://osf.io/rmj8b/.

Conditional GANs (CGANs) and their extensions are primarily chosen for SOH and RUL estimation, especially when data must be generated under specific conditions or for reconstructing missing or incomplete information. These models are valued for their flexibility in conditioning on various battery states, degradation modes, or operational scenarios, enabling more targeted and relevant data augmentation for complex or data-limited settings. The increasing use of conditional and cluster-based GANs reflects a trend towards more specialized and context-aware data generation strategies in battery state estimation and prognostics.

Other notable architectures, such as BiLSTM-based GANs, Transformer-GANs, and hybrid models (including VAE-GAN, CR-GAN, and CLSTM-CGAN), are selected for their ability to exploit spatial–temporal patterns, extract features from complex data, or leverage

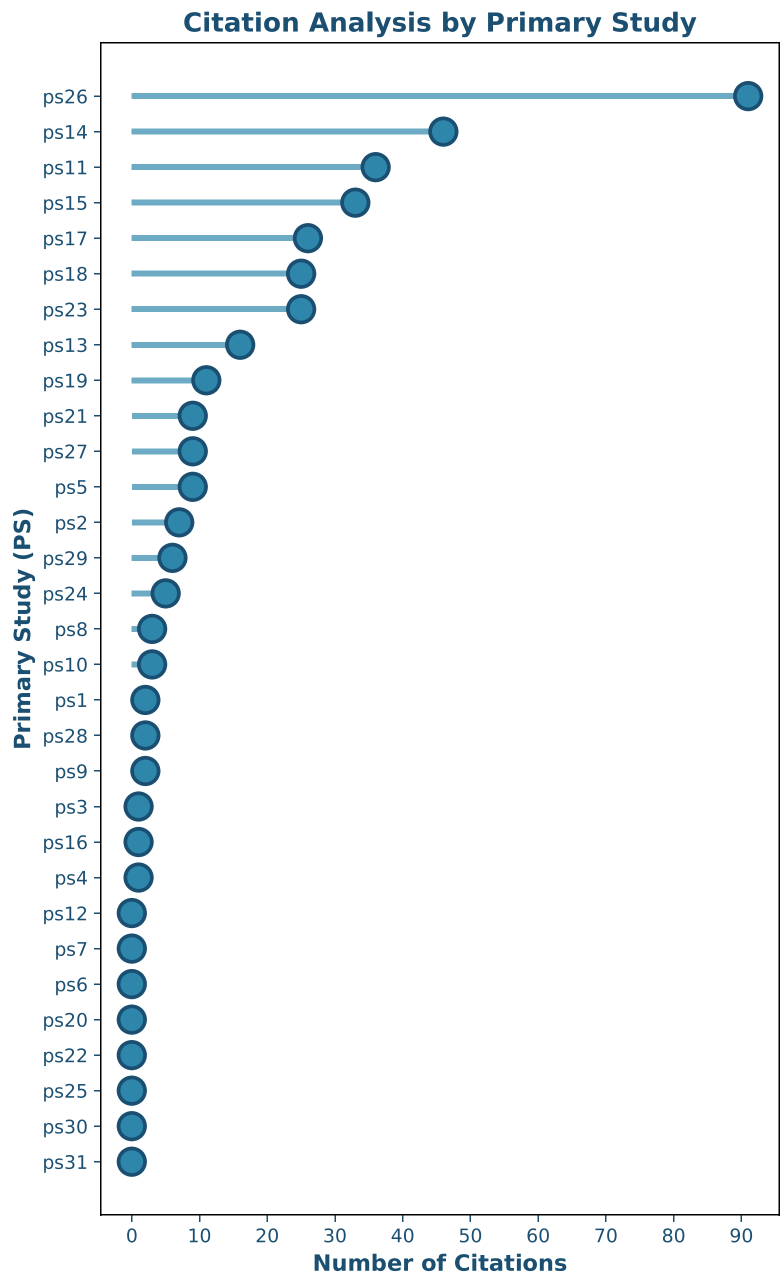

Fig. 10. Citation analysis for the 31 included primary studies (PS). Each point shows the total Scopus citation count for a primary study (y-axis lists PS identifiers), and the horizontal stems emphasize the magnitude of citations on the x-axis; studies are ordered by citation count to highlight the most-cited contributions. This visualization provides a quick view of citation skew (a small number of highly cited studies versus many low-citation or recent studies).
Data source: Scopus; snapshot: 14 Feb 2025; artifacts on the Open Science Framework (OSF).

Table 7
Summary of reported Generative Adversarial Network (GAN) training and model-configuration choices across the primary studies. Rows list configuration categories (e.g., GAN family, learning rate, batch size, optimizer, regularization, downstream architecture, and training duration), with the most commonly reported setting, an interpretive takeaway, and representative primary studies (PS) supporting each entry.

Category	Most common (%)	Key takeaway	PS references
GAN architecture	WGAN variants (29.0%)	WGAN: stability; TimeGAN: temporal fidelity	PS2, PS7, PS8, PS10, PS17, PS18, PS20, PS22
Learning rate	1e−3 (45.2%)	Lower LR improves stability	PS2, PS17, PS18, PS20, PS21, PS22, PS29, PS30
Batch size	64 (38.7%)	Larger batches aid convergence	PS2, PS17, PS19, PS21, PS22
Optimizer	Adam (35.5%)	Adam balances speed/stability	PS2, PS6, PS14, PS15, PS16, PS17, PS20, PS21, PS22
Regularization	Gradient penalty (25.8%)	GP essential for WGAN stability	PS2, PS17, PS19, PS20
Deep learning architecture	LSTM (48.4%)	Sequential models most effective	PS2, PS3, PS5, PS7, PS9, PS10, PS11, PS13, PS15, PS17, PS19, PS21, PS22, PS26, PS27, PS28, PS29, PS30
Training duration	30k–50k iters (25.8%)	Longer training helps	PS2, PS5, PS21, PS22, PS24

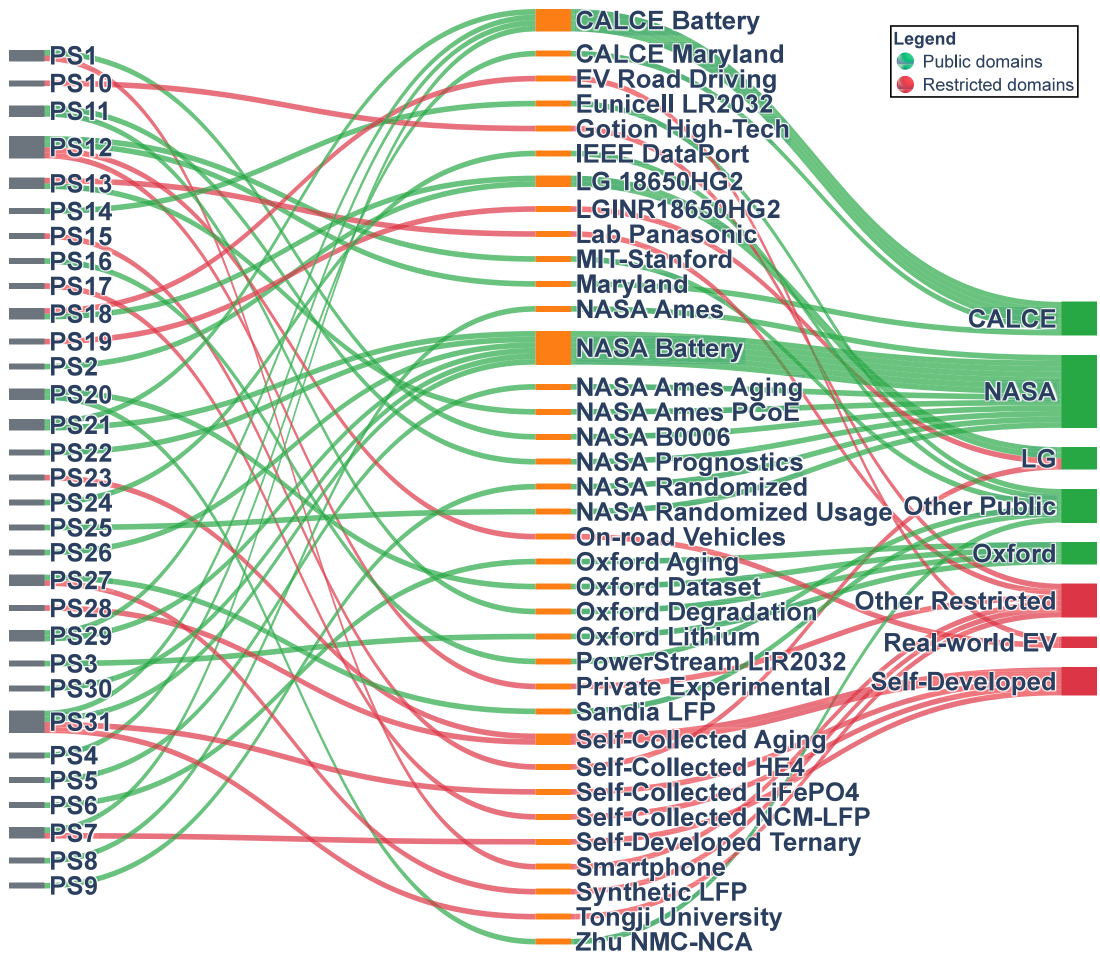

Fig. 11. Overview of dataset usage patterns in Generative Adversarial Network (GAN)-based lithium-ion battery state estimation and prognostics research. Flows summarize task usage for State of Charge (SOC), State of Health (SOH), and Remaining Useful Life (RUL) across the National Aeronautics and Space Administration (NASA), Center for Advanced Life Cycle Engineering (CALCE), and Oxford datasets; edge widths indicate frequency across primary studies (PS).

unsupervised learning for feature extraction. These choices are motivated by the need to address specific data characteristics, such as multi-signal sequences, impedance spectra, or run-to-failure patterns, and to enhance the generalization and adaptability of prognostic models. The diversity of GAN architectures in the literature highlights the importance of matching model capabilities to the unique challenges posed by different battery datasets and prognostic objectives.

Overall, the rationale for the widespread use of TimeGAN, WGAN, CGAN, and hybrid GAN architectures in battery state estimation and prognostics is their methodological strength in addressing data scarcity, modeling complex temporal and conditional relationships, and supporting the development of reliable and generalizable battery state estimation and prognostic systems. As the field advances, there is a clear movement towards integrating multiple deep learning paradigms and designing GANs that are increasingly tailored to the nuanced requirements of battery health monitoring and prediction.

4.4. RQ3: How effective is GAN-based data augmentation in improving state estimation and prognostics tasks (SOC, SOH, and RUL)?

Generative Adversarial Networks (GANs) have been shown to be effective for data augmentation in the field of lithium-ion battery state estimation and prognostics, particularly for tasks such as state of charge (SOC), state of health (SOH), and remaining useful life (RUL) estimation. The core motivation behind employing GAN-based augmentation is to address the scarcity and imbalance of high-quality battery datasets, which are critical for training robust deep learning models. By generating realistic synthetic data that mimics the statistical properties of real measurements, GANs can enhance model generalizability, reduce overfitting, and improve predictive accuracy. The following sections systematically compare the performance of base models and their GAN-augmented counterparts, highlighting the tangible benefits and limitations observed across a range of studies.

We organize RQ3 by task and report matched baseline versus GAN-augmented results.

4.4.1. Preprocessing and its impact on GAN performance

Preprocessing can materially affect GAN training stability, synthetic-data fidelity, and downstream SOC, SOH, and RUL accuracy because it alters the scale, noise structure, and temporal coherence presented to the generator and discriminator. This sensitivity is particularly relevant in our corpus because commonly used battery benchmarks span heterogeneous operating conditions and sampling regimes (Section 4,

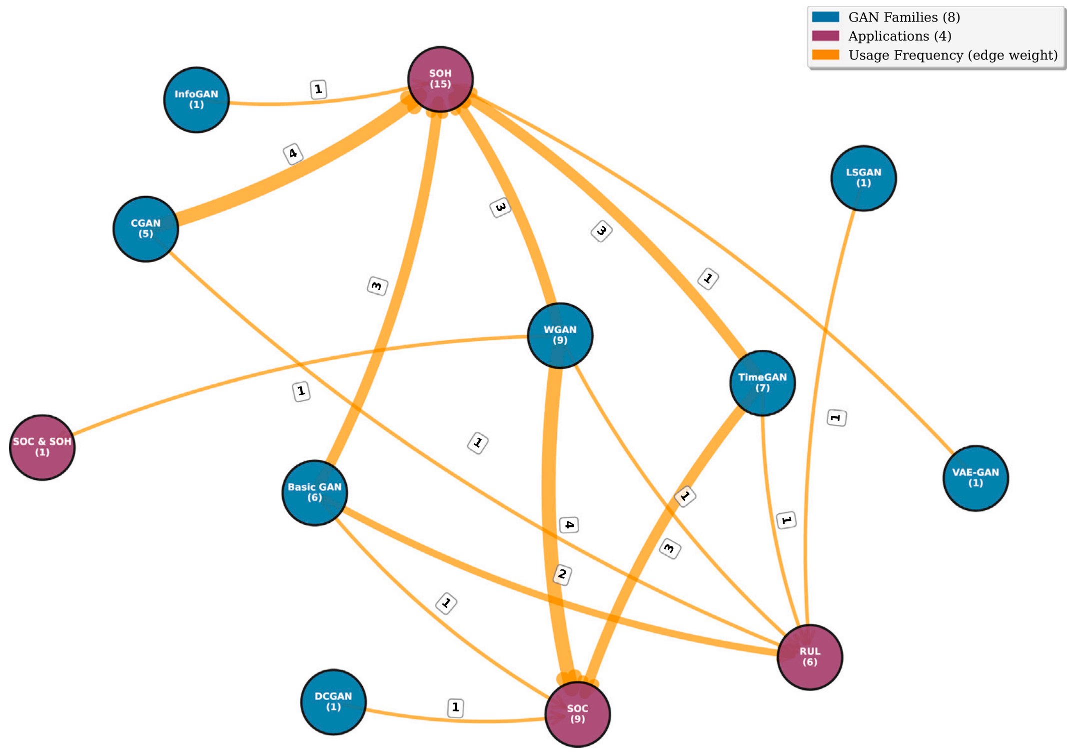

Fig. 12. Flow diagram illustrating the distribution of Generative Adversarial Network (GAN) architectures and their applications in lithium-ion battery state estimation across 31 primary studies (PS). The diagram links GAN families—including Wasserstein GAN (WGAN), Time-series GAN (TimeGAN), conditional GAN (CGAN), Deep Convolutional GAN (DCGAN), Least-Squares GAN (LSGAN), Information Maximizing GAN (InfoGAN), and Variational Autoencoder–GAN (VAE-GAN)—to target tasks: State of Charge (SOC), State of Health (SOH), Remaining Useful Life (RUL), and mixed tasks. Edge weights indicate usage frequency.

Table 8
Preprocessing reporting across the 31 primary studies (PS1–PS31), where PS denotes primary study. The Count column gives the number of studies that explicitly report each preprocessing element. Examples list representative operations as described by the source papers (e.g., denoising/cleaning methods, windowing choices, scaling policies, and split granularity). Leakage safeguards refer to procedures such as fitting scalers on the training split only and restricting augmentation to training folds; these were not explicitly reported in the corpus.

Item	Count	Representative evidence (examples)
Denoising/cleaning	3	PS4 (variational mode decomposition (VMD)); PS22 (Savitzky–Golay, imputation); PS28 (abnormal-data removal + normalization).
Windowing/sequence formatting	8	PS3 (window = 11, stride = 1); PS11 (sequence length = 24); PS30 (convolutional sliding window, stride = 1).
Scaling/normalization	10	PS9 (min–max); PS22 (standardization); PS25 (scaler comparison); PS26 (min–max by variable).
Split granularity	9	PS4 and PS30 (70/30 time-wise); PS6 (cell-wise); PS22 (cross-cell train/validate); PS24 (trip-wise).
Leakage safeguards (train-only fit; aug train-only)	0	Not explicitly reported in the corpus.

RQ1), making normalization, windowing, cleaning, and split design consequential for both comparability and reproducibility.

Across the corpus, explicit reporting of preprocessing is uneven (Table 8). Only three studies report a concrete denoising or cleaning step (e.g., VMD decomposition in PS4; Savitzky–Golay smoothing and imputation in PS22; cleaning and normalization in PS28). Windowing or explicit sequence formatting is reported in eight studies (e.g., PS3, PS8–PS11, PS26, PS30), and scaling/normalization in ten studies (e.g., PS3, PS4, PS9, PS18, PS21, PS22, PS24–PS26, PS28). Split schemes are reported in nine studies (e.g., time-wise 70/30 in PS4 and PS30; cell-wise in PS6; trip-wise in PS24; cross-cell validation in PS22). Notably, none of the studies explicitly states leakage safeguards such as fitting scalers/transforms on training data only and restricting augmentation to training folds; this under-reporting limits causal interpretation of reported gains and hinders replication.

Evidence that preprocessing affects GAN outcomes is clearest when preprocessing is varied directly. PS25 evaluates multiple scaling options before TimeGAN training and reports sizable changes in fidelity metrics (discriminative loss 0.135→0.0625; predictive loss 0.195→0.1135) when moving beyond the default min–max configuration, indicating that the scaling policy can alter both realism diagnostics and downstream utility. In contrast, most studies report preprocessing only as part of an end-to-end pipeline (e.g., incremental capacity (IC)-derived health factors in PS10; smoothing/standardization with feature selection in PS22; decomposition-based denoising in PS4) without an ablation that isolates preprocessing from other modeling choices. Accordingly, our corpus-level evidence supports a conservative interpretation: preprocessing is a plausible contributor to GAN stability and performance, but it is rarely documented—or experimentally varied—in enough detail to support strong causal attribution or robust cross-study comparison.

To improve comparability and reproducibility, a minimal preprocessing reporting checklist should include: (i) the exact scaler/normalization policy and whether it is fit on the training split only; (ii) windowing/sequence length and stride (if applicable); (iii) cleaning/denoising and outlier rules; (iv) split granularity (time-wise, cell-wise, trip-wise) and leakage safeguards (augmentation restricted to training folds); and (v) whether synthetic samples are filtered/selected prior to fusion.

For full per-study preprocessing fields and page-level evidence, see the OSF extraction table (https://osf.io/rmj8b/).

Table 9
State of Health (SOH) results from matched baseline versus Generative Adversarial Network (GAN)-augmented comparisons. PS denotes the primary study identifier. Base and GAN specify the downstream estimator and the augmentation method used in that study. Metric lists the reported error measure(s); w/o and w/ report the baseline (without GAN augmentation) and GAN-augmented values (with GAN augmentation), and *Δ* reports the relative improvement (decrease) within the same study/metric (units may differ across studies, e.g., mAh vs. %).

PS	Base	GAN	Metric	w/o	w/	Δ
PS6	NN (capacity estimation)	CGAN-SQD	RMSE	12.16 mAh	1.20 mAh	↓90%
PS7	BiLSTM	WGAN-BiLSTM	RMSE; MAPE	RMSE 1.93%; MAPE 1.42%	RMSE 0.42%; MAPE 0.29%	RMSE ↓78%; MAPE ↓80%
PS9	LSTM/GRU	TimeGAN	RMSE	0.0451	0.0336	↓26%
PS10	LSTM	WTimeGAN	MAE; RMSE	MAE 0.047; RMSE 0.216	MAE 0.012; RMSE 0.107	RMSE ↓50%; MAE ↓75%
PS14	GPR	EISGAN	MAE; RMSE	MAE 3.1–5.1 mAh; RMSE 3.2–5.1 mAh	MAE ≤1.74 mAh; RMSE ≤1.87 mAh	MAE ↓44%–66%; RMSE ↓42%–63%
PS16	VAE	VAEGAN	RMSE; MAE	RMSE 3.24 mAh; MAE 2.59 mAh	RMSE 2.91 mAh; MAE 2.38 mAh	↓10%
PS21	LSTM-FC	GAN-LSTM-TL	RMSE; MAE	RMSE 0.0390; MAE 0.0284	RMSE 0.0275; MAE 0.0208	RMSE ↓30%; MAE ↓27%
PS22	BiLSTM	BiLSTM-WGAN	RMSE; MAE	RMSE 0.391; MAE 0.315	RMSE 0.215; MAE 0.136	RMSE ↓45%; MAE ↓57%
PS23	LSTM	RGAN	RMSE; MAE	RMSE 0.004; MAE 0.003	RMSE 0.0036; MAE 0.0028	↓10%
PS27	LSTM	CTSGAN	RMSE; MAPE	RMSE 0.357%; MAPE 0.042	RMSE 0.193%; MAPE 0.019	RMSE ↓46%; MAPE ↓55%
PS30	TCNN-Trans	TimeGAN	Average root mean square error (ARMSE)	0.007	0.0032	↓54%

4.4.2. State of Health (SOH)

For SOH, Table 9 reports matched baseline-versus-augmented comparisons, where w/o denotes training without GAN-generated samples, w/ denotes training with GAN-augmented data, and *Δ* denotes the relative improvement. Because SOH is reported using heterogeneous metrics and scales across studies (e.g., mAh-based errors versus percentage-based errors), interpretation is restricted to within-study matched comparisons rather than cross-study ranking.

Across the included SOH studies, GAN-augmented training is associated with error reductions in multiple matched comparisons across datasets and downstream estimators. The largest gain is reported in PS6, where RMSE decreases from 12.16 mAh to 1.20 mAh (90% reduction) under CGAN-SQD, indicating that conditional synthesis combined with explicit sample-quality control can substantially improve SOH estimation under scarcity or imbalance. A representative mid-range gain is reported in PS22, where RMSE decreases from 0.391 to 0.215 (45% reduction) and MAE decreases from 0.315 to 0.136 under a BiLSTM-WGAN configuration, supporting the pattern that augmentation can improve multiple error measures concurrently. Modality-aligned evidence is also reported for EIS-based SOH: in PS14, MAE is reduced by 44%–66% and RMSE by 42%–63% under EISGAN, consistent with the importance of aligning the generative design to the measurement modality.

Overall, larger gains tend to be reported when augmentation is guided by conditioning and quality selection (e.g., SQD or similar gating) rather than indiscriminate oversampling, and when the generative design is aligned to the measurement modality (cycling time-series versus EIS) so that task-relevant structure is preserved. Gains are also reported across heterogeneous downstream estimators (e.g., GPR, LSTM/BiLSTM, and Transformer-based models), suggesting that augmentation benefits are not confined to a single estimator family.

4.4.3. State of Charge (SOC)

SOC comparisons in Table 10 indicate that GAN-based augmentation can reduce estimation error under matched baseline-versus-augmented evaluations. The largest gains are reported in studies that combine temporally aware modeling with stabilized adversarial training. For instance, PS18 reports RMSE decreasing from ∼2.5% to ∼0.75% and MAE from ∼2.2% to ∼0.6% under TS-WGAN, while PS17 reports RMSE decreasing from ∼2% to <1% and MAE from ∼1% to <0.5% under C-LSTM-WGAN-GP. Conditioning can further amplify gains: PS15 (GAN-CLS + BiLSTM) reports an MSE reduction from 0.011 to 0.0013 (88%).

Collectively, the SOC evidence suggests that augmentation is most effective when the generated samples preserve drive-cycle-dependent temporal structure and when training is stabilized to avoid mode collapse in time-dependent synthesis (e.g., via WGAN-GP-style regularization). Conditioning that leverages cross-signal structure (e.g., SOC–OCV context) is also associated with stronger gains in the matched comparisons.

Because SOC metrics are reported on heterogeneous scales across studies (percentage-based errors versus absolute-scale errors), the evidence is interpreted within-study rather than used for cross-study ranking. These patterns are consistent with the configuration summary in Table 12.

4.4.4. Remaining Useful Life (RUL)

For long-horizon RUL prediction, the matched comparisons in Table 11 show that GAN-augmented training is associated with error reductions despite scarce run-to-failure trajectories. The largest gain is reported in PS8, where WGAN-fusion reduces absolute error from 6 to 1 (83% reduction) and halves RMSE from 0.020 to 0.010. Additional improvements are reported under both conditional and recurrent generators: PS12 (CGAN) reduces RMSE from 7.7% to 6.4% and MAE from 6.3% to 5.1% (17% and 19% reductions, respectively), while PS13 (GRU-LSGAN) reduces MAE from 2.4% to 0.66% (72% reduction). Hybrid convolutional–recurrent generation also shows benefits under more complex patterns (PS26: RMSE 23.57→18.39; 22% reduction).

Taken together, these matched comparisons suggest that RUL gains are most often reported when generation remains stable over long sequences and preserves degradation-trend structure. In particular, stabilized objectives (e.g., Wasserstein or least-squares formulations) and temporal encoders/decoders (e.g., GRU/BiLSTM or convolutional–recurrent hybrids) appear repeatedly in higher-gain pipelines, consistent with the need to model long-range dependencies in aging trajectories.

Because RUL outcomes are reported with heterogeneous metrics and scales across studies (e.g., AE, RMSE, MAE, and percentage-based errors), the evidence is interpreted within-study rather than used for cross-study ranking. These RUL patterns align with the configuration trends summarized in Table 12.

Across SOH, SOC, and RUL, the matched comparisons (Tables 9, 10, and 11) indicate that GAN-based augmentation is frequently associated with lower estimation and prognostic errors relative to non-augmented baselines. Where direct before–after comparisons are available, reported improvements commonly fall in the tens-of-percent range across RMSE, MAE, MAPE, and MSE, with larger gains reported in a subset of studies. Because metrics and scales vary across studies, this synthesis is interpreted at the within-study level rather than used for cross-study ranking.

Table 10
State of Charge (SOC) results from matched baseline versus Generative Adversarial Network (GAN)-augmented comparisons. PS denotes the primary study identifier. Base and GAN specify the estimator and augmentation method used in that study. Metric lists the reported error measure(s); w/o and w/ report the baseline (without GAN augmentation) and GAN-augmented values (with GAN augmentation), and *Δ* reports the within-study relative improvement (decrease) for the same metric (metrics/scales vary across studies, e.g., % vs. absolute).

PS	Base	GAN	Metric	w/o	w/	*Δ*
PS4	BiLSTM	VMD-Transformer-GAN	RMSE; MAE; MAPE	RMSE 0.547; MAE 0.363; MAPE 21.98	RMSE 0.446; MAE 0.264; MAPE 18.59	RMSE ↓18%; MAE ↓27%; MAPE ↓15%
PS15	EKF	GAN-CLS + BiLSTM	MSE	0.011	0.0013	↓88%
PS17	LSTM	C-LSTM-WGAN-GP	RMSE; MAE	RMSE ∼2%; MAE ∼1%	RMSE <1%; MAE <0.5%	↓50%
PS18	Trans-former	TS-WGAN	RMSE; MAE	RMSE ∼2.5%; MAE ∼2.2%	RMSE ∼0.75%; MAE ∼0.6%	↓70%
PS19	LSTM	TS-DCGAN	RMSE; MAE	RMSE 2.10%; MAE 1.65%	RMSE 1.74%; MAE 1.38%	↓17%

Table 11
Remaining Useful Life (RUL) results from matched baseline versus Generative Adversarial Network (GAN)-augmented comparisons. PS denotes the primary study identifier. Base and GAN specify the predictor and augmentation method used in that study. Metric lists the reported error measure(s), such as absolute error (AE), root mean square error (RMSE), and mean absolute error (MAE). The w/o and w/ columns report baseline (without GAN augmentation) and GAN-augmented values (with GAN augmentation), and *Δ* reports the within-study relative improvement (decrease) for the same metric (metrics/scales vary across studies).

PS	Base	GAN	Metric	w/o	w/	*Δ*
PS3	BiLSTM	GAN-BiLSTM	MSE-v	0.00343	0.00154	↓55%
PS8	Empirical degradation model	WGAN-fusion	AE; RMSE	AE 6; RMSE 0.020	AE 1; RMSE 0.010	AE ↓83%; RMSE ↓50%
PS12	Transformer	CGAN	RMSE; MAE	RMSE 7.7%; MAE 6.3%	RMSE 6.4%; MAE 5.1%	RMSE ↓17%; MAE ↓19%
PS13	GRU	GRU-LSGAN	MAE	2.4%	0.66%	↓72%
PS26	TCMN	CR-GAN	RMSE	23.57	18.39	↓22%

Table 12
Representative Generative Adversarial Network (GAN) configurations and reported error reductions for research question 3 (RQ3). Each row groups a commonly reported configuration family observed in the corpus. Reported root mean square error (RMSE) ↓ indicates the reduction level reported by the referenced primary studies for matched baseline versus GAN-augmented comparisons under that configuration family (values are taken from the source studies and are not meta-analytic averages). Key advantage summarizes the typical rationale reported for the configuration; PS references list representative primary studies (PS) using the configuration.

Configuration	Reported RMSE ↓	Key advantage	PS references
WGAN-GP + Adam + Batch 64	45%	High training stability	PS2, PS17, PS19, PS20, PS22
TimeGAN + LSTM + LR 1e−3	38%	Effective temporal modeling	PS1, PS9, PS11, PS24, PS25, PS30
Conditional GAN + Gradient Penalty	52%	Large accuracy gains	PS6, PS12, PS15, PS17, PS19, PS20, PS28, PS29
Transfer Learning (with GAN aug)	40%	Cross-dataset generalization	PS20, PS21, PS27, PS31
Quality Filtering for Synth Data	28%	Improved data reliability	PS6, PS10, PS12, PS31

The cross-task evidence further indicates that effectiveness depends not only on the GAN family but also on stability controls, conditioning, and data hygiene (e.g., sample-quality selection and preprocessing consistency). To make these recurring implementation choices explicit, Table 12 summarizes representative configurations and reported error-reduction levels across the corpus; the complete study-level inventory is available in the OSF extraction table (https://osf.io/rmj8b/).

Table 12 summarizes commonly reported effective configurations, highlighting the recurrent co-occurrence of (i) stability-oriented objectives (e.g., WGAN-GP-style regularization), (ii) temporally aware components for sequential signals (e.g., LSTM/Transformer modules), and (iii) conditioning or quality filtering when data are scarce, imbalanced, or heterogeneous. Transfer learning is also recurrent in studies that explicitly evaluate cross-dataset or cross-condition generalization.

4.5. RQ4: What technical challenges and limitations exist for GAN-based methods in this field?

This research question examines the technical challenges and limitations that researchers face when applying GAN-based methods to lithium-ion battery state estimation. The analysis focuses on identifying the key obstacles that hinder the widespread adoption and effectiveness of GAN approaches in battery state estimation and prognostics. Understanding these challenges is crucial for guiding future research directions and improving the practical implementation of GAN-based data augmentation techniques in battery management systems.

As summarized in Fig. 13, data scarcity dominates, followed by synthetic-data quality, GAN instability, and domain shift. Categories are non–mutually exclusive, so percentages need not sum to 100%. Per-study categorizations and mitigation mappings are available in the OSF extraction table (https://osf.io/rmj8b/). Challenge categories were dual-coded by two reviewers with consensus resolution; agreement on initial independent coding was monitored and discrepancies were resolved by discussion (see OSF coding ledger).

4.5.1. Analysis of technical challenges and mitigation strategies

Four themes dominate across the corpus: data scarcity (∼94%), synthetic-data quality (∼48%), GAN instability (∼26%), and domain shift (∼19%). The systematic analysis of 31 primary studies reveals distinct patterns in the technical challenges faced by GAN-based lithium-ion battery state estimation approaches and the corresponding mitigation strategies employed. Data scarcity emerges as the most pervasive

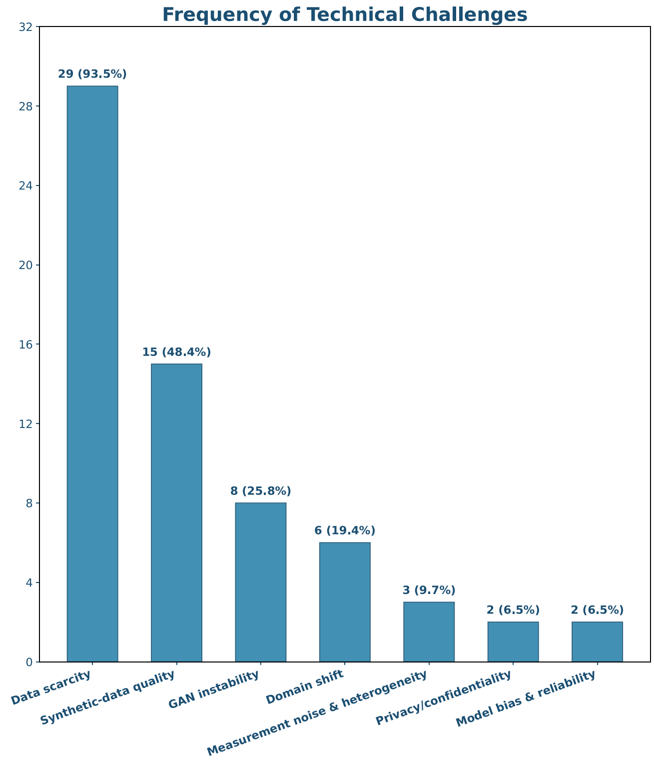

Fig. 13. Frequency of technical challenge categories across the 31 primary studies (PS) for research question 4 (RQ4). Bars report the number (and percentage) of studies mentioning each challenge category (non-mutually exclusive coding): data scarcity, synthetic-data quality, Generative Adversarial Network (GAN) instability, domain shift, measurement noise/heterogeneity, privacy/confidentiality, and model bias/reliability. Counts and percentages are annotated above each bar.

challenge, affecting 29 out of 31 studies (93.5%), followed by synthetic-data quality concerns in 15 studies (48.4%), GAN instability in 8 studies (25.8%), and domain shift issues in 6 studies (19.4%). These statistics highlight the fundamental nature of data limitations in battery state estimation and prognostics and the critical importance of addressing synthetic data quality and training stability.

4.5.2. Data scarcity: The primary challenge

Data scarcity represents the most significant technical challenge, with 29 studies explicitly addressing this limitation. The high prevalence of this challenge reflects the fundamental constraint in battery state estimation and prognostics research, where obtaining comprehensive, high-quality experimental data is both time-consuming and costly. Studies such as PS1, PS2, PS3, PS5, PS6, PS9, PS11, PS12, PS15, PS17, PS19, PS20, PS21, PS22, PS26, PS28, PS29, PS30, and PS31 all implement various GAN architectures specifically designed to generate synthetic data that can supplement limited real-world measurements. The most common mitigation strategies include TimeGAN variants (PS1, PS9, PS11, PS25, PS30), Conditional GANs (PS6, PS12, PS15, PS17, PS28, PS29), and Wasserstein GAN approaches (PS2, PS7, PS10, PS20, PS22). These approaches demonstrate that GANs can effectively address data scarcity by generating realistic synthetic samples that maintain the statistical properties and temporal patterns of real battery data.

4.5.3. Synthetic-data quality: Ensuring reliability

Synthetic-data quality concerns appear in 15 studies (48.4%), indicating that while GANs can generate data, ensuring its quality and reliability remains a significant challenge. Studies such as PS2, PS6, PS7, PS12, PS22, and PS31 implement various quality control mechanisms. The most effective strategies include Sample Quality Determination (SQD) frameworks (PS6), diversity-selection modules (PS12), and quality evaluation mechanisms (PS31). These approaches demonstrate that raw GAN outputs often require additional filtering and validation to ensure they meet the stringent requirements of battery state estimation and prognostics applications. The implementation of conditional architectures and quality assessment modules has proven particularly effective in maintaining the electrochemical characteristics and temporal consistency of generated data.

4.5.4. GAN instability: Training challenges

GAN instability affects 8 studies (25.8%), representing a significant technical challenge that can hinder the widespread adoption of GAN-based approaches. Studies such as PS4, PS7, PS10, PS13, PS24, and PS31 address this challenge through various stabilization techniques. The most effective mitigation strategies include gradient penalty mechanisms (PS4, PS7, PS10), alternative loss functions such as Least-Squares loss (PS13), and robust training approaches with Huber/MAE loss functions (PS24). These techniques demonstrate that while GANs can suffer from mode collapse, vanishing gradients, and training instability, these issues can be effectively addressed through careful architectural design and loss function selection.

4.5.5. Domain shift: Generalization limitations

Domain shift challenges appear in 6 studies (19.4%), highlighting the difficulty of applying GAN-based approaches across different battery chemistries, operating conditions, and aging patterns. Sub-cases include cross-chemistry (e.g., PS20, PS31), cross-dataset (e.g., PS21), and cross-temperature/drive-cycle transfer (e.g., PS27). Studies such as PS20, PS21, PS27, and PS31 implement transfer learning approaches to address this limitation. The most effective strategies include W-DC-GAN-GP with transfer learning (PS20), GAN-LSTM with transfer learning (PS21), and Deep CORAL–based knowledge transfer (PS27). These approaches demonstrate that while GANs can struggle with domain adaptation, transfer learning techniques can significantly improve generalization capabilities across different battery types and operating conditions.

4.5.6. Measurement noise and data heterogeneity

Measurement noise and data heterogeneity challenges appear in 3 studies (9.7%), including PS14, PS16, and PS23. These studies implement specialized approaches such as InfoGAN-based EISGAN (PS14), VAEGAN for robust feature extraction (PS16), and Regression GAN for anomaly detection (PS23). These approaches demonstrate that GANs can effectively handle noisy and heterogeneous data by learning robust latent representations and identifying anomalous patterns that might otherwise degrade model performance.

4.5.7. Privacy and confidentiality constraints

Privacy and confidentiality constraints appear in 2 studies (6.5%), specifically PS5 and PS19. These studies implement cloud-based digital twin approaches (PS5) and TS-DCGAN for synthetic data generation (PS19) to address the challenge of limited access to proprietary battery data. These approaches demonstrate that GANs can effectively address privacy concerns by generating synthetic data that maintains the statistical properties of real data without revealing proprietary information.

4.5.8. Model bias and prediction reliability

Model bias and prediction reliability challenges appear in 2 studies (6.5%), specifically PS8 and PS18. These studies implement fusion approaches combining multiple models (PS8) and hybrid architectures capturing both local and global features (PS18) to address the limitations of single-model approaches. These strategies demonstrate that GANs can be effectively integrated with other modeling approaches to improve prediction reliability and reduce model bias.

Table 13
Training stability techniques reported across the primary studies and their stated roles in improving Generative Adversarial Network (GAN) training behavior. Each row lists a stabilization technique (e.g., gradient penalty, optimizer choice, batch size, training duration, quality filtering, transfer learning), the qualitative effect described by the study authors (e.g., reduced mode collapse, improved convergence, improved synthetic-data reliability), and representative primary studies (PS) reporting the technique. Effects are synthesized from study narratives and are not treated as causal estimates.

Technique	Observed effect	PS references
Gradient Penalty (WGAN-GP)	*Mode collapse* ↓, stability ↑	PS2, PS7, PS15, PS17, PS19, PS20, PS22
Adam Optimizer	Convergence ↑ (faster, more stable)	PS2, PS6, PS14, PS15, PS16, PS17, PS18, PS20, PS21, PS22, PS23, PS30
Larger Batch Sizes (≥64)	Training instability ↓	PS2, PS5, PS17, PS19, PS21, PS22, PS24
Longer Training (30k–50k iters)	Performance ↑; diminishing returns >50k	PS2, PS5, PS21, PS22, PS24
Quality Filtering (MMDSQD)	Synthetic data reliability ↑	PS6, PS10, PS12, PS31
Transfer Learning	Cross-dataset generalization ↑	PS20, PS21, PS27, PS31

4.5.9. Comprehensive analysis and implications

The analysis reveals that GAN-based approaches in lithium-ion battery state estimation face multiple interconnected challenges that require sophisticated mitigation strategies. The high prevalence of data scarcity (93.5%) highlights the fundamental need for synthetic data generation in this field, while the significant presence of synthetic-data quality concerns (48.4%) indicates that quality control mechanisms are essential for practical deployment. The moderate prevalence of GAN instability (25.8%) suggests that while training challenges exist, they can be effectively addressed through proper architectural design and loss function selection. Notably, studies that employ gradient-penalty regularization, quality filtering/selection, or transfer learning (e.g., PS2, PS6, PS10, PS20, and PS22) co-occur with sizeable RMSE/MAE gains in RQ3 (Tables 9–11); causal attribution is not claimed. Furthermore, the key stabilizers and controls are summarized in Table 13.

The effectiveness of mitigation strategies varies across challenge domains. For data scarcity, TimeGAN variants and Conditional GANs have proven highly effective, with 19 studies implementing these approaches. For synthetic-data quality, quality determination frameworks and diversity-selection modules have shown positive results, with 15 studies implementing various quality control mechanisms. For GAN instability, gradient penalty mechanisms and alternative loss functions have demonstrated significant effectiveness, with 8 studies implementing these stabilization techniques. Briefly, commonly reported stabilizers include WGAN-GP with gradient penalty ($\lambda_{GP} \approx 10$) and Adam (learning rate $\approx 10^{-3}$); batch sizes ≥ 64 and longer training (30k–50k iterations) further reduce instability, while quality filtering (MMDSQD) improves synthetic-data reliability. These patterns align with the corpus-level hyperparameter analysis in the OSF extraction table (https://osf.io/rmj8b/).

The analysis also reveals important trends in the evolution of GAN-based approaches. Early studies focused primarily on addressing data scarcity through basic GAN architectures, while more recent studies have incorporated sophisticated quality control mechanisms, transfer learning approaches, and hybrid architectures. This evolution suggests that the field is moving towards more robust and practical implementations that can address multiple challenges simultaneously.

4.5.10. Engineering considerations for BMS deployment

Deployment-oriented metadata were synthesized from the per-study extraction table on OSF (https://osf.io/rmj8b/) and summarized as coverage in Fig. 14. Here, *deployability* is treated as the extent to which a study reports the engineering information needed to assess feasibility in a battery management system (BMS), including device class and deployment venue, latency and memory/footprint, power/energy, integration placement, deployment optimizations, and portability artifacts.

Across the 31 studies, reporting is sparse for the metrics that are most consequential for on-vehicle deployment. Device class is specified in 11/31 (35.5%) studies and is reported almost exclusively as training hardware (e.g., Tesla P100, RTX 2080 Ti/3060/3080, i7/Xeon/4090), rather than embedded MCU/DSP/FPGA targets. Deployment venue is reported in 13/31 (41.9%) studies and is dominated by offline or training-only settings; system-level placement is articulated in only two cases (PS23: cloud-to-edge SOH workflow; PS5: embedded-plus-cloud digital-twin context). Latency is reported in 2/31 (6.5%) studies: PS7 reports 16 ms inference for a WGAN-BiLSTM SOH estimator, whereas PS15 reports generation time (100–120 ms for 120 synthetic samples) rather than estimator latency. No study reports model footprint (parameters/MB), runtime power/energy, or portability artifacts (e.g., ONNX/TFLite). By contrast, qualitative integration statements are common (28/31), but they remain largely conceptual and do not quantify performance under hardware constraints (latency/footprint/power) required for BMS deployment.

Overall, the corpus primarily validates augmentation benefits in offline regimes rather than establishing engineering readiness for ISO~26262 (functional safety)-constrained BMS deployment. The absence of footprint, power/energy, and portability reporting—together with the lack of embedded-class deployment targets—precludes feasibility assessment against on-pack compute, memory/flash, and thermal budgets. Compression toolchains (quantization, pruning, distillation) and batch = 1 latency under fixed device/clock conditions are also not reported, although they are prerequisites for real-time integration on automotive microcontrollers and gateways.

To support meaningful deployment assessment, a minimal *deployment card* should be reported: device class and vendor/model; venue (edge/gateway/cloud); batch = 1 latency (ms) at the stated device/clock; model footprint (parameters and MB); applied compression (quantization/pruning/distillation) and toolchain; runtime power or energy/inference; and a portability artifact (e.g., ONNX or TFLite) with a commit hash/DOI. Within this corpus, quantitative deployment evidence is limited to one strong real-time datapoint (PS7) and two architecture-level integration exemplars (PS5 and PS23); the remainder provide conceptual integration without metrics. Consequently, while accuracy gains from GAN-based augmentation are demonstrated, hardware-constrained readiness for on-vehicle BMS deployment is not yet established. This gap motivates the reporting and engineering guidance in Section 6.

4.6. RQ5: What future research directions are proposed for GAN-based data augmentation in battery state estimation and prognostics?

This section synthesizes future directions for GAN-based augmentation in lithium-ion battery state estimation and prognostics, focusing on actionable priorities and evaluation practices. Analysis of 31 primary studies reveals ten distinct research directions, which we organize into five thematic categories to provide structured guidance for advancing the field. For per-study mappings and the full future directions table, see the OSF extraction table (https://osf.io/rmj8b/).

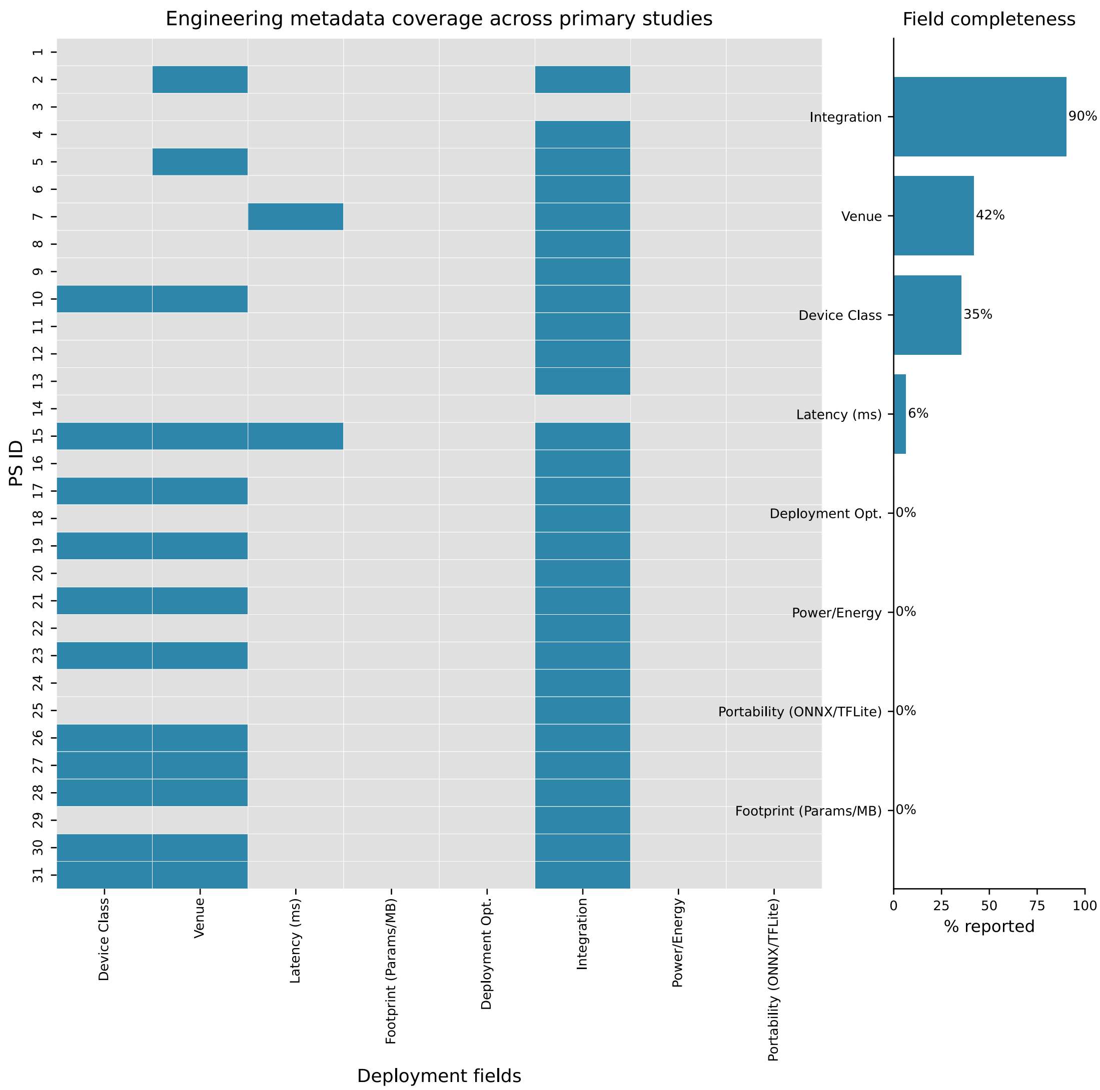

Fig. 14. Engineering metadata coverage for Battery Management System (BMS) deployment across the 31 primary studies. Each row corresponds to a primary study (PS), and each column corresponds to a deployment-relevant reporting field (e.g., device class, deployment venue, latency, model footprint, power/energy, deployment optimizations, integration placement, and portability artifacts). Blue cells indicate that the field was explicitly reported in the paper; blank/gray cells indicate not reported (NR). The right-hand panel aggregates these markings to show the percentage of studies reporting each field, highlighting which deployability-critical metrics are routinely reported versus largely missing. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.6.1. *Physics-informed and constraint-based approaches*

The most frequently identified research direction (10 studies: PS1, PS3, PS4, PS6, PS9, PS12, PS13, PS18, PS20, and PS23) emphasizes integrating physics-informed and electrochemical constraints into GAN architectures to ensure generated data adheres to fundamental battery principles. PS1 proposes physics-informed conditional GANs for realistic SOC curves with domain adaptation across chemistries, while PS3 advocates integrating GAN-based augmentation with uncertainty-aware BiLSTM models while enforcing cycle-life physics constraints in the generator. PS4 extends existing VMD-Transformer-GAN architectures by embedding physics-guided SOC consistency losses during training, and PS6 builds on CGAN-SQD by adding physics-informed electrochemical consistency losses during generation.

The rationale for this emphasis stems from the current data-driven nature of GAN approaches, which often generate statistically plausible but physically implausible battery behaviors. PS12 specifically calls for injecting physics-informed electrochemical constraints into both generator and prediction loss, while PS13 emphasizes embedding physics-informed electrochemical constraints during adversarial training. PS18 proposes injecting physics-guided voltage/current reconstruction losses so the generator learns thermodynamic consistency, not just SOC patterns. PS20 advocates embedding electro-thermal/aging constraints in the adversarial loss to ensure physically plausible synthetic sequences of voltage, current, and impedance.

These studies collectively identify critical physical constraints that must be enforced: SOC-OCV relationships, monotonic capacity fade, voltage–current feasibility, thermodynamic consistency, and electrochemical degradation patterns. The integration of such constraints requires novel loss function designs that combine adversarial objectives with physics-based penalties, potentially leveraging differentiable electrochemical models or constraint satisfaction networks.

4.6.2. Robustness and generalization through domain adaptation

Domain adaptation and transfer learning represent the second most prominent research direction (11 studies: PS1, PS4, PS6, PS9, PS10, PS12, PS20, PS21, PS25, PS27, PS31), addressing the critical challenge of generalizing GAN-based approaches across different battery chemistries, operating temperatures, and drive cycles. PS1 specifically targets domain adaptation across chemistries, while PS4 proposes adding domain-adaptation layers to generalize across temperatures and drive profiles. PS6 leverages transfer or meta-learning to adapt generators to new chemistries with few real samples, and PS9 advocates leveraging transfer learning across chemistries.

PS20 presents a comprehensive approach, extending the W-DC-GAN-GP-TL framework by scaling cross-chemistry via meta-transfer so a single generator adapts from NMC datasets to LFP, LCO, and solid-state cells with less than 5% real data. PS21 advances GAN-LSTM transfer learning by incorporating meta-learning for quick adaptation to new chemistries, while PS27 generalizes degradation-knowledge transfer from LFP to NMC, LCO, and solid-state chemistries via multi-source domain adaptation. PS31 introduces graph-based cross-battery feature alignment to model relational structure among hundreds/thousands of cells in EV fleets.

The domain shift challenge is particularly acute in battery applications due to the sensitivity of electrochemical behavior to chemistry, temperature, and usage patterns. Current studies demonstrate that GANs trained on one chemistry (e.g., NMC) often fail to generalize to others (e.g., LFP), highlighting the need for robust domain adaptation mechanisms. The proposed solutions include adversarial domain adaptation, meta-learning for few-shot adaptation, and continual learning approaches that enable online adaptation as new data becomes available.

4.6.3. Uncertainty quantification and reliability assessment

Uncertainty quantification emerges as a critical research direction (9 studies: PS3, PS8, PS12, PS13, PS18, PS21, PS23, PS27, PS30), addressing the safety-critical nature of battery management systems where prediction confidence is essential. PS3 integrates GAN-based augmentation with uncertainty-aware BiLSTM models, while PS8 incorporates uncertainty quantification in generated degradation trajectories. PS12 endows transformers with Bayesian or quantile heads for calibrated uncertainty in long-horizon RUL forecasts, and PS13 extends pipelines to knee-point detection and calibrated uncertainty estimation for safer cycle-life forecasts.

PS18 adds Bayesian or attention-based uncertainty heads to quantify confidence under real-road distribution shift, while PS21 fuses uncertainty-aware loss to improve reliability. PS23 integrates explainable AI techniques (e.g., SHapley Additive exPlanations (SHAP), attention maps) to interpret detected aging indicators and generator outputs, and PS27 expands frameworks into unified health-assessment systems that combine path-specific degradation indicators with uncertainty-aware SOH/RUL forecasts. PS30 quantifies the benefit of synthetic data on downstream SOC/RUL estimators under extreme drive and temperature scenarios.

The absence of uncertainty quantification in current GAN-generated synthetic data represents a significant limitation for safety-critical applications. Battery management systems require not only accurate predictions but also reliable confidence estimates to enable appropriate safety margins and failure prevention strategies. The proposed approaches include Bayesian neural networks, quantile regression, ensemble methods, and calibration techniques that provide well-calibrated uncertainty estimates for both synthetic data quality and downstream prediction reliability.

4.6.4. Real-time deployment and edge computing

Real-time and edge deployment represents a crucial practical research direction (12 studies: PS2, PS4, PS5, PS14, PS16, PS17, PS19, PS20, PS21, PS28, PS30, PS31), addressing the computational constraints of battery management systems in electric vehicles and portable devices. PS2 couples GAN generators with differentiable SOC estimators in end-to-end training and explores lightweight architectures and incremental learning to accelerate synthesis for practical deployment. PS4 compresses/quantizes transformer generators for real-time on-board deployment, while PS5 enables streaming/online retraining across the edge-cloud hierarchy for continual SOC prediction.

PS14 explores on-board fast EIS plus GAN denoising for rapid SOH updates, and PS16 deploys lightweight models for in-situ onboard diagnostics. PS17 investigates edge deployment for real-time SOH monitoring, while PS19 develops adaptive or automated scaling selection and robust M-estimator losses to further raise generation fidelity. PS20 enables continual learning that updates the generator-estimator pair online as new fleet data arrive without catastrophic forgetting. PS21 investigates edge deployment for real-time SOH monitoring, and PS28 enables online SOH tracking by updating the generator with incremental data.

PS30 distills models and prunes attention layers for resource-constrained edge/BMS deployment with sub-10 ms inference, while PS31 stress-tests scalability and latency on million-cycle, multi-factory datasets to validate industrial deployment. The computational complexity of GAN training and inference poses significant challenges for real-time applications, where response times must be sub-second and computational resources are limited. The proposed solutions include model compression, quantization, pruning, knowledge distillation, and specialized hardware acceleration techniques.

4.6.5. Advanced learning paradigms and standardization

Multi-modal learning, active learning, and benchmarking frameworks represent active research directions that address the complexity and diversity of battery data. Multi-modal and multi-task learning (7 studies: PS11, PS15, PS17, PS20, PS22, PS30, PS31) propose fusing voltage, current, temperature, and EIS data with cross-attention mechanisms and jointly learning SOC/SOH/RUL with ablations for modality contribution. PS11 develops multi-task cGANs to synthesize data for concurrent SOH, SOC indicators, while PS15 extends GAN-CLS image-to-signal synthesis with vision-language diffusion models and combines a bidirectional LSTM decoder with a transformer for long-range SOC prediction.

Active learning and data efficiency (4 studies: PS7, PS10, PS22, PS31) address the challenge of reducing labeling costs and improving data efficiency through query strategies and quality-based selection. PS7 integrates active-learning loops to iteratively label most informative cycles and reduce data-collection overhead, while PS10 employs multi-task learning to jointly model capacity fade and internal resistance and studies active learning loops to select informative cycles for augmentation.

Benchmarking and evaluation frameworks (9 studies: PS1, PS10, PS17, PS20, PS24, PS25, PS26, PS30, PS31) emphasize the need for standardized evaluation protocols, including standardized splits, preprocessing pipelines, fidelity checks (MMD/DTW/PCA+t-SNE), code release, and stress-tests for scalability/latency and cross-chemistry transfer. PS1 proposes open benchmarks for synthetic-data quality, while PS10 releases open benchmark workloads for small-data capacity-estimation research. PS20 benchmarks joint SOC-SOH estimation on high-power dynamic drive cycles and fast-charge scenarios, and PS31 evaluates cross-dataset transfer for condition evaluation.

Privacy-preserving and federated learning (3 studies: PS5, PS20, PS31) address the challenge of enabling fleet-scale collaboration without raw-data sharing, with explicit privacy budgets and performance gaps versus centralized baselines. PS5 incorporates federated learning for privacy-preserving fleet-wide data sharing, while PS20 and PS31 enable privacy-preserving federated transfer so estimators adapt to new vehicles without sharing raw data.

These advanced learning paradigms collectively address the need for more sophisticated, efficient, and standardized approaches to GAN-based battery state estimation, moving beyond basic data augmentation towards comprehensive frameworks that integrate multiple data modalities, optimize data efficiency, and ensure reproducible evaluation across diverse battery systems and operating conditions.

5. Conclusion

Lithium-ion battery state estimation and prognostics are safety-critical for electric vehicles, grid storage, and portable electronics. This systematic review screened 319 records and included 31 primary studies (PSs), showing that GAN-based data augmentation consistently improves SOC, SOH, and RUL tasks. Across heterogeneous datasets and models, reported error reductions range from 17%–90% (RMSE/MAE/MAPE). Large gains were achieved using temporal architectures such as TimeGAN (22.6%) and stability-oriented designs such as WGAN variants (29.0%) integrated with sequential estimators including LSTM, GRU, and Transformer models.

The evidence highlights both strengths and constraints. Research predominantly relies on public datasets—NASA (13 PSs), CALCE (6 PSs), and Oxford (5 PSs)—which enhance reproducibility but limit cross-domain generalization. Persistent challenges include data scarcity (93.5%), synthetic-data fidelity concerns (48.4%), GAN training instability (25.8%), and domain shift across chemistries, temperatures, and drive cycles (19.4%). These obstacles underline the need for stronger fidelity checks, stability mechanisms, and cross-dataset validation before industrial deployment.

Six research priorities emerge to advance reliable and interpretable battery management systems: (1) integrating physics-informed constraints, (2) developing domain adaptation methods across chemistries and conditions, (3) advancing uncertainty quantification and calibration, (4) enabling real-time deployment with efficient computation, (5) expanding multimodal learning from voltage, current, temperature, and impedance signals, and (6) improving data efficiency through active and federated learning. Addressing these gaps will bridge experimental advances with trustworthy large-scale applications.

Overall, this systematic analysis establishes GAN-based augmentation as a robust and effective strategy for battery state estimation and prognostics under data constraints. By combining fidelity validation with scalable architectures, future work can accelerate the transition towards accurate, reliable, and safe battery management systems for EVs, renewable energy storage, and portable devices.

6. Recommendations

Based on the findings of this review, we outline targeted recommendations for advancing the use of GAN-based augmentation in battery state estimation and prognostics. For researchers, the emphasis is on reproducibility, model fidelity, stability, and methods that address domain adaptation, multimodality, and data efficiency. For practitioners, the focus is on deploying robust GAN variants with sequential estimators, enforcing strict data hygiene, applying fidelity checks, and ensuring reliability under real-world conditions and constraints. Together, these recommendations aim to bridge methodological progress with trustworthy and scalable applications in battery management systems.

Key recommendations for researchers

- Standardize benchmarks and report complete protocols: fixed splits, key preprocessing choices (e.g., scaling/windowing/denoising), and leakage controls; release code and artifacts for reproducibility.
- Integrate physics-informed constraints and battery priors into GAN objectives or architectures to improve realism and identifiability.
- Develop domain adaptation methods across chemistries, temperatures, and drive cycles; evaluate cross-dataset transfer and robustness to shift.
- Quantify and report uncertainty and calibration of downstream estimators when trained with synthetic data.
- Improve training stability (WGAN-GP, spectral normalization); report random seeds, training length (30 k–50 k iterations), batch size ($\geq$64), and ablations.
- Advance multimodal and multitask learning (voltage, current, temperature, EIS) with explicit modality-contribution analyses.
- Explore active learning and federated learning to reduce labeling cost and enable privacy-preserving collaboration.

Key recommendations for practitioners

- Use temporally aware GANs (e.g., TimeGAN) or stable variants (WGAN-GP) with sequential estimators (e.g., LSTM, GRU, and Transformer) for time-series battery data.
- Enforce strict data hygiene: generate only from training data, validate on held-out sets, and prevent cross-split augmentation leakage.
- Apply quantitative fidelity checks (e.g., MMD and DTW), and use PCA + t-SNE for qualitative inspection; apply SQD before mixing.
- Start with robust defaults (Adam optimizer, learning rate $\approx 10^{-3}$, batch size ≥ 64, gradient penalty $\lambda_{GP} \approx 10$); tune per dataset with small grids.
- Tune synthetic-to-real ratios via validation and report a sensitivity analysis; avoid over-synthesis.
- Evaluate robustness under domain shift (chemistry, temperature, drive cycles) and report latency/footprint for on-vehicle or edge deployment.
- When data sharing is constrained, adopt federated or privacy-preserving setups rather than centralized pooling.

CRediT authorship contribution statement

Md. Sulyman Islam Sifat: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ummy Sawda:** Resources, Investigation, Data curation. **Md Alamgir Kabir:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Conceptualization. **Mobyen Uddin Ahmed:** Writing – review & editing, Writing – original draft. **M.S. Hossain Lipu:** Writing – review & editing. **M.M. Manjurul Islam:** Writing – review & editing.

Primary studies

The complete set of primary studies (PS1–PS31), with extraction sheets, is available on the OSF project hub: https://osf.io/yshnq/.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors used ChatGPT to support language editing during manuscript preparation. After its use, the authors critically reviewed and revised the text and accept responsibility for the content of this publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix. List of abbreviations

Abbreviation	Full form	Abbreviation	Full form
ACGAN	Auxiliary Classifier GAN	LSGAN	Least Squares Generative Adversarial Network
AE	Absolute Error	LSTM	Long Short-Term Memory
ARMSE	Average Root Mean Square Error	MAE	Mean Absolute Error
BiGRU	Bidirectional Gated Recurrent Unit	MAPE	Mean Absolute Percentage Error
BiLSTM	Bidirectional Long Short-Term Memory	MLP	Multi-Layer Perceptron
BMS	Battery Management System	MMD	Maximum Mean Discrepancy
CAGR	Compound Annual Growth Rate	MSE	Mean Squared Error
CALCE	Center for Advanced Life Cycle Engineering	NASA	National Aeronautics and Space Administration
CASP	Critical Appraisal Skills Programme	NMC	Nickel Manganese Cobalt
C-rate	Charge/discharge rate relative to capacity	NN	Neural Network
CGAN	Conditional Generative Adversarial Network	OCV	Open-Circuit Voltage
CLSTM-CGAN	Convolutional LSTM Conditional GAN	OSF	Open Science Framework
CNN	Convolutional Neural Network	PCA	Principal Component Analysis
CORAL	CORrelation ALignment	PHM	Prognostics and Health Management
CR-GAN	Convolutional Recurrent GAN	PICOS	Population, Intervention, Comparator, Outcomes, Study design
CSV	Comma-Separated Values	PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
CTSGAN	Conditional Time-Series GAN	PRISMA-S	PRISMA extension for Searching
CycleGAN	Cycle-Consistent Generative Adversarial Network	PS	Primary Study identifier (e.g., PS1–PS31)
DCGAN	Deep Convolutional GAN	QA	Quality Assurance
DFN	Doyle–Fuller–Newman model	R^2	Coefficient of Determination
DRAGAN	Deep Regret Analytic GAN	RC	Resistor–Capacitor (ladder) model
DTW	Dynamic Time Warping	RGAN	Regression Generative Adversarial Network
DA	Domain Adaptation	soft-DTW	soft Dynamic Time Warping
ECM	Equivalent Circuit Model	RIS	Research Information Systems (citation export format)
EIS	Electrochemical Impedance Spectroscopy	RNN	Recurrent Neural Network
EKF	Extended Kalman Filter	RMSE	Root Mean Square Error
RQ	Research Question	UQ	Uncertainty Quantification
EMD	Earth Mover’s Distance	RUL	Remaining Useful Life
ESS	Effective Sample Size	SHAP	SHapley Additive exPlanations
EV	Electric Vehicle	SLR	Systematic Literature Review
FC	Fully Connected	SMC	Sequential Monte Carlo
FID	Fréchet Inception Distance	SMOTE	Synthetic Minority Over-sampling Technique
GAN	Generative Adversarial Network	SOC	State of Charge
GDA	Gaussian Data Augmentation	SOH	State of Health
GP	Gradient Penalty	SQD	Sample Quality Determination
GPR	Gaussian Process Regression	SSIM	Structural Similarity Index Measure
GRU	Gated Recurrent Unit	t-SNE	t-Distributed Stochastic Neighbor Embedding
IEEE	Institute of Electrical and Electronics Engineers	TCNN	Temporal Convolutional Neural Network
InfoGAN	Information Maximizing GAN	TFT	Temporal Fusion Transformer
IQR	Interquartile Range	TimeGAN	Time-series Generative Adversarial Network
IS	Inception Score	TL	Transfer Learning
JSD	Jensen–Shannon Divergence	Transformer	Transformer Neural Network
KL	Kullback–Leibler	VAE	Variational Autoencoder
LCO	Lithium Cobalt Oxide	VAE-GAN	Variational Autoencoder–Generative Adversarial Network
LFP	Lithium Iron Phosphate	VMD	Variational Mode Decomposition
Li-ion	Lithium-ion	WGAN	Wasserstein Generative Adversarial Network
LR	Learning Rate	WGAN-GP	Wasserstein GAN with Gradient Penalty
CC	Constant-Current	CCCV	Constant-Current Constant-Voltage
IC	Incremental Capacity	PCoE	Prognostics Center of Excellence
DSP	Digital Signal Processor	FPGA	Field-Programmable Gate Array
MCU	Microcontroller Unit	ONNX	Open Neural Network Exchange
TFLite	TensorFlow Lite	ISO 26262	Functional safety standard for road vehicles

Data availability

Data will be made available on request.

References

[1] W. Chen, J. Liang, Z. Yang, G. Li, A review of lithium-ion battery for electric vehicle applications and beyond, Energy Procedia 158 (2019) 4363–4368, http://dx.doi.org/10.1016/j.egypro.2019.01.783, URL https://www.sciencedirect.com/science/article/pii/S1876610219308215, Innovative Solutions for Energy Transitions.

[2] K.A. Smith, Electrochemical control of lithium-ion batteries [applications of control], IEEE Control Syst. Mag. 30 (2) (2010) 18–25, http://dx.doi.org/10.1109/MCS.2010.935882.

[3] L. Barote, C. Marinescu, Reserved memory for Li-ion battery SOC estimation method in applications with EV, in: 2021 International Aegean Conference on Electrical Machines and Power Electronics (ACEMP) & 2021 International Conference on Optimization of Electrical and Electronic Equipment, OPTIM, 2021, pp. 22–27, http://dx.doi.org/10.1109/OPTIM-ACEMP50812.2021.9590063.

[4] E. Kim, M. Kim, J. Kim, J. Kim, J.-H. Park, K.-T. Kim, J.-H. Park, T. Kim, K. Min, Data-driven methods for predicting the state of health, state of charge, and remaining useful life of Li-ion batteries: A comprehensive review, Int. J. Precis. Eng. Manuf. 24 (2023) 1281–1304, URL https://api.semanticscholar.org/CorpusID:258645252.

[5] J. Zhao, Y. Zhu, B. Zhang, M. Liu, J. Wang, C. Liu, X. Hao, Review of state estimation and remaining useful life prediction methods for lithium–ion batteries, Sustainability 15 (2023) 5014, http://dx.doi.org/10.3390/su15065014.

[6] U. Saleem, W. Li, W. Liu, I. Ahmad, M.M. Aslam, H.U. Lateef, Investigation of deep learning based techniques for prognostic and health management of lithium-ion battery, in: 2023 15th International Conference on Electronics, Computers and Artificial Intelligence, ECAI, 2023, pp. 01–06, http://dx.doi.org/10.1109/ECAI58194.2023.10194122.

[7] T. Pattnaik, M. Kumar, M.S. Ballal, Predicting SoH and RUL of Li-ion batteries in electric vehicles: A comprehensive review and chemistry insights, in: 2023 IEEE International Conference on Power Electronics, Smart Grid, and Renewable Energy, PESGRE, 2023, pp. 1–6, URL https://api.semanticscholar.org/CorpusID:267340012.

[8] W. Gomez, F.-K. Wang, J.-H. Chou, Li-ion battery capacity prediction using improved temporal fusion transformer model, Energy 296 (2024) 131114, http://dx.doi.org/10.1016/j.energy.2024.131114, URL https://www.sciencedirect.com/science/article/pii/S0360544224008867.

[9] C. Tripp-Barba, J.A. Aguilar-Calderón, L. Urquiza-Aguiar, A. Zaldívar-Colado, A. Ramírez-Noriega, A systematic mapping study on state estimation techniques for lithium-ion batteries in electric vehicles, World Electr. Veh. J. 16 (2) (2025) http://dx.doi.org/10.3390/wevj16020057, URL https://www.mdpi.com/2032-6653/16/2/57.

[10] B. Mullapudi, S. R, Unified approach for estimation of state of charge and remaining useful life of Li-ion batteries using deep learning, in: 2022 IEEE 19th India Council International Conference, INDICON, 2022, pp. 1–6, http://dx.doi.org/10.1109/INDICON56171.2022.10039759.

[11] D. Li, X. Liu, H. Zhou, Z. Cheng, A hybrid model based joint estimation of multi-state parameters for lithium-ion batteries, in: 2023 42nd Chinese Control Conference, CCC, 2023, pp. 6304–6308, URL https://api.semanticscholar.org/CorpusID:262063873.

[12] S. Gupta, P.K. Mishra, Estimation of SoC, SoH and RUL of Li-ion battery: A review, in: 2023 5th International Conference on Energy, Power and Environment: Towards Flexible Green Energy Technologies, ICEPE, 2023, pp. 1–6, URL https://api.semanticscholar.org/CorpusID:260741174.

[13] P. Dini, A. Colicelli, S. Saponara, Review on modeling and SOC/SOH estimation of batteries for automotive applications, Batteries 10 (1) (2024) http://dx.doi.org/10.3390/batteries10010034, URL https://www.mdpi.com/2313-0105/10/1/34.

[14] S. Saxena, Y. Xing, M.G. Pecht, PHM of Li-ion batteries, in: Prognostics and Health Management of Electronics, John Wiley & Sons, Ltd, 2018, pp. 349–375, http://dx.doi.org/10.1002/9781119515326.ch13, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119515326.ch13. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119515326.ch13.

[15] G. Shinde, R. Mohapatra, P. Krishan, H. Garg, S. Prabhu, S. Das, M. Masum, S. Sengupta, The state of lithium-ion battery health prognostics in the CPS era, 2024, arXiv, abs/2403.19816, URL https://api.semanticscholar.org/CorpusID:268793730.

[16] K. Das, R. Kumar, Electric vehicle battery capacity degradation and health estimation using machine-learning techniques: A review, Clean Energy 7 (6) (2023) 1268–1281, http://dx.doi.org/10.1093/ce/zkad054, arXiv:https://academic.oup.com/ce/article-pdf/7/6/1268/53588600/zkad054.pdf.

[17] S.S. Madani, C. Ziebert, P. Vahdatkhah, S.K. Sadrnezhaad, Recent progress of deep learning methods for health monitoring of lithium-ion batteries, Batteries 10 (6) (2024) http://dx.doi.org/10.3390/batteries10060204, URL https://www.mdpi.com/2313-0105/10/6/204.

[18] M.M. Islam, A.E. Prosvirin, J.-M. Kim, Data-driven prognostic scheme for rolling-element bearings using a new health index and variants of least-square support vector machines, Mech. Syst. Signal Process. 160 (2021) 107853.

[19] R. Mehta, A. Gupta, Mathematical modelling of electrochemical, thermal and degradation processes in lithium-ion cells—A comprehensive review, Renew. Sustain. Energy Rev. 192 (2024) 114264, http://dx.doi.org/10.1016/j.rser.2023.114264, URL https://www.sciencedirect.com/science/article/pii/S136403212301122X.

[20] K.S. Hariharan, V. Senthil Kumar, A nonlinear equivalent circuit model for lithium ion cells, J. Power Sources 222 (2013) 210–217, http://dx.doi.org/10.1016/j.jpowsour.2012.08.090, URL https://www.sciencedirect.com/science/article/pii/S0378775312013973.

[21] P. Rodríguez-Iturriaga, D. Anseán, S. Rodríguez-Bolívar, V.M. García, M. González, J.A. López-Villanueva, Modeling current-rate effects in lithium-ion batteries based on a distributed, multi-particle equivalent circuit model, Appl. Energy 353 (2024) 122141, http://dx.doi.org/10.1016/j.apenergy.2023.122141, URL https://www.sciencedirect.com/science/article/pii/S0306261923015052.

[22] M. Doyle, T.F. Fuller, J. Newman, Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell, J. Electrochem. Soc. 140 (6) (1993) 1526, http://dx.doi.org/10.1149/1.2221597.

[23] Y. Peng, H. Hu, A. Geng, Y. Chen, Z. Zhao, An electrochemical impedance spectroscopy-based model for lithium-ion batteries with solid phase diffusion process, in: 2023 IEEE 2nd International Power Electronics and Application Symposium, PEAS, 2023, pp. 887–891, URL https://api.semanticscholar.org/CorpusID:267260746.

[24] C. Fan, T. Zhao, W.D. Widanage, A comparative study of equivalent circuit models with diverse nonlinear dynamics for Li-ion batteries, in: IECON 2023-49th Annual Conference of the IEEE Industrial Electronics Society, 2023, pp. 1–6, http://dx.doi.org/10.1109/IECON51785.2023.10312148.

[25] H. Pan, X. Wang, L. Zhang, R. Wang, H. Dai, X. Wei, Online broadband impedance identification for lithium-ion batteries based on a nonlinear equivalent circuit model, World Electr. Veh. J. 14 (7) (2023) http://dx.doi.org/10.3390/wevj14070168, URL https://www.mdpi.com/2032-6653/14/7/168.

[26] A. Graule, F. Oehler, J. Schmitt, J. Li, A. Jossen, Development and evaluation of a physicochemical equivalent circuit model for lithium-ion batteries, J. Electrochem. Soc. 171 (2) (2024) 020503, http://dx.doi.org/10.1149/1945-7111/ad1ec7.

[27] B. Sun, J. Liu, H. Ruan, W. Zhang, H. Li, J. Wang, Study on fractional order modeling and equivalent stress of AC–dc superposition condition for lithium-ion batteries, in: W. Chen, Q. Yang, L. Wang, D. Liu, X. Han, G. Meng (Eds.), The Proceedings of the 9th Frontier Academic Forum of Electrical Engineering, Springer Singapore, Singapore, 2021, pp. 115–127.

[28] W. Xu, K. Yang, Y. Zhang, S. Sun, S. Mao, T. Xue, DiffLiB: High-fidelity differentiable modeling of lithium-ion batteries and efficient gradient-based parameter identification, 2025, arXiv:2504.20674, URL https://arxiv.org/abs/2504.20674.

[29] D. Gao, X. Liu, Z. Zhu, Q. Yang, A hybrid CNN-BiLSTM approach for remaining useful life prediction of EVs lithium-ion battery, Meas. Control. 56 (1–2) (2023) 371–383, http://dx.doi.org/10.1177/00202940221103622, arXiv:https://doi.org/10.1177/00202940221103622.

[30] B. Zhou, C. Cheng, G. Ma, Y. Zhang, Remaining useful life prediction of lithium-ion battery based on attention mechanism with positional encoding, IOP Conf. Ser.: Mater. Sci. Eng. 895 (1) (2020) 012006, http://dx.doi.org/10.1088/1757-899X/895/1/012006.

[31] A.H. Shakil, M.A. Orin, M. Rahman, M.A. Kabir, AttNet: An attention-based BiGRU network for remaining useful life prediction, 2024, http://dx.doi.org/10.1109/STI64222.2024.10951153, URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-105003430669&doi=10.1109%2fSTI64222.2024.10951153&partnerID=40&md5=2a179e91c946f33ec748c4388d0feaa9, Scopus; Cited by: 0, 2024 6th International Conference on Sustainable Technologies for Industry 5.0 (STI 2024).

[32] S.T. Aggani, S. Mishra, Deep-learning-based SoC estimation for $\mathbf{LiFePO_4}$ batteries using CNN-GRU-TCN hybrid architecture, in: 2025 Emerging Technologies for Intelligent Systems, ETIS, 2025, pp. 1–7, http://dx.doi.org/10.1109/ETIS64005.2025.10961441.

[33] M.Z. Aqil, Q.A. Farahmand, A.H. Shujaie, J. Chen, Application of deep learning for prediction of remaining useful life of Li-ion batteries, in: J. Wu, A. Ma'aram (Eds.), Ninth International Conference on Electromechanical Control Technology and Transportation, ICECTT 2024, Vol. 13251, SPIE, International Society for Optics and Photonics, 2024, p. 132511B, http://dx.doi.org/10.1117/12.3040336.

[34] B. Wu, S. Qiu, W. Liu, Addressing sensor data heterogeneity and sample imbalance: A transformer-based approach for battery degradation prediction in electric vehicles, Sensors 25 (11) (2025) http://dx.doi.org/10.3390/s25113564, URL https://www.mdpi.com/1424-8220/25/11/3564.

[35] M.R. Fauzi, N. Yudistira, W.F. Mahmudy, State-of-health prediction of lithium-ion batteries using exponential smoothing transformer with seasonal and growth embedding, IEEE Access 12 (2024) 14659–14670, http://dx.doi.org/10.1109/ACCESS.2024.3357736.

[36] C. Hu, F. Cheng, Y. Zhao, S. Guo, L. Ma, State of charge estimation for lithium-ion batteries based on data augmentation with generative adversarial network, J. Energy Storage 80 (2024) 110004, http://dx.doi.org/10.1016/j.est.2023.110004, URL https://www.sciencedirect.com/science/article/pii/S2352152X23034035.

[37] J. Liu, X. Wang, Y. Zhang, W. Chen, Time-series wasserstein GAN for battery state estimation: A novel approach to synthetic data generation, J. Energy Storage 72 (2023) 108456, http://dx.doi.org/10.1016/j.est.2023.108456.

[38] L. Zhang, M. Li, H. Zhao, P. Wang, Deep convolutional generative adversarial networks for lithium-ion battery degradation trajectory synthesis, Appl. Energy 355 (2024) 122234, http://dx.doi.org/10.1016/j.apenergy.2023.122234.

[39] X. Gu, K. See, Y. Liu, B. Arshad, L. Zhao, Y. Wang, A time-series wasserstein GAN method for state-of-charge estimation of lithium-ion batteries, J. Power Sources 581 (2023) 233472, http://dx.doi.org/10.1016/j.jpowsour.2023.233472, URL https://www.sciencedirect.com/science/article/pii/S0378775323008480.

[40] G. Yang, Q. Ma, H. Sun, X. Zhang, State of health estimation based on GAN-LSTM-TL for lithium-ion batteries, Int. J. Electrochem. Sci. 17 (11) (2022) 221128, http://dx.doi.org/10.20964/2022.11.07, URL https://www.sciencedirect.com/science/article/pii/S145239812302655X.

[41] H. Zhang, W. Tang, W. Na, P. Lee, J. Kim, Implementation of generative adversarial network-CLS combined with bidirectional long short-term memory for lithium-ion battery state prediction, J. Energy Storage 31 (2020) 101489, http://dx.doi.org/10.1016/j.est.2020.101489.

[42] C. Wang, S. Jiao, P. Wu, S. Kou, X. Wang, SOC estimation of Li-ion battery using variational mode decomposition and transformer–generative adversarial network, 2023, http://dx.doi.org/10.1051/e3sconf/202346601006, E3S Web of Conferences.

[43] M.S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, IEEE Trans. Signal Process. 50 (2) (2002) 174–188, http://dx.doi.org/10.1109/78.978374.

[44] A. Doucet, N. de Freitas, N. Gordon (Eds.), Sequential Monte Carlo Methods in Practice, Springer, New York, 2001, http://dx.doi.org/10.1007/978-1-4757-3437-9.

[45] Y. Ma, Y. Chen, X. Zhou, H. Chen, Remaining useful life prediction of lithium-ion battery based on Gauss–Hermite particle filter, IEEE Trans. Control Syst. Technol. 27 (4) (2019) 1788–1795, http://dx.doi.org/10.1109/TCST.2018.2819965.

[46] K.A. Severson, P.M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M.H. Chen, M. Aykol, P.K. Herring, D. Fraggedakis, M.Z. Bazant, S.J. Harris, W.C. Chueh, R.D. Braatz, Data-driven prediction of battery cycle life before capacity degradation, Nat. Energy 4 (2019) 383–391, http://dx.doi.org/10.1038/s41560-019-0356-8.

[47] M. Berecibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, P. Van den Bossche, Critical review of state of health estimation methods of Li-ion batteries for real applications, Renew. Sustain. Energy Rev. 56 (2016) 572–587, http://dx.doi.org/10.1016/j.rser.2015.11.042.

[48] V.M. Nagulapati, H. Lee, D. Jung, B. Brigljević, Y. Choi, H. Lim, Capacity estimation of batteries: Influence of training dataset size and diversity on data driven prognostic models, Reliab. Eng. Syst. Saf. 216 (2021) 108048, http://dx.doi.org/10.1016/j.ress.2021.108048.

[49] X. Cui, T. Hu, State of health diagnosis and remaining useful life prediction for lithium-ion battery based on data model fusion method, IEEE Access 8 (2020) 207298–207307, http://dx.doi.org/10.1109/ACCESS.2020.3038182.

[50] D. Liu, Y. Luo, J. Liu, Y. Peng, L. Guo, M.G. Pecht, Lithium-ion battery remaining useful life estimation based on fusion nonlinear degradation AR model and RPF algorithm, Neural Comput. Appl. 25 (2013) 557–572, URL https://api.semanticscholar.org/CorpusID:18076906.

[51] A. Bracale, P. de Falco, L.P.D. Noia, R. Rizzo, Probabilistic state of health and remaining useful life prediction for Li-ion batteries, in: 2021 IEEE Texas Power and Energy Conference, TPEC, 2021, pp. 1–6, http://dx.doi.org/10.1109/TPEC51183.2021.9384939.

[52] W. Yao, R. Lai, Y. Tian, X. Li, J. Tian, State of health estimation of lithium-ion batteries using data augmentation and feature mapping, IEEE Trans. Transp. Electrification 11 (2025) 4895–4905, http://dx.doi.org/10.1109/TTE.2024.3471867.

[53] H. Zhang, M. Cisse, Y.N. Dauphin, D. Lopez-Paz, Mixup: Beyond empirical risk minimization, 2018, arXiv:1710.09412, URL https://arxiv.org/abs/1710.09412.

[54] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: Synthetic minority over-sampling technique, J. Artificial Intelligence Res. 16 (2002) 321–357, http://dx.doi.org/10.1613/jair.953.

[55] H.I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, P.-A. Muller, Data augmentation using synthetic data for time series classification with deep residual networks, 2018, URL https://arxiv.org/abs/1808.02455, arXiv preprint arXiv:1808.02455.

[56] C. Zhang, Y. Zhang, Z. Li, Z. Zhang, M.S. Nazir, T. Peng, Enhancing state of charge and state of energy estimation in lithium-ion batteries based on a TimesNet model with Gaussian data augmentation and error correction, Appl. Energy 359 (2024) 122669, http://dx.doi.org/10.1016/j.apenergy.2024.122669, URL https://www.sciencedirect.com/science/article/pii/S0306261924000527.

[57] T. Shi, H. Pei, Research on data enhancement analysis of battery management system based on generative adversarial network, in: 2024 4th Asia-Pacific Conference on Communications Technology and Computer Science, ACCTCS, 2024, pp. 464–468, http://dx.doi.org/10.1109/ACCTCS61748.2024.00087.

[58] J. Sun, A. Gu, J. Kainz, A solution framework for the experimental data shortage problem of lithium-ion batteries: Generative adversarial network-based data augmentation for battery state estimation, J. Energy Chem. 103 (2025) 476–497, http://dx.doi.org/10.1016/j.jechem.2024.12.010, URL https://www.sciencedirect.com/science/article/pii/S2095495624008623.

[59] F. Hu, C. Dong, M. Wang, Q. Xiao, Y. Mu, H. Jia, WGAN-GP with residual network model for lithium battery thermal image data expansion with quantitative metrics, in: Proc. IEEE Int. Electrical and Energy Conf., CIEEC, 2023, pp. 4030–4035, http://dx.doi.org/10.1109/CIEEC58067.2023.10166002.

[60] Y. Liu, Q. Zhang, R. Wang, Conditional WGAN-GP-based renewable energy scenario generation method, in: Proc. Int. Conf. Smart Grid and Green Energy, ICSGGE, 2025, pp. 91–95, http://dx.doi.org/10.1109/ICSGGE64667.2025.10982692.

[61] K. Cheng, H. Liu, J. Liu, B. Xu, X. Liu, Frequency-auxiliary one-shot domain adaptation of generative adversarial networks, Electronics 13 (13) (2024) http://dx.doi.org/10.3390/electronics13132643, URL https://www.mdpi.com/2079-9292/13/13/2643.

[62] J. Tao, S. Wang, W. Cao, P. Takyi-Aninakwa, C. Fernandez, J.M. Guerrero, A comprehensive review of state-of-charge and state-of-health estimation for lithium-ion battery energy storage systems, Ionics 30 (2024) 5903–5927, http://dx.doi.org/10.1007/s11581-024-05686-z.

[63] X. Lei, F. Xie, J. Wang, C. Zhang, A review of lithium-ion battery state of health and remaining useful life estimation methods based on bibliometric analysis, J. Traffic Transp. Eng. (English Edition) 11 (6) (2024) 1420–1446, http://dx.doi.org/10.1016/j.jtte.2024.09.004.

[64] Y. Liu, Y. He, H. Bian, W. Guo, X. Zhang, A review of lithium-ion battery state of charge estimation based on deep learning: Directions for improvement and future trends, J. Energy Storage 52 (2022) 104664, http://dx.doi.org/10.1016/j.est.2022.104664.

[65] A.G. Li, A.C. West, M. Preindl, Towards unified machine learning characterization of lithium-ion battery degradation across multiple levels: A critical review, Appl. Energy 316 (2022) 119030, http://dx.doi.org/10.1016/j.apenergy.2022.119030.

[66] Y. Zhang, Y.-F. Li, Prognostics and health management of lithium-ion battery using deep learning methods: A review, Renew. Sustain. Energy Rev. 161 (2022) 112282, http://dx.doi.org/10.1016/j.rser.2022.112282.

[67] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in Neural Information Processing Systems, NeurIPS, 2014, pp. 2672–2680.

[68] S. Nowozin, B. Cseke, R. Tomioka, F-GAN: Training generative neural samplers using variational divergence minimization, in: Advances in Neural Information Processing Systems, NeurIPS, 2016, pp. 271–279.

[69] L. Mescheder, A. Geiger, S. Nowozin, Which training methods for GANs do actually converge? in: Proceedings of the 35th International Conference on Machine Learning, ICML, 2018, pp. 3481–3490.

[70] X. Mao, Q. Li, H. Xie, R.Y. Lau, Z. Wang, S. Paul Smolley, Least squares generative adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, ICCV, 2017, pp. 2794–2802.

[71] A. Odena, C. Olah, J. Shlens, Conditional image synthesis with auxiliary classifier GANs, 2017, arXiv:1610.09585, URL https://arxiv.org/abs/1610.09585.

[72] T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida, Spectral normalization for generative adversarial networks, 2018, arXiv:1802.05957, URL https://arxiv.org/abs/1802.05957.

[73] M. Mirza, S. Osindero, Conditional generative adversarial nets, 2014, arXiv:1411.1784, URL https://arxiv.org/abs/1411.1784.

[74] M. Pooyandeh, I. Sohn, Smart lithium-ion battery monitoring in electric vehicles: An AI-empowered digital twin approach, Mathematics 11 (23) (2023) http://dx.doi.org/10.3390/math11234865, URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85178926549&doi=10.3390%2fmath11234865&partnerID=40&md5=e1effca812b487c3753e9cd308e0cd5d, Cited by: 36; All Open Access, Gold Open Access.

[75] X. Zhang, Y. Qin, C. Yuen, L. Jayasinghe, X. Liu, Time-series regeneration with convolutional recurrent generative adversarial network for remaining useful life estimation, IEEE Trans. Ind. Informatics 17 (2021) 6820–6831, http://dx.doi.org/10.1109/TII.2020.3046036.

[76] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein GAN, in: International Conference on Machine Learning, ICML, 2017, pp. 214–223.

[77] C. Villani, Optimal Transport: Old and New, Springer, 2009.

[78] G. Peyré, M. Cuturi, Computational optimal transport, 2020, arXiv:1803.00567, URL https://arxiv.org/abs/1803.00567.

[79] J. Yang, T. Shu, L. Zhang, Improving noised gradient penalty with synchronized activation function for generative adversarial networks, IEICE Trans. Inf. Syst. E105.D (2022) 1537–1545, http://dx.doi.org/10.1587/transinf.2022EDP7019.

[80] J. Chang, F. Hu, H. Xu, X. Mao, Y. Zhao, L. Huang, Towards generating realistic wrist pulse signals using enhanced one-dimensional wasserstein GAN, Sensors 23 (3) (2023) 1450, http://dx.doi.org/10.3390/s23031450.

[81] J. Yoon, D. Jarrett, M. van der Schaar, Time-series generative adversarial networks, 2019, Advances in Neural Information Processing Systems (NeurIPS).

[82] C. Esteban, S.L. Hyland, G. Rätsch, Real-valued (medical) time series generation with recurrent conditional GANs, 2017, arXiv:1706.02633, URL https://arxiv.org/abs/1706.02633.

[83] O. Mogren, C-RNN-GAN: Continuous recurrent neural networks with adversarial training, 2016, arXiv:1611.09904, URL https://arxiv.org/abs/1611.09904.

[84] H.I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, P.-A. Muller, Deep learning for time series classification: A review, Data Min. Knowl. Discov. 33 (2019) 917–963, http://dx.doi.org/10.1007/s10618-019-00619-1.

[85] B. Lim, S. Zohren, Time-series forecasting with deep learning: A survey, Philos. Trans. R. Soc. A 379 (2194) (2021) 20200209, http://dx.doi.org/10.1098/rsta.2020.0209.

[86] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.

[87] F.A. Gers, J. Schmidhuber, F. Cummins, Learning to forget: Continual prediction with LSTM, Neural Comput. 12 (10) (2000) 2451–2471, http://dx.doi.org/10.1162/089976600300015015, URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-0034293152&doi=10.1162%2f089976600300015015&partnerID=40&md5=202d5c0234434c0dbc4a038fe3a5bc79, Cited by: 4671.

[88] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016, URL http://www.deeplearningbook.org.

[89] P. Malhotra, V. TV, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, G. Shroff, Multi-sensor prognostics using an unsupervised health index based on LSTM encoder-decoder, 2016, arXiv:1608.06154, URL https://arxiv.org/abs/1608.06154.

[90] K. Park, Y. Choi, W.J. Choi, H.-Y. Ryu, H. Kim, LSTM-based battery remaining useful life prediction with multi-channel charging profiles, IEEE Access 8 (2020) 20786–20798, http://dx.doi.org/10.1109/ACCESS.2020.2968939.

[91] M. Schuster, K. Paliwal, Bidirectional recurrent neural networks, IEEE Trans. Signal Process. 45 (11) (1997) 2673–2681, http://dx.doi.org/10.1109/78.650093.

[92] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, 2023, arXiv:1706.03762, URL https://arxiv.org/abs/1706.03762.

[93] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (1) (2019) 60, http://dx.doi.org/10.1186/s40537-019-0197-0.

[94] Q. Wen, L. Sun, F. Yang, X. Song, J. Gao, X. Wang, H. Xu, Time series data augmentation for deep learning: A survey, in: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, International Joint Conferences on Artificial Intelligence Organization, 2021, pp. 4653–4660, http://dx.doi.org/10.24963/ijcai.2021/631.

[95] H. He, E.A. Garcia, Learning from imbalanced data, IEEE Trans. Knowl. Data Eng. 21 (9) (2009) 1263–1284, http://dx.doi.org/10.1109/TKDE.2008.239.

[96] M. Buda, A. Maki, M.A. Mazurowski, A systematic study of the class imbalance problem in convolutional neural networks, Neural Netw. 106 (2018) 249–259, http://dx.doi.org/10.1016/j.neunet.2018.07.011.

[97] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, G. Klambauer, S. Hochreiter, GANs trained by a two time-scale update rule converge to a local Nash equilibrium, in: Advances in Neural Information Processing Systems, NeurIPS, 2017, pp. 6626–6637.

[98] M.S.M. Sajjadi, O. Bachem, M. Lucic, O. Bousquet, S. Gelly, Assessing generative models via precision and recall, 2018, arXiv:1806.00035, URL https://arxiv.org/abs/1806.00035.

[99] T. Kynkäänniemi, T. Karras, S. Laine, J. Lehtinen, T. Aila, Improved precision and recall metric for assessing generative models, 2019, arXiv:1904.06991, URL https://arxiv.org/abs/1904.06991.

[100] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training GANs, in: Advances in Neural Information Processing Systems, NeurIPS, 2016, pp. 2234–2242.

[101] H. Eivazi, A. Hebenbrock, R. Ginster, S. Blömeke, S.H.A. Wittek, C. Hermann, T.S. Spengler, T. Turek, A. Rausch, DiffBatt: A diffusion model for battery degradation prediction and synthesis, 2024, http://dx.doi.org/10.48550/arXiv.2410.23893, arXiv preprint arXiv:2410.23893.

[102] S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Stat. 22 (1) (1951) 79–86.

[103] J. Lin, Divergence measures based on the Shannon entropy, IEEE Trans. Inform. Theory 37 (1) (1991) 145–151.

[104] I. Jolliffe, Principal Component Analysis, Springer, 2002.

[105] L. Van Der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2625, URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-57249084011&partnerID=40&md5=234c45663c3957ae2268abc330a03438, Cited by: 43819.

[106] W. Zhang, J. Hu, B. Lin, D. Liu, M. Wang, D. Mu, Y. Lu, Cross-domain state-of-health estimation of Li-ion batteries based on transfer neural network with soft-dynamic time warping, Energy Sci. Eng. 11 (2023) 3137–3148, http://dx.doi.org/10.1002/ese3.1509.

[107] A.B. Owen, Monte Carlo Theory, Methods and Examples, Stanford University, 2013, URL https://statweb.stanford.edu/~owen/mc/.

[108] K. Roth, A. Lucchi, S. Nowozin, T. Hofmann, Stabilizing training of generative adversarial networks through regularization, 2017, arXiv:1705.09367, URL https://arxiv.org/abs/1705.09367.

[109] N. Kodali, J. Abernethy, J. Hays, Z. Kira, On convergence and stability of GANs, 2017, arXiv:1705.07215, URL https://arxiv.org/abs/1705.07215.

[110] A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Sch"olkopf, A. Smola, A kernel two-sample test, J. Mach. Learn. Res. 13 (2012) 723–773.

[111] K. Simonyan, A. Vedaldi, A. Zisserman, Deep inside convolutional networks: Visualising image classification models and saliency maps, 2013, arXiv:1312.6034, URL https://arxiv.org/abs/1312.6034.

[112] R. Xiong, L. Li, J. Tian, Towards a smarter battery management system: A critical review on battery state of health monitoring methods, J. Power Sources 405 (2018) 18–29, http://dx.doi.org/10.1016/j.jpowsour.2018.10.019.

[113] X. Hu, F. Feng, K. Liu, L. Zhang, J. Xie, B. Liu, State estimation for advanced battery management: Key challenges and future trends, Renew. Sustain. Energy Rev. 114 (2019) 109334, http://dx.doi.org/10.1016/j.rser.2019.109334.

[114] J. Kong, J. Liu, J. Zhu, X. Zhang, K.-L. Tsui, Z. Peng, D. Wang, Review on lithium-ion battery PHM from the perspective of key PHM steps, Chin. J. Mech. Eng. 37 (2024) 71, http://dx.doi.org/10.1186/s10033-024-01055-z.

[115] J. Zhao, X. Han, M. Ouyang, A.F. Burke, Specialized deep neural networks for battery health prognostics: Opportunities and challenges, J. Energy Chem. 87 (2023) 416–438, http://dx.doi.org/10.1016/j.jechem.2023.08.047.

[116] A. Shah, K. Shah, C. Shah, M. Shah, 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. 42 (2022) 146–164, http://dx.doi.org/10.1016/j.ref.2022.06.001.

[117] E.D. Foster, A. Deardorff, Open science framework (OSF), J. Med. Libr. Assoc. : JMLA 105 (2017) 203–206, URL https://api.semanticscholar.org/CorpusID:33681693.

[118] N.P.C. of Excellence, NASA prognostics data repository: Battery data sets, 2025, URL https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/, (Accessed 30 August 2025).

[119] C. for Advanced Life Cycle Engineering (CALCE), CALCE battery data repository, 2025, URL https://calce.umd.edu/battery-data, (Accessed 30 August 2025), CALCE, University of Maryland.

[120] U. of Oxford, Oxford battery degradation/aging dataset, 2017, URL https://ora.ox.ac.uk/objects/uuid:03ba4b01-cfed-46d3-9b1a-7d4a7bdf6fac, (Accessed 30 August 2025).

[121] P. Kollmeyer, C. Vidal, M. Naguib, M. Skells, LG 18650HG2 Li-ion battery data and example deep neural network xEV SOC estimator script, 2020, http://dx.doi.org/10.17632/cp3473x7xv.3, URL https://data.mendeley.com/datasets/cp3473x7xv/3, Mendeley Data, V3.

[122] Automotive Li-ion cell usage data set, 2018, http://dx.doi.org/10.21227/ce9q-jr19, IEEE DataPort.

[123] PowerStream, PowerStream LiR2032 coin cell battery dataset, 2025, URL https://www.powerstream.com/licoin.htm, (Accessed 30 August 2025), PowerStream.

[124] S.N. Laboratories, Sandia national laboratories LFP battery dataset, 2025, URL https://www.sandia.gov/energystoragesafety/rd-data-repository/, (Accessed 30 August 2025), Sandia National Laboratories.