batteries

MDPI

Review

Advanced Machine Learning and Deep Learning Approaches for Estimating the Remaining Life of EV Batteries—A Review

Daniel H. de la Iglesia [1,*], Carlos Chinchilla Corbacho [1,2], Jorge Zakour Dib [1,2], Vidal Alonso-Secades [2] and Alfonso J. López Rivero [2]

[1] Expert Systems and Applications Lab—ESALAB, Faculty of Science, University of Salamanca, Plaza de los Caídos s/n, 37008 Salamanca, Spain; cchinchilla@usal.es (C.C.C.); jzakourdi@upsa.es (J.Z.D.)

[2] Computer Science Faculty, Universidad Pontificia de Salamanca, Salamanca, 37002 Salamanca, Spain; valonsose@upsa.es (V.A.-S.); ajlopezri@upsa.es (A.J.L.R.)

* Correspondence: danihiglesias@usal.es

Abstract: This systematic review presents a critical analysis of advanced machine learning (ML) and deep learning (DL) approaches for predicting the remaining useful life (RUL) of electric vehicle (EV) batteries. Conducted in accordance with PRISMA guidelines and using a novel adaptation of the Downs and Black (D&B) scale, this study evaluates 89 research papers and provides insights into the evolving landscape of RUL estimation. Our analysis reveals an evolving landscape of methodological approaches, with different techniques showing distinct capabilities in capturing complex degradation patterns in EV batteries. While recent years have seen increased adoption of DL methods, the effectiveness of different approaches varies significantly based on application context and data characteristics. However, we also uncover critical challenges, including a lack of standardized evaluation metrics, prevalent overfitting problems, and limited dataset sizes, that hinder the field's progress. To address these, we propose a comprehensive set of evaluation metrics and emphasize the need for larger and more diverse datasets. The review introduces an innovative clustering approach that provides a nuanced understanding of research trends and methodological gaps. In addition, we discuss the ethical implications of DL in RUL estimation, addressing concerns about privacy and algorithmic bias. By synthesizing current knowledge, identifying key research directions, and suggesting methodological improvements, this review serves as a central guide for researchers and practitioners in the rapidly evolving field of EV battery management. It not only contributes to the advancement of RUL estimation techniques but also sets a new standard for conducting systematic reviews in technology-driven fields, paving the way for more sustainable and efficient EV technologies.

Keywords: lithium-ion battery; electric vehicle; battery management system; remaining useful life; machine learning; deep learning

check for updates

Academic Editor: Zhenbo Wang

Received: 18 November 2024

Revised: 26 December 2024

Accepted: 31 December 2024

Published: 3 January 2025

Citation: de la Iglesia, D.H.; Corbacho, C.C.; Dib, J.Z.; Alonso-Secades, V.; López Rivero, A.J. Advanced Machine Learning and Deep Learning Approaches for Estimating the Remaining Life of EV Batteries—A Review. *Batteries* **2025**, *11*, 17. https://doi.org/10.3390/batteries11010017

1. Introduction

The EV market has experienced a remarkable surge, driven largely by advancements in lithium-ion battery (LIB) technology. According to the International Energy Agency's Global EV Outlook 2024 [1], nearly 14 million new electric cars were registered globally in 2023, bringing their total number on the roads to 40 million. Electric cars now account for around 18% of all cars sold in 2023, up from 14% in 2022, with particularly strong adoption in China, Europe, and the United States, which together represent 95% of global electric car sales. This unprecedented growth underscores their growing acceptance as a viable alternative to internal combustion vehicles and marks a pivotal shift towards sustainable mobility.

The reliance of EVs on efficient and durable batteries has prompted the development of methodologies for predicting their RUL [2]. Accurate RUL estimation of paramount importance for the reliability and safety of EVs, while also promoting a circular economy through battery reuse [3]. This approach has the potential to extend battery life, reduce the necessity for new production, and mitigate the environmental impact of manufacturing and recycling.

The incorporation of ML and DL into LIB lifetime and degradation modeling has expanded the available tools for addressing the complex chemical and physical processes involved. Different methodological approaches offer distinct advantages depending on specific application requirements, data availability, and operational constraints. As demonstrated by Surucu et al. [4], these technologies have proven effective in developing intelligent monitoring systems and enhancing predictive maintenance across various industries, including the EV sector. Nevertheless, despite the potential of these technologies, there is a notable lack of comprehensive literature reviews to guide the selection of the most effective ML and DL methodologies for RUL analysis in batteries. This gap hinders the advancement of the field and the implementation of these technologies in real-world settings, ultimately impeding the full potential of EVs in the transition to sustainable mobility.

This review addresses the optimization of RUL estimation in EV batteries using ML and DL, with a particular focus on effective methodologies, prediction accuracy, and emerging trends. Our comprehensive analysis identifies critical research gaps including data limitations, the lack of standardized evaluation metrics, reproducibility issues, and insufficient critical analysis of model limitations. By evaluating various approaches and identifying research opportunities, this review aims to provide a clear guide for the scientific community and industry professionals, contributing to the advancement of EV battery management and sustainable mobility.

Contributions and Novelty

This systematic review is distinguished by its methodological approach and comprehensive analysis. Although previous reviews such as in [5,6] have made valuable contributions, our work addresses several gaps in the existing literature through the following:

- Conducting a PRISMA-compliant systematic review of 89 studies on ML and DL approaches for battery RUL estimation in EVs.
- Employing an adapted D&B scale for the quality analysis of ML and DL research.
- Implementing advanced clustering techniques to analyze literature trends and challenges.
- Explicitly evaluating study reproducibility.
- Providing practical implementation guidance and best practices for ML and DL in RUL estimation.

To further elucidate the distinctive features of our review and demonstrate its contribution to the existing literature, we present a comparative analysis in Table A1. This table provides a comprehensive overview of our approach in relation to other relevant reviews, highlighting the novel contributions and methodological innovations introduced in this study.

The remainder of this paper is organized as follows: Section 2 outlines the methodology employed in this study, including a detailed account of the PRISMA compliance and the adaptation of the D&B scale. Section 3 provides background information on EV batteries and the fundamentals of RUL estimation. In Section 4, we present a critical assessment of RUL estimation approaches, including an analysis of ML and DL studies and the identification of gaps. Section 5 synthesizes the results for RUL estimation, followed by a discussion of findings and implications in Section 6. Finally, Section 7 concludes the review and outlines directions for future work.

2. Methodology used in the Literature Review

This systematic review was conducted in accordance with the PRISMA guidelines [7] and with adaptations to the D&B scale [8], specifically for ML and DL studies. The objective of the study was to identify pertinent articles by addressing the following research questions:

1. How do different ML and DL methodologies perform in estimating the RUL of EV batteries, and what factors influence their effectiveness across different operational contexts and datasets?
2. How do input characteristics and parameters influence the accuracy of ML and DL models for RUL estimation in EV batteries? Which characteristics and parameters are most commonly used in the literature?
3. What are the main current challenges and limitations in the application of ML and DL for RUL estimation in EV batteries, and how are these challenges being addressed?
4. What emerging trends in research and development of ML and DL methodologies for RUL estimation in EV batteries are identified in recent literature?

Furthermore, in order to ensure the effective structuring of this systematic review, the components of the PICO framework (Population, Intervention, Comparison, Outcome) were established as follows:

- Population (P): LIBs installed or withdrawn from EVs or equivalent LIBs in terms of specifications.
- Intervention (I): Use of ML and DL algorithms for RUL estimation.
- Comparison (C): Evaluation of differences in data sources, sample volume, input parameters, ML/DL models, and complementary techniques.
- Outcome (O): Quantitative studies, with performance metrics as root mean squared error (RMSE).

The methodology employed in this systematic review is illustrated in Figure 1, which provides an overview of the process utilized in the review.

Identification: The search strategy employed in the ScienceDirect, IEEE Xplore, and Web of Science databases was oriented towards articles published between January 2016 and December 2023, with a particular emphasis on high-impact papers. The following search query was utilized, adapted to the specific syntax of each database:

```
(‘‘RUL estimation’’ OR ‘‘Remaining Useful Life’’)
AND (‘‘Electric vehicle’’ OR ‘‘Electric car’’ OR ‘‘EV’’)
AND (‘‘Deep Learning’’ OR ‘‘Machine Learning’’)
```

Despite initially exploring terms related to battery reuse, the limited number of articles found led us to broaden the spectrum of the review, excluding these terms to capture a wider range of relevant studies.

Screening: Initially, 518 studies were identified, of which 179 were excluded because they were not aligned with the focus of the review. Subsequently, 5 inaccessible studies were excluded.

Eligibility: In the eligibility phase, of the remaining 333 studies, 244 were excluded for focusing on battery condition, State of Health (SOH) prediction, battery degradation, or for being reviews or perspectives.

Inclusion: Finally, 89 studies were selected for the review.

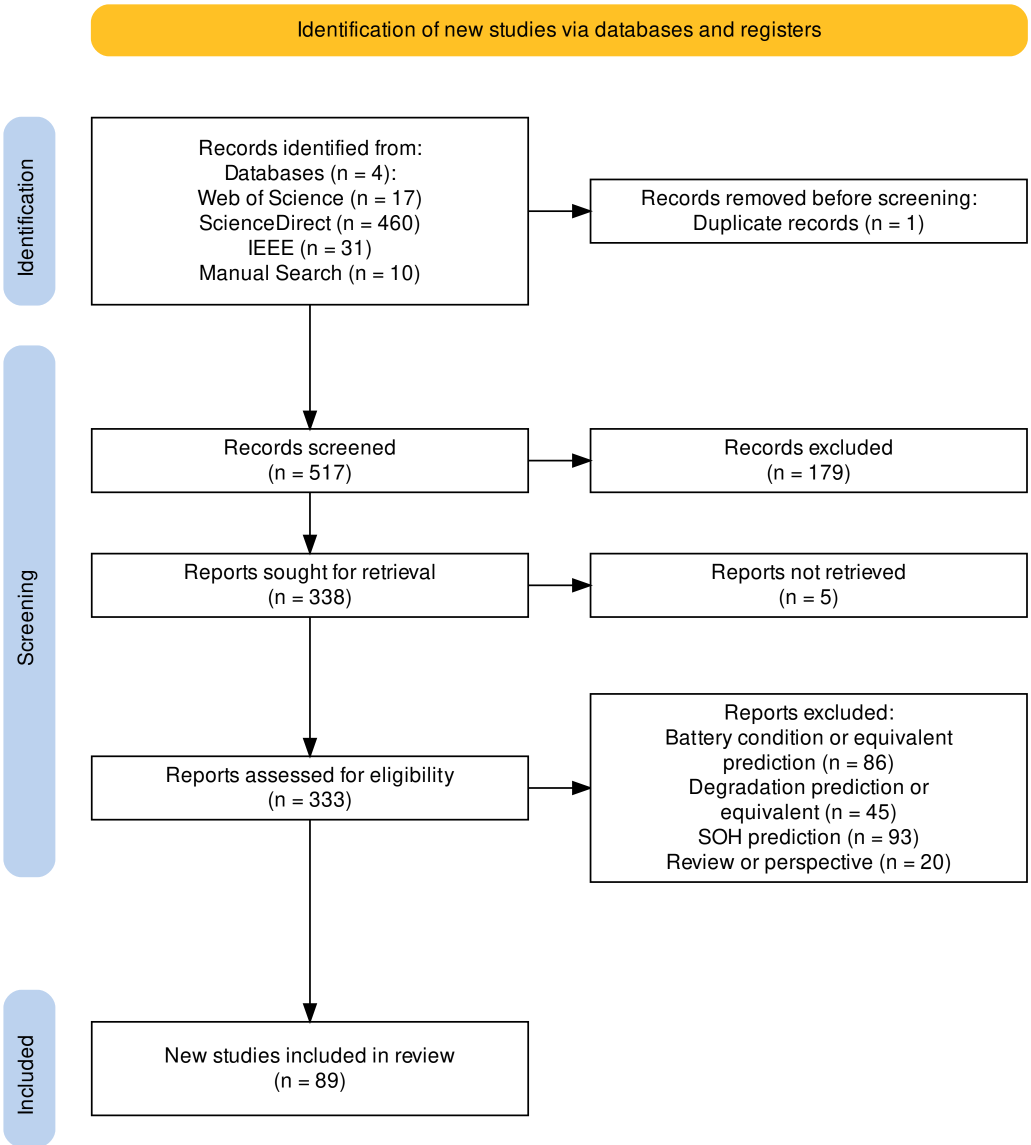

Figure 1. Methodology employed in review process. Generated using the tool [9].

The data extraction process was conducted in a systematic manner, with the use of a standardized form. Each study was reviewed by two investigators, and any discrepancies were resolved by a third party. The evaluation of key aspects included an assessment of the study characteristics, methodologies, data used, and reproducibility. The risk of bias was assessed using an adapted D&B scale, with a particular focus on reporting quality, external validity, bias, confounding, and statistical power. Further details on the categorization of reproducibility and the adapted scale can be found in Appendices A and B, respectively.

Clustering Techniques

The application of clustering techniques in this review represents a methodological advancement over that of conventional systematic reviews. By employing a range of characteristics, including sample size, publication date, study quality (D&B scale), reproducibility, and methodology, this approach enables the identification of innovations and future research directions.

The process employed the *K-means* algorithm to categorize studies into groups based on shared characteristics, followed by a *Random Forest* model to ascertain the most influential variables in cluster formation. Decision trees were then utilized to develop specific rules delineating the defining characteristics of each cluster. This advanced analysis provides detailed insights into trends in EV battery RUL research, establishing a solid foundation for identifying key areas for future innovation and development.

3. Preliminary Information

3.1. Fundamentals of EV Batteries and Their Critical Role

EV batteries, primarily LIBs, operate by converting chemical energy into electrical energy through electrochemical reactions. LIBs are composed of an anode, cathode, electrolyte, and separator. During the discharge phase, lithium ions are transported from the anode to the cathode, thereby generating an electron flow within the external circuit. The aforementioned process is reversed during the charging phase.

LIBs have become the predominant technology in the EV market due to their high energy density and durability. However, they face challenges related to cost, safety, and environmental impact. Research into alternative technologies has led to developments in solid-state batteries, which offer enhanced safety and potentially higher energy density, though they currently face challenges in production costs and RUL [10]. Sodium-ion batteries have emerged as a potential alternative, primarily motivated by concerns about lithium resource scarcity, though they currently demonstrate lower energy density compared with LIBs [11]. While these technologies show promise in addressing specific limitations of LIBs, significant technological and economic barriers remain before they can achieve widespread commercial adoption.

The EV battery lifecycle presents a number of challenges that must be addressed. A comprehensive lifecycle assessment is essential for understanding and mitigating the ecological footprint of EV batteries, from manufacturing to disposal [12]. Techniques for advancing battery recycling and exploring second-life applications in less demanding systems are essential for sustainability [13,14]. Accurate monitoring of the SOH and state of charge (SOC) is critical for efficient battery management. Additionally, innovative solutions for equitable energy distribution and smart charging systems are necessary, especially during peak demand periods [15].

3.2. LIBs Degradation Mechanism Phenomena

The deterioration of LIBs is the result of complex physical and chemical interactions that influence their functionality and operational lifespan. These processes include, but are not limited to, the formation of the solid–electrolyte interface at the electrodes and particle fracture, which results in the loss of electronic contact. Such phenomena are critical because they directly affect battery capacity and efficiency. This analysis is based on the understanding that degradation is a multifaceted process, affected by the structure and composition of the battery materials, as well as by the operating conditions under which it is used. The interaction between these factors determines the rate and degree of degradation, underscoring the importance of optimizing both the materials and the battery design to minimize these adverse effects [16].

3.3. Understanding RUL: Definition and Significance

As has been emphasized throughout this review, an understanding of the RUL is of paramount importance for the optimization of the use and maintenance of LIBs in critical applications such as EVs. The RUL of a battery is the amount of time or cycles of use remaining before the battery reaches the end of its operational life. This is commonly defined as the point at which the battery's capacity drops below 80% of its original capacity. Research and development in this field aim to enhance RUL estimation methodologies, addressing challenges such as variability in usage conditions and heterogeneous battery aging, to guarantee optimal performance and extend battery life [17].

The RUL plays a foundational role throughout the entire life cycle of a LIB. During its period of use, the RUL is a key indicator for determining when a battery is no longer fit for use in EVs, thus enabling its reassignment to a second life cycle. In the collection and post-disposal evaluation phase, the RUL is a crucial factor in determining whether batteries should be recycled or remanufactured for reuse. Furthermore, the RUL provides guidance in the recycling process, identifying the optimal point at which a battery should be treated to recover essential materials.

3.4. Key Datasets

To facilitate the validation and comparison of different RUL estimation methods, it is crucial to have publicly available datasets. Based on our review and the comprehensive analysis performed by [18], we identified four predominant datasets in EV battery RUL research:

- NASA Ames Prognostics Center of Excellence (PCoE) Battery Dataset [19]: This set contains data from 34 18650 lithium-ion cells with a nominal capacity of 2 Ah. The cells were cycled at different ambient temperatures (4 °C, 24 °C, 43 °C) with different discharge regimes. The kit includes measurements of current, voltage, temperature, discharge capacity, and EIS impedance measurements. It is particularly valuable for studying battery aging under different operating conditions (https://data.nasa.gov/dataset/Li-ion-Battery-Aging-Datasets/uj5r-zjdb/about_data, accessed on 30 December 2024).
- CALCE Battery Group Dataset [20]: Provides data from several LCO cell types, including 15 CS2 prismatic cells and 12 CX2 prismatic cells. Experiments explore different depths of discharge, partial charge ranges, and charge/discharge protocols. This set is particularly useful for studying the effects of different usage patterns on battery aging (https://web.calce.umd.edu/batteries/data.html, accessed on 30 December 2024).
- Toyota Research Institute dataset [21]: This comprehensive set contains data from 357 commercial LFP/graphite cells. It focuses on the effect of fast-charging protocols on cell aging, with 124 cells subjected to 72 different fast-charging profiles and 233 additional cells used to optimize these protocols. It is particularly valuable for investigating charging strategies that maximize battery life (https://data.matr.io/1/projects/5c48dd2bc625d700019f3204, accessed on 30 December 2024).
- Dataset from the Battery Intelligence Lab at the University of Oxford [22]: This dataset, known as the "path dependence battery degradation dataset", is the result of a 3-year project (2017–2020) to study path dependence in aging lithium-ion batteries. It contains data from 28 commercial NCA/graphite 18650 3Ah cells divided into 10 groups. The study combines calendar and cyclic aging periods, providing valuable information on how the order and periodicity of these types of aging affect battery degradation. Data include time, current, voltage, capacity, temperature, and results from the Reference Performance Test and EIS impedance (https://ora.ox.ac.uk/objects/uuid:03ba4b01-cfed-46d3-9b1a-7d4a7bdf6fac, accessed on 30 December 2024).

These datasets have been instrumental in the development and validation of RUL estimation models and provide a solid basis for comparing different methodological approaches. Each set offers unique insights into different aspects of battery aging, from different operating conditions to specific usage patterns and charging strategies. We encourage researchers to use these resources to validate the effectiveness of their estimation methods and gain a more complete understanding of the factors that influence EV battery life.

4. Critical Assessment of RUL Estimation Approaches

Our analysis indicates a notable increase in research activity on RUL estimation for EV batteries since 2016, with a particularly pronounced surge from 2020 onward. The Supplementary Materials (Tables S1–S3) provides an overview of studies employing ML techniques, DL approaches, and combinations of both methodologies. These data reveal a gradual adoption of these techniques over time, with DL applications accelerating more rapidly in recent years compared with the relatively stable trend in ML studies.

4.1. Gaps in Literature

Despite the growing body of research, our review has identified several crucial areas that require further investigation. These gaps underscore the complexity of RUL estimation and point to opportunities for significant advancements in the field. By addressing these challenges, researchers can contribute to the development of more accurate, reliable, and generalizable RUL estimation methods, which will ultimately enhance the performance and longevity of EV batteries.

4.1.1. Data and Sample Limitations

- There is a paucity of studies examining the minimum number of cycles required for effective RUL predictions.
- The capacity of models to generalize is constrained by the limited size of the sample sets. This limitation impedes the effective learning of battery degradation patterns across a range of operational scenarios and conditions. Most studies that have been conducted depend on datasets comprising fewer than 24 battery samples, a number that is inadequate for fully capturing the intricacies of degradation behaviors.
- It is evident that a considerable discrepancy exists between the conditions observed in laboratory testing and the actual operational environments of vehicles in real-world settings. While laboratory data offers controlled experimental conditions, it frequently lacks the capacity to capture the intricate interactions of real-world factors, including variable driving patterns, environmental conditions, and charging behaviors. This underscores the imperative for the incorporation of real vehicle operational data to develop more robust and practical models for predicting the RUL of vehicles.

4.1.2. Methodological Issues

- The comparison of different methodologies is significantly hindered by the use of diverse, often private datasets, making it difficult to determine whether performance differences stem from the methodologies themselves or from the characteristics of the underlying data.
- The issue is further exacerbated by the absence of standardization in evaluation metrics, which hinders the ability to make direct comparisons between studies, even when similar datasets are utilized.
- There is an inadequate assessment and mitigation of overfitting, as well as a dearth of empirical evaluation of preventive techniques.

4.1.3. Transparency and Reproducibility

- The majority of studies demonstrate only moderate reproducibility, indicating a need for enhanced documentation and the open sharing of research resources.

4.1.4. Limited Critical Analysis

- There is an insufficient discussion of the limitations, biases, and generalizability of the models to unseen conditions or real-world scenarios.

While the challenges are common across various domains employing ML and DL techniques, the estimation of RUL in EV batteries presents its own unique set of domain-specific challenges that require particular attention and specialized solutions. These challenges arise from the inherent complexity of battery systems, their operational environments, and the long-term nature of degradation studies.

4.1.5. Domain-Specific Challenges

- Data quality issues are prevalent due to inherent noise in long-term battery testing, with equipment calibration drift and environmental variations affecting measurements over extended periods (spanning years) required to reach end-of-life conditions. This is particularly evident in datasets like NASA and Toyota, where long-term monitoring introduces measurement uncertainties and inconsistencies that must be carefully addressed.
- Accurate labeling is particularly challenging in real EV environments due to the following:
  - Dynamic operating conditions that affect battery performance, including varying discharge rates, depth of discharge, and charging protocols.
  - Variable temperature profiles during operation, ranging from extreme cold to hot conditions, which significantly impact battery behavior and aging patterns.
  - Difficulties in obtaining ground truth SOH values under diverse usage patterns, as traditional capacity measurement methods may not be feasible during normal vehicle operation.
  - Sensor drift and calibration issues in operational environments, compounded by the cost and complexity of maintaining high-precision measurements in commercial vehicles.
- The time-dependent nature of degradation mechanisms presents unique challenges for prediction models, as the same usage patterns may lead to different degradation rates depending on the battery's age and history. This path dependence of degradation makes it particularly challenging to develop models that can accurately capture the complex relationships between usage patterns and aging effects.
- There is significant variability between individual cells, even within the same batch, making it difficult to generalize degradation patterns. This manufacturing variability, combined with different usage histories, creates a complex prediction landscape where models must account for both inherent cell differences and operational factors.

It is of the utmost importance to address these gaps in order to facilitate advancements in the field of RUL estimation in EV batteries. The D&B scale analysis, which is detailed in the Supplementary Material, and Figure 2, demonstrate a discrepancy between the observed quality of studies and the maximum possible score, with an average of 16 out of 26 points. This indicates that the quality of studies is acceptable, but that there is significant room for improvement. This emphasizes the need for enhanced research practices in this field.

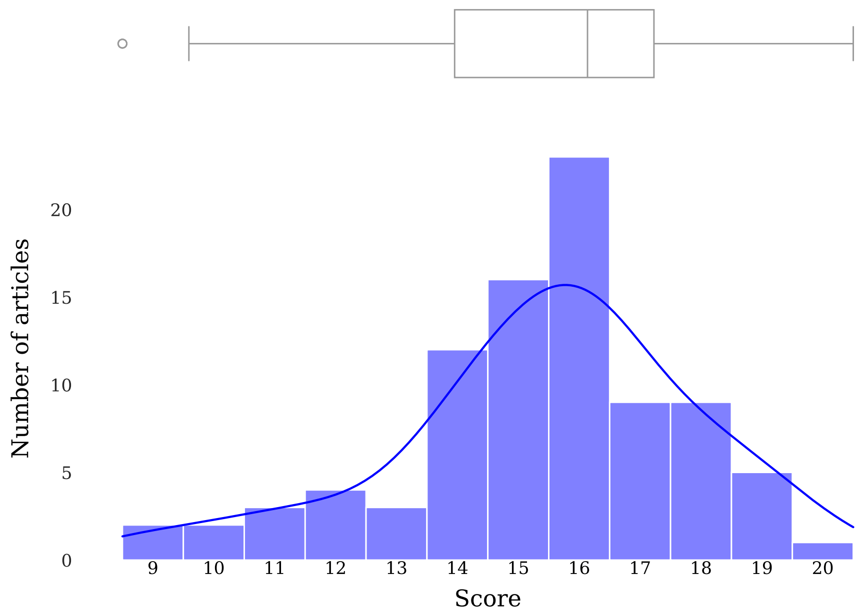

Figure 2. Histogram and boxplot of the score obtained by the articles in the D&B scale.

4.2. Overcoming RUL Estimation Challenges

The estimation of RUL of LIBs is a challenging undertaking due to the complex electrochemical processes and variable operating conditions involved. While conventional methods have inherent limitations, the integration of physical models with machine learning algorithms and statistical techniques has led to significant advancements. These approaches have facilitated more accurate and adaptive RUL estimates by enhancing the interpretation of historical and real-time data.

The advancement of computational capabilities has facilitated the adoption of diverse methodological approaches for RUL estimation. Traditional ML methods often offer advantages in scenarios with limited data or computational resources, while DL approaches can excel in capturing complex temporal patterns when sufficient data is available. Hybrid approaches combining multiple techniques have emerged as a promising direction, though their effectiveness depends heavily on the specific application context and available resources.

The choice of methodology involves careful consideration of various factors:

- Data availability and characteristics.
- Computational resources and real-time processing requirements.
- Need for model interpretability.
- Specific operational conditions and constraints.
- Required prediction horizons and accuracy levels.

Recent research explores the potential of digital twin models for EV battery management [23]. These models, which utilize incremental learning and cloud computing for SOC and SOH estimation, represent a promising direction for future RUL research and predictive maintenance. However, challenges remain, particularly in addressing heterogeneity in battery aging and uncertainties in input data. The ongoing development of predictive models and data fusion techniques to integrate information from various sources and sensors is crucial for effective battery lifecycle management.

4.3. Leveraging Advances in Deep Learning

DL, an advanced subset of ML, employs deep neural networks to analyze data at multiple levels of abstraction. In contrast to traditional ML, which necessitates the manual extraction of features, DL is capable of autonomously learning these features, thereby enhancing the accuracy and efficiency of complex tasks.

In the estimation of RUL, DL approaches have demonstrated various capabilities in RUL estimation, with their effectiveness particularly dependent on data availability and specific operational requirements. While these methods can capture complex patterns in large datasets, their performance advantages must be weighed against computational requirements and the need for extensive training data. Different architectural choices (LSTM, CNN, Transformers) present distinct trade-offs between accuracy, computational efficiency, and interpretability.

A notable consideration in RUL prediction approaches is the visualization and interpretation of battery degradation patterns. Couture and Lin [24] pioneered an innovative approach by exploring the use of graphical data interpretation through DL models and pretrained neural networks (see Figure 3), opening new possibilities in the field of RUL prediction. While this early work marked an interesting direction for feature extraction innovation, subsequent research by Lee et al. [25] has highlighted important considerations about the generalizability of such visual representations. Their analysis demonstrates how voltage, current, and capacity trends during charging and discharging cycles, though following characteristic patterns as batteries degrade, can vary significantly depending on data processing approaches and experimental conditions. Different interpolation methods can lead to varying trend interpretations, including unexpected patterns that may arise from measurement artifacts or physical phenomena such as SEI layer stabilization. Although the field has evolved towards other approaches, the initial exploration by Couture and Lin [24] represents an important milestone in expanding the boundaries of feature extraction methods in battery degradation analysis.

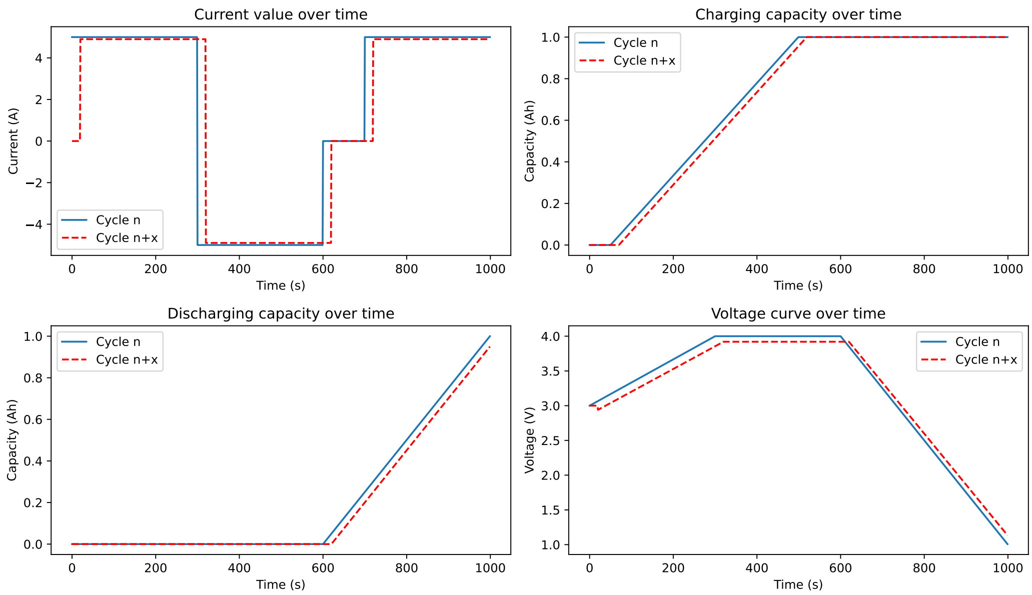

Figure 3. Simplified visual representation of the degradation occurring in LIBs, used as input to the RUL estimation models. Based on the study in [24].

4.4. Comparative Analysis of Methodological Efficacy

Our analysis of the literature reveals distinct patterns in the evolution and effectiveness of RUL estimation methodologies. Based on the systematic review of 89 studies and their reported results (detailed in Supplementary Tables S1–S3), we can identify several key methodological trends and their relative performance.

DL approaches, particularly LSTM-based architectures, have shown promising results in temporal pattern recognition, with studies like the one in [26] achieving RMSE values below 0.01 and MAPE under 1%. CNN-based models have demonstrated effectiveness in feature extraction from raw signals, especially when combined with attention mechanisms [24]. More recently, Transformer architectures have emerged as powerful tools for handling long-term dependencies, though with increased computational demands.

Hybrid models have gained significant traction, with CNN-LSTM combinations showing particularly balanced performance. Models incorporating physical constraints or domain knowledge (e.g., [27]) have demonstrated improved generalizability. The integration of optimization algorithms (PSO, WOA) with deep learning has also shown potential in enhancing model adaptability.

Traditional ML approaches maintain their relevance, with Gaussian Process Regression demonstrating robust performance with limited data. Ensemble methods (Random Forest, XGBoost) and Support Vector Regression have shown consistent performance across different operating conditions, often with better interpretability than their deep learning counterparts.

However, several significant limitations must be acknowledged when comparing methodological efficacy across studies. The prevalent issue of overfitting, as discussed in Section 5, poses a substantial challenge to direct performance comparisons. Many studies report near-zero error rates, which raises concerns about the generalizability of their results and complicates the assessment of true methodological effectiveness. Furthermore, the lack of standardized evaluation metrics across studies, as evidenced by our analysis of the adapted D&B scale results (Table A2), introduces additional complexity in making meaningful comparisons between different approaches.

The heterogeneity of datasets employed across studies presents another significant barrier to comparative analysis. The use of private or proprietary data in many cases makes it particularly challenging to determine whether observed performance differences are attributable to the methodologies themselves or merely reflect variations in the underlying data characteristics. This challenge is further compounded by the considerable variation in experimental conditions and preprocessing methods across different studies, which introduces additional confounding factors in the assessment of methodological efficacy.

Given these limitations, while general trends in methodological efficacy can be identified, any conclusions must be interpreted with careful consideration of these contextual factors. To address these challenges and provide a more structured approach to methodology evaluation and selection, we propose a comprehensive set of evaluation metrics and implementation guidelines in Section 6. This framework aims to establish a more standardized approach for assessing and comparing RUL estimation methodologies while accounting for both theoretical performance and practical implementation considerations.

5. Synthesis of Results for RUL Estimation

The results section of this study is organized into three principal sections. The initial section, designated as Section 5.1, offers a comprehensive overview of the prevailing trends and methodologies employed in the estimation of RUL. The second part, Section 5.2, presents an analysis of the results obtained by applying the adapted D&B scale. This analysis is primarily concerned with the methodological rigor employed. Finally, Section 5.3 presents a discussion of the findings on the clustering techniques used to identify specific patterns and segmentations in the existing literature. The complete results of the data collected from each study, including detailed information on methodologies, datasets, and performance metrics, can be found in the Supplementary Material, specifically in Tables S4–S8.

5.1. General Observations

The Supplementary Material provides an overview of the academic output on RUL research using ML and DL methodologies over the eight years examined (Table S9), wherein a notable polarization in the distribution of publications is discernible. A small number of journals are distinguished by the considerable volume of articles published therein, which suggests that they have served as the primary channels for research and debate in the field. In contrast, the majority of journals have contributed only one or two articles, indicating a broad but diluted attention to the topic in question. With regard to geographic distribution (Table S10), China not only has the largest number of publications, which evinces its dominant position in research in this domain, but was also a pioneer in initiating exploration of this topic.

5.1.1. Dataset

The role of datasets in ML and DL methodologies for EV RUL estimation is of paramount importance. The most prominent datasets in this field are those developed by NASA Ames PCoE [19] and MIT Stanford-Toyota Research Institute [21,28]. The dataset from the NASA Ames PCoE, employed in 55 of the 89 studies, furnishes indispensable information, including in-cycle and electrochemical impedance spectroscopy measurements for 34 lithium-ion cells. The MIT Stanford-Toyota dataset, utilized in 26 studies, encompasses comprehensive data for 357 lithium iron phosphate cells, with a particular emphasis on the impact of diverse charging protocols on cell aging. Moreover, the Center for Advanced Life Cycle Engineering (CALCE) [20] has a set of 27 lithium cobalt oxide cells that have been utilized in 20 studies. This set is focused on the analysis of the influence of different charging and discharging protocols on degradation.

Other datasets of relevance include those from the University of Oxford [22] and four additional datasets [29–32], which illustrate the diversity of resources available for research purposes. Nevertheless, the inclusion of proprietary datasets in 20 studies underscores the imperative for promoting transparency and open access to data, which is vital to accelerate progress in RUL research. This represents a significant challenge in this field.

Figure 4 illustrates the prevalence of studies that utilize a limited number of samples due to constraints related to cost and data availability. However, there has also been a notable effort to perform more extensive analyses. This reflects a growing tendency toward experiments with a reduced number of samples, coupled with a notable increase in the mid-range. Of particular note is the MIT dataset, which is of a public nature and comprises a considerable number of samples.

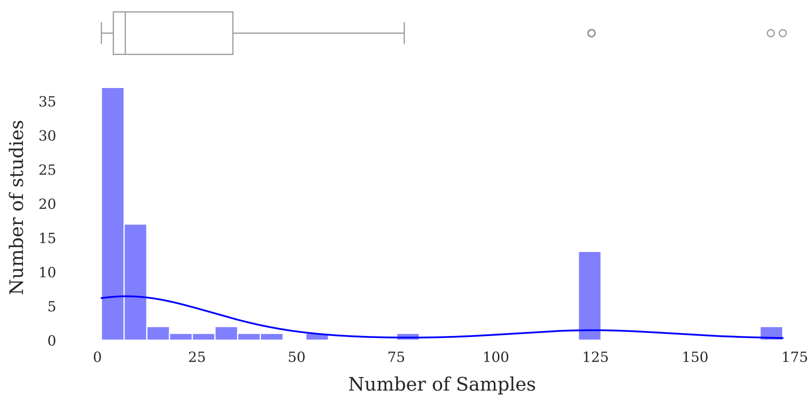

Figure 4. Histogram and boxplot of the number of samples used in the studies.

5.1.2. Variables Analyzed

In the modeling of the RUL, there is a notable preference for the use of scalar characteristics, which are employed in approximately 95% of the studies. These characteristics are of particular importance as reliable indicators in predictive modeling, thereby underscoring their significance for accurate RUL estimation.

Moreover, the inclusion of temporal and sequential metrics in approximately 32% of the studies underscores the significance of dynamic information and temporal sequences. This integration demonstrates that the evaluation of RUL benefits considerably from the inclusion of these data, especially when combined with DL methodologies, which are specialists in the analysis of these types of variables. Additionally, there is considerable interest in derived characteristics and signal decomposition techniques, as evidenced by studies that calculate battery status indicators, such as SOH and SOC.

The study of the SOH of batteries is addressed in 22 studies, which demonstrates its relevance in assessing the long-term viability of batteries. The study of the SOC, which reflects the available charge relative to the total capacity of the battery, is central to eight studies. These derived indicators are of great importance in order to enhance the precision of RUL estimation and to provide a more comprehensive representation of battery health. They serve to exemplify how transformed data can enrich RUL modeling. These strategies, in conjunction with advanced physical measurements [26,27,33–36], as well as early explorations in image analysis [24,37], opened new perspectives on feature extraction methodologies in RUL modeling, contributing to the evolution of the field.

5.1.3. Methodology

Research in RUL estimation of EV battery RUL shows a strong bias towards the use of DL techniques, which were employed in 46% of the studies identified. This approach aligns with the significance of temporal and sequential metrics in the input data, which illustrates the capacity of neural networks and hybrid methods to process temporal sequences and complex patterns. These capabilities render them particularly well suited for modeling the nonlinear dynamics of battery degradation.

The methodological landscape of RUL estimation shows considerable diversity, with different approaches demonstrating varying strengths across different application contexts. While hybrid models have gained increased attention, their adoption reflects a broader trend toward combining complementary techniques rather than a universal solution. The effectiveness of any approach appears highly dependent on factors such as dataset characteristics, operational requirements, and specific implementation constraints.

The importance of genetic and optimization algorithms is highlighted. These methods have been employed in 22 of the reviewed studies. Specific algorithms, including Particle Swarm Optimization [38–42] and the Whale Optimization Algorithm [42,43], are examples of strategies that seek to optimize feature selection and model parameters with the goal of improving the accuracy and efficiency of predictions.

In contrast, signal decomposition techniques, employed in 13 studies, align with the observed trend, underscoring the significance of temporal and sequential input data. These techniques, such as Empirical Mode Decomposition (EMD) [44,45] and Complete Ensemble EMD [46–48], facilitate the decomposition of complex signals into intrinsic components. This process enables a more comprehensive understanding and characterization of battery signals, which is essential for predictive models that rely on temporal precision and subtle pattern detection.

5.1.4. Metrics

The analysis of the RUL time is based on a range of metrics, which reflect the diversity of methods that researchers have adopted. The most prevalent metrics are root mean square error (RMSE), which is expressed in ampere-hours (Ah), and absolute error (AE), which is expressed in charge/discharge cycles. These metrics are featured in 40 and 37 studies, respectively. Notwithstanding their pervasive use, these metrics have not reached a consensus within the scientific community, indicating a lack of a unified standard for evaluating battery RUL. The diversity of techniques employed extends beyond these two metrics, which presents a challenge for comparison between studies.

Figure 5 illustrates a notable concentration of low root mean square error (RMSE) values (Ah), which suggests a high degree of accuracy in numerous RUL predictions. However, the presence of outliers indicates considerable variability in some models, which could be attributed to differences in experimental designs or data interpretation. In contrast, the AE metric (cycles) and the MAE (Ah) exhibit a larger scatter, together with persistent outliers, suggesting that, although certain models predict the RUL accurately, others show notable deviations from the true value. This may be indicative of generalization problems in the models.

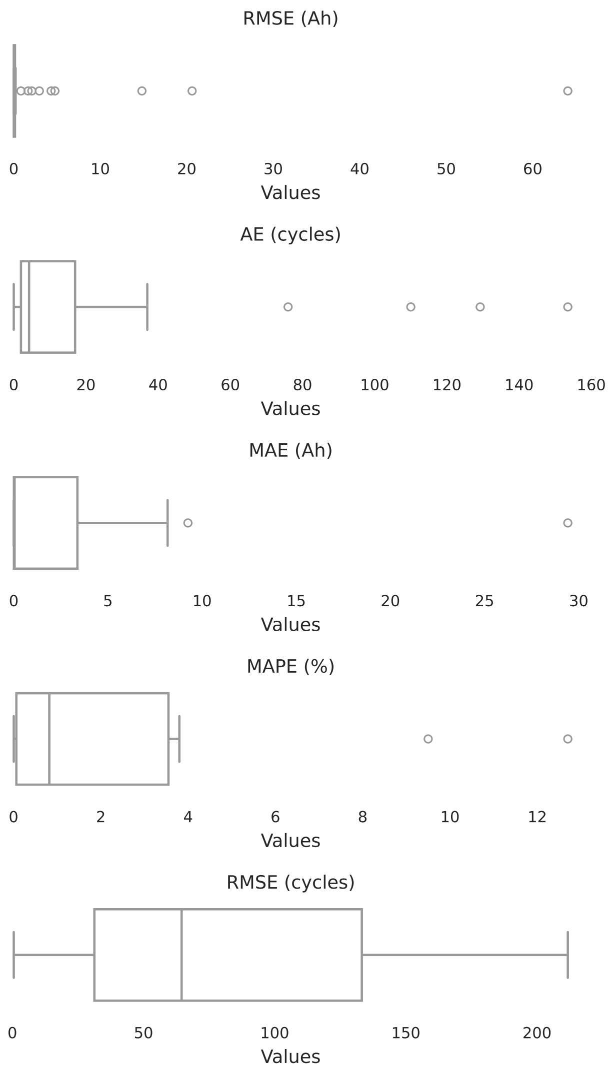

Figure 5. Boxplot of the results of the five most used techniques in the analyzed studies.

5.1.5. Reproducibility

The majority of studies have been classified as exhibiting moderate reproducibility. This suggests that, although the methodology and data are typically accessible, essential elements such as the source code or trained models, which are necessary for a complete replication of the investigation, are frequently unavailable. A smaller percentage of studies demonstrate limited reproducibility, wherein the methodology is not sufficiently detailed and private data are utilized, thereby further complicating the process of accurate replication. This situation highlights the need for more comprehensive documentation and dissemination of data resources to reinforce transparency and replicability in scientific research. It constitutes a call to the scientific community to enhance accessibility and clarity in the publication of research resources, which is pivotal to validating and replicating studies effectively. Table A2 shows the detailed percentage of studies meeting each individual item of the D&B scale assessment, providing a granular view of methodology quality across all evaluation criteria.

5.2. Adapted D&B Scale

Figure 2 illustrates a notable concentration around a high level of quality, with a mean of 15.75, as previously discussed. This reflects a central tendency within the spectrum of good quality, indicating that the majority of studies evaluated possess an acceptable level of methodological quality. However, there is considerable scope for improvement. In instances where critical areas emerge that require a more rigorous approach, such as the mitigation of overfitting, studies tend to employ techniques to reduce overfitting without sufficient empirical evaluation of their effectiveness. This gives rise to concerns regarding the generalizability of the models and their performance under conditions that were not observed during the training phase. Moreover, an analysis of the intrinsic constraints of the studies is notably absent. This is a crucial issue, as a thorough understanding of the limitations is essential for an accurate interpretation of the research outcomes and their applicability in real-world scenarios, where the models must be capable of accommodating a diverse range of operational conditions and battery types to be genuinely useful in practical applications. Table A3 presents the clustering analysis results, showing how studies are distributed across different methodological characteristics and quality dimensions.

5.3. Application of Clustering Techniques

The application of clustering techniques to the review of studies on the RUL estimation of EV batteries has revealed key patterns and methodologies in the field. By segmenting the data based on specific characteristics, this analysis offers a structured overview of the evolution of modeling techniques and underscores the significance of research quality and reproducibility. Table A3 illustrates the variability of the identified groups, reflecting the diversity and dynamism of the methodologies employed. This approach not only elucidates current trends and specific research areas of interest but also, by highlighting future research directions, reaffirms the critical value of clustering in the synthesis and analysis of the scientific literature, enhancing the understanding and orientation of future research in this technologically advanced field.

- **First clustering** (1st row)
  - Cluster 0: This group is distinguished by its inclusion of research utilizing a substantial quantity of data, with a minimum of 66 samples. This indicates that these studies may possess greater representativeness and applicability, as the utilization of a substantial dataset is pivotal for the effective training of predictive RUL models. The majority of these studies utilize the MIT database, which is the sole repository offering such a vast quantity of samples.

  - Cluster 1: This group is composed of studies with fewer than 66 samples, suggesting a more detailed approach or smaller-scale experiments. This may indicate research exploring new methodologies or specific applications in the RUL field.
  - Cluster 2: Includes studies that do not specify the number of samples used, indicating a potential lack of rigor or transparency in the methodological documentation.
- **Second clustering** (2nd row) A clustering analysis focused on the publication date reveals the evolution of methodologies and modeling approaches over time. This variable becomes pivotal in the classification of studies once the number of samples is disregarded. A noteworthy aspect observed in previous analyses is the considerable increase in the adoption of DL techniques in comparison with ML.
- **Third clustering** (3rd row) The third clustering analysis, which focuses on the score according to the D&B scale and disregards the number of samples and date of publication, reveals two groupings of studies based on their quality.
  - Cluster 0: Grouping of studies with a score of 14.5 or higher. This group is distinguished by its representation of high-quality research, characterized by sound methodology and significant results. The high score indicates that these studies adhere to rigorous standards in terms of experimental design, data analysis, and reporting of results, rendering them particularly relevant and reliable in the field of RUL research.
  - Cluster 1: Includes studies with a score equal to or lower than 14.5. This classification may indicate studies with certain limitations in their design, methodology, or the relevance of their findings. The lower score may reflect deficiencies in experimental robustness or in the detailed presentation of methods and results, which would affect the interpretation and applicability of their conclusions.

  This clustering approach, which focuses on the methodological quality of the studies according to the D&B scale adapted to ML and DL, provides insight into the level of scientific rigor present in RUL research. By considering the study score as a key indicator, this analysis helps to distinguish between high- and low-quality research, facilitating the identification of those studies that provide more robust and reliable contributions to the field of RUL in EV batteries.
- **Fourth clustering** (4th row) The fourth clustering analysis is concerned with the reproducibility and publication source of the studies, with the number of samples, publication date, and score being excluded from consideration. This approach underscores the significance of reproducibility and study provenance in investigating the RUL of EV batteries.
  - Cluster 0: This group is composed of studies with high or moderate reproducibility, primarily from the IEEE and a few manually aggregated. The manually aggregated studies were identified as eight in IEEE and two in ScienceDirect. This cluster suggests greater methodological rigor and transparency in the publications of these sources, which may facilitate replication and validation by other researchers.
  - Cluster 1: Composed of studies with low or limited reproducibility. These studies may present deficiencies in their experimental design, methodology, or in the detailed documentation of their procedures and results. These limitations may affect the reliability and validity of their findings, underscoring the necessity to enhance rigor and transparency.

  - Clusters 2: Groups studies with moderate or high reproducibility, whose primary sources are Web of Science (WOS) and Science Direct. This grouping indicates that studies published on these platforms also tend to maintain an adequate level of methodological rigor and data reporting.

  The distribution of studies across publication platforms suggests that a significant set of studies maintain a high level of methodological rigor and transparency, facilitating their replication and validation. However, the differentiation by source could indicate subtle variations in editorial standards related to other variables analyzed, such as the approach, the focus of the study, and the type of publication.

- **Fifth clustering** (5th row) By omitting variables such as the number of samples, the date of publication, the score, reproducibility, and the source of publication, this approach allows for a clearer understanding of the methodology and technical aspects employed in studies on the estimation of RUL. This allows for the identification of the methodologies that are currently predominant and evolving in the field of research.
  - Cluster 0: This group comprises studies that focus on the utilization of DL or the combination of DL and ML. The preponderance of studies employing either DL or a combination of DL and ML signals a pronounced inclination toward the development of sophisticated models.
  - Cluster 1: Comprising studies utilizing ML techniques, this cluster represents a more conventional or established methodology in RUL research. Despite the growing prominence of DL, ML remains a pertinent approach and is being employed in the research.

6. Discussion

Our systematic review reveals a significant paradigm shift in the field of RUL estimation for EV batteries, with a clear trend towards DL methods over traditional ML approaches. This transition reflects the increasing complexity of battery data and the need for more sophisticated modeling techniques capable of capturing intricate degradation patterns. However, this shift brings with it a set of challenges that require careful consideration.

A primary concern identified in our analysis is the lack of standardized evaluation metrics, which significantly hinders the comparative assessment of different methodologies. The absence of a unified evaluation framework makes it challenging to objectively determine the most effective approaches, potentially slowing progress in the field. To address this, we propose a comprehensive set of evaluation metrics:

- Root Mean Square Error (RMSE): To assess the overall accuracy of predictions.
- Mean Absolute Error (MAE): To measure the average magnitude of prediction errors.
- Mean Absolute Percentage Error (MAPE): To provide a relative measure of prediction error.
- Coefficient of Determination (R^2): To evaluate how well the model explains the variability in the data.
- Maximum Error: To identify the worst predictions and evaluate the robustness of the model.
- Accuracy within a threshold (e.g., ±10% of the actual RUL): To assess the practical reliability of the model.
- Computational time: To assess the computational efficiency of the model.

The implementation of these metrics would not only facilitate more meaningful comparisons between studies but also provide a more comprehensive assessment of model performance, encompassing both accuracy and practical applicability.

Our analysis has revealed a significant issue, namely the prevalence of overfitting, as evidenced by near-zero error rates in key metrics such as RMSE and AE. This raises significant concerns about the generalizability of current models to real-world scenarios, particularly given the limited datasets used in many studies. The adapted D&B scale further emphasizes this problem, with most studies failing to adequately address overfitting (item 7 of the scale). This underscores the need for more rigorous validation techniques and the development of models that can maintain performance across diverse operational conditions.

The limited sample sizes observed in many studies further complicate the process of generalization. Our analysis reveals a stark dichotomy: while 17.98% of studies utilize large datasets comprising over 66 samples, the majority (70.79%) employ more limited data. This discrepancy underscores the imperative for larger, more diverse datasets that accurately reflect real-world operational conditions and battery life cycles. The MIT dataset emerges as a valuable resource in this context, offering a substantial number of samples that facilitate more robust model training. However, the field would benefit from additional comprehensive datasets that capture the full spectrum of battery types, usage patterns, and environmental conditions encountered in practical EV applications.

Notwithstanding these challenges, DL exhibits considerable promise for addressing the intricate, large-scale datasets that are typical in EV battery applications. Its capacity to automatically extract pertinent features and model complex relationships renders it particularly well suited to the multifaceted nature of battery degradation processes. Nevertheless, ML techniques retain distinct advantages in scenarios that require computational efficiency, interpretability, and resilience to overfitting. This indicates that hybrid approaches, which leverage the strengths of both ML and DL, may represent a particularly fruitful avenue for future research.

The field is actively addressing these challenges through the implementation of advanced optimization algorithms, including regularization techniques, dropout mechanisms, and adaptive optimization methods. Of particular note is the emergence of genetic algorithms as optimization tools, observed in 22 of the reviewed studies. This innovative approach represents a promising direction for enhancing both the accuracy and generalizability of DL models, potentially leading to more reliable RUL estimates across a wider range of operational scenarios.

A notable discrepancy exists in the literature with regard to the minimum quantity of operational data required for accurate RUL estimation. The fact that only 15 studies specified the number of cycles needed underscores the necessity for focused research into models capable of making accurate predictions with minimal data. This is of particular importance for real-world applications where comprehensive historical data may not be available. It would be beneficial for future research to prioritize the development of techniques that can provide reliable RUL estimates with limited data, potentially through the application of transfer learning, data augmentation, or innovative feature extraction methods that maximize the information gleaned from available data points.

The issue of reproducibility has emerged as a significant concern in the field of RUL estimation for EV batteries. Our analysis indicates a correlation between data sources and reproducibility, with studies from certain sources (e.g., ScienceDirect) exhibiting higher reproducibility. This finding underscores the pivotal role that publication venues play in influencing research practices and highlights the necessity for consistent, rigorous standards across the field. To enhance the transparency and replicability of research in this domain, the scientific community must prioritize improved documentation practices and the open sharing of datasets and code.

The field is actively developing approaches to address these domain-specific challenges. To handle measurement noise and sensor drift, researchers are implementing sophisticated preprocessing techniques and robust feature extraction methods. For example, Wang et al. [34] proposes an improved anti-noise adaptive LSTM architecture specifically designed to handle noisy battery data. The challenge of accurate labeling in real-world conditions is being addressed through semi-supervised learning approaches and transfer learning techniques that can leverage limited labeled data from laboratory tests to improve predictions on real-world data. Nguyen and Bae [49] and Couture and Lin [24] demonstrate the effectiveness of transfer learning in this context. To account for cell-to-cell variations, ensemble methods and probabilistic approaches are being explored, with Najera-Flores et al. [33] showing promising results using Bayesian Neural Networks that can capture uncertainty in predictions.

6.1. Practical Implementation Guidelines

Building upon our analysis of methodological trends and challenges, we present a structured framework for implementing RUL estimation systems. The following guidelines, derived from successful implementations documented in the Supplementary Material (Tables S1–S3), offer practical recommendations addressing both technical and operational considerations:

- Selection of input variables
  - *Scalar vs. derived features:* Multiple works (e.g., [26,42,50]) report that combining basic signals (current, voltage, temperature) with derived features (e.g., voltage gradients, impedance) can improve the accuracy of RUL predictions.
  - *Temporal and sequential metrics:* In many recent studies (e.g., [51,52]), models leverage time-series and sequential indices (such as cycle counts or partial-discharge events) to capture the progressive nature of battery degradation.
- Signal preprocessing and decomposition
  - *Decomposition techniques:* Methods such as EEMD (Ensemble Empirical Mode Decomposition) or VMD (Variational Mode Decomposition) (see [53,54]) are often used to isolate noise and enhance the prognostic features.
  - *Transformations and filtering:* Wavelet transforms, Box–Cox, and CEEMDAN have been employed to stabilize variance and reveal latent patterns before training the RUL estimation model (e.g., [46,55,56]).
- Learning models and training strategies
  - *Neural networks and hybrid approaches:* Advanced architectures (LSTM, CNN, Transformers) are frequently used to handle complex, nonlinear relationships and long-term dependencies, while hybrid models (CNN–LSTM, attention-based) combine multiple structures for enhanced performance.
  - *Regression and Bayesian models:* When data are scarce or uncertain, some works favor Gaussian Process Regression (GPR) or ensemble methods (XGBoost, CatBoost, RF) because of their balance between interpretability and robustness (e.g., [50,53]).
  - *Hyperparameter optimization:* Techniques such as PSO, GA, or Bayesian search can improve model performance by systematically tuning hyperparameters, though at the cost of additional computation (see [57–59]).
- Case studies and scalability
  - *Data size and diversity:* Experiments under varied temperature and load conditions help models generalize to new scenarios. While certain investigations rely on

large datasets (e.g., [21]), others employ limited samples and sometimes adopt transfer learning ([37,49]).
   - *Real-time applications:* For on-board applications, models like ELM, PF, or efficient attention-based networks can meet real-time constraints, as presented in [47,60].
   - *Validation and metrics:* As identified in our analysis, the lack of standardized evaluation metrics hinders the comparative assessment of different methodologies. To address this, we recommend a comprehensive set of metrics—including RMSE, MAE, MAPE, R^2, Maximum Error, Accuracy within a threshold (e.g., $\pm 10\%$ of the actual RUL), and Computational Time—to capture both predictive accuracy and practical applicability. In certain contexts, advanced statistical tools (e.g., survival analysis) may also provide deeper insights into the long-term reliability of the model.
- Physical and experimental considerations
   - *Integration of physical models:* Combining data-driven techniques with electrochemical representations can yield more robust and interpretable RUL predictions (see [33,49]).
   - *Sensitivity analysis:* Varying operational conditions (current, temperature, C-rate) affect battery aging differently, and identifying the most influential factors enhances model accuracy and reliability.
   - *Scalability and repeatability:* Scaling laboratory results to industrial use requires additional validation and standardized calibration protocols to ensure consistent RUL forecasts.

Overall, these guidelines provide a structured approach to choosing input variables, preprocessing steps, modeling techniques, and validation practices for reliable battery RUL estimation. The diverse methods reported in the supplementary material (Tables S1–S3) confirm that no single strategy fits all needs; the ultimate choice should align with data availability, application constraints, and required prediction accuracy.

6.2. Ethical Considerations

The ethical implications of utilizing DL for RUL estimation are becoming increasingly pertinent. The potential for these models to process detailed vehicle usage data raises concerns regarding individual privacy protection. Additionally, there is a risk of algorithmic bias if the training data does not adequately represent diverse user profiles and operating conditions. To address these issues, future research should focus on developing ethical and fair DL models, potentially incorporating techniques such as federated learning to protect user privacy and explainable AI to enhance model transparency and interpretability.

7. Conclusions and Future Works

This review introduces a novel approach to RUL estimation for EV batteries. This approach employs the PRISMA methodology and adapts the D&B scale for ML and DL research. Our comprehensive analysis, enhanced by clustering techniques, has facilitated a more nuanced understanding of the current trends and challenges in the field. Notable findings include a notable shift towards DL methodologies, though comparisons are hindered by inconsistent evaluation metrics. Furthermore, we have identified a prevalent issue of overfitting, particularly in studies with limited datasets. This underscores the necessity for improved generalization across diverse operating conditions and battery types. Furthermore, our analysis indicated a lack of attention to the principles of reproducibility and transparency in research practices.

In light of these findings, we put forth several avenues for future research. It is imperative to develop techniques for accurate RUL estimation with limited data, potentially through data augmentation or transfer learning. Furthermore, future studies should integrate advanced physical characteristics into ML and DL models to enhance their robustness and applicability. Improving model interpretability, particularly for complex DL approaches, remains a crucial challenge. Additionally, it is becoming increasingly important to investigate the ethical implications and potential biases in ML/DL models for RUL estimation as these technologies advance.

The adapted D&B scale introduced in this study has the potential to be applied more broadly in the assessment of research quality in the field of machine learning and deep learning across a range of domains. Its emphasis on universal principles, reproducibility, and comprehensive evaluation provides a foundation for the standardization of best practices in the field.

To further advance this research, we intend to validate the generalizability of various ML/DL models by training them on established datasets and testing them on a new, diverse dataset of LIBs with varying chemistries and operating conditions. This approach is intended to facilitate a more equitable comparison of model performance and applicability. In accordance with our emphasis on reproducibility, we pledge to disseminate this novel dataset, along with the associated code and pretrained models, to the scientific community.

Furthermore, our analysis highlights the unique challenges faced in RUL estimation for EV batteries, particularly regarding data quality issues in long-term testing, accurate labeling in dynamic operating conditions, and cell-to-cell variations. Future research should focus on developing robust approaches to handle measurement noise and sensor drift, as well as methods that can effectively transfer knowledge from controlled laboratory conditions to real-world EV operations. The development of standardized protocols for handling these domain-specific challenges will be crucial for advancing the field.

By addressing these challenges and pursuing these research directions, we aim to contribute to the advancement of sustainable and efficient solutions in EV battery management, thereby supporting the broader transition to green mobility. Our work not only highlights the current state of RUL estimation techniques but also sets a new standard for systematic reviews in this rapidly evolving field, thereby paving the way for more rigorous and impactful research in the future.

For researchers new to this field, we particularly recommend the following studies for their high methodological quality and significant contributions: Chen et al. [26] (score: 20, high reproducibility) presents an innovative approach using federated learning for RUL estimation; Couture and Lin [24] (score: 19, moderate reproducibility) stands out for its innovative early exploration of computer vision techniques for feature extraction; Hong et al. [61] (score: 19, moderate reproducibility) provides a robust implementation of dilated convolutional networks for large datasets; Zhang et al. [62] (score: 19, moderate reproducibility) introduces an effective TCN-DCN fusion model with innovative filtering techniques; Fei et al. [63] (score: 19, limited reproducibility) presents a notable deep attention-assisted temporal convolutional approach for scenarios with limited data; and Ardeshiri et al. [64] (score: 18, moderate reproducibility) develops a robust multivariate stacked bidirectional LSTM architecture. Although not covered in this review, we also strongly recommend the work by Lee et al. [25] for its comprehensive analysis of various neural network architectures and its insights into the importance of robust feature engineering over visual pattern analysis in RUL prediction. These works stand out for their methodological rigor, clarity in presentation, and reproducible results, providing a solid foundation for understanding the current state of the field.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/batteries11010017/s1.

Author Contributions: Conceptualization, C.C.C. and D.H.d.l.I.; Methodology, J.Z.D., D.H.d.l.I., C.C.C. and A.J.L.R.; Software, J.Z.D.; Validation, V.A.-S., D.H.d.l.I. and A.J.L.R.; Investigation, D.H.d.l.I. and C.C.C.; Data curation, C.C.C.; Writing—original draft, J.Z.D., C.C.C. and D.H.d.l.I.; Writing— review & editing, C.C.C.; Supervision, V.A.-S. and A.J.L.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work is part of the CERBEA project "Intelligent recycling of lithium batteries for the manufacture of domestic energy accumulators based on circular economy". Grant TED2021-131981A-I00 funded by MCIN/AEI/10.13039/501100011033 and by the "European Union NextGenerationEU/PRTR".

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

Abbreviations

The following abbreviations are used in this manuscript:

AE	Absolute Error
Ah	Ampere-hour
BMS	Battery Management System
CALCE	Center for Advanced Life Cycle Engineering
CYCLE	Charge/Discharge Cycle
D&B	Downs and Black
DL	Deep Learning
DMS	Data Management System
EMD	Empirical Mode Decomposition
EV	Electric Vehicle
GDPR	General Data Protection Regulation
IEEE	Institute of Electrical and Electronics Engineers
K-means	K-means Clustering Algorithm
LIB	Lithium-ion Battery
LCO	Lithium Cobalt Oxide
LFP	Lithium Iron Phosphate
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MIT	Massachusetts Institute of Technology
ML/DL	Machine Learning and Deep Learning
NASA	National Aeronautics and Space Administration
PCoE	Prognostics Center of Excellence
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
R^2	Coefficient of Determination
RMSE	Root Mean Squared Error
RUL	Remaining Useful Life
SOH	State of Health
SOC	State of Charge
WOS	Web of Science

Appendix A. Categorisation of Reproducibility

In assessing the reproducibility of battery RUL estimation studies, a proposed scale has been adopted that classifies reproducibility into four categories: Low, Limited, Moderate, and High. This classification is based on several key criteria, including detailed methodological description, data accessibility, and availability of additional resources such as source code or pretrained models.

- Low reproducibility: Deficiency in methodological details and/or data accessibility, hindering independent replication.
- Limited reproducibility: Partially detailed methodology but with shortcomings that limit complete and accurate replication.
- Moderate reproducibility: Access to methodology and datasets, but without additional resources such as source code or pretrained models.
- High reproducibility: Comprehensive methodology, accessible datasets, and supplementary resources that facilitate accurate replication.

Appendix B. Adapted D&B Scale

The adapted items and categories corresponding to the D&B scale specifically adapted for ML and DL studies are presented below:

Category 1: Report Quality (10 items)

1. Clarity in the statement of research objectives.
2. Detailed description of the dataset, its origin, and representativeness.
3. Detail of model architectures, algorithms, and parameters.
4. Specification of performance metrics and results obtained.
5. Discussion of limitations and potential biases of the models.
6. Description and justification of statistical and model evaluation methods.
7. Assessment and mitigation of overfitting with appropriate techniques.
8. Details to allow independent reproduction of the results.
9. Statement of sources of financing.
10. Declaration of possible conflicts of interest.

Category 2: External Validity (3 items)

11. Representativeness of datasets for real applications.
12. Comparisons of models with relevant studies.
13. Discussion of reproducibility of results in different settings and with different data.

Category 3: Bias (7 items)

14. Fair comparison between different models.
15. Prevention of bias in the selection of training and test data.
16. Objective evaluation of the models.
17. Consideration of the diversity and complexity of the data.
18. Use of valid methods to measure performance.
19. Strategies to prevent overadjustment.
20. Verification of the consistency of results in multiple executions.

Category 4: Confusion (6 items)

21. Comparison of model results with benchmarks.
22. Analysis of the generalization capacity and robustness of the model.
23. Attempts at replication or independent verification of results.
24. Methods for handling outliers and missing data.
25. Justification of methods for multivariate analysis.

Category 5: Statistical Power (1 item)

26. Assess statistical power to ensure generalization and avoid overfitting.

Each item on the scale is scored as follows:

- Yes: The study meets the criteria specified in the item.
- No: The study does not meet the criteria specified in the item.
- Cannot be determined: There is not enough information in the study to evaluate the item.

The score obtained for each study under this adapted scale is interpreted as follows:

- Excellent quality: More than 18 points.
- Good quality: Between 14 and 18 points.
- Fair quality: Between 9 and 13 points.
- Poor quality: Fewer than 9 points

Criteria for a positive evaluation of some of the key items are detailed below:

- Discussion of limitations and potential biases (item 5): Limitations inherent in your study methodology, potential bias in data selection or treatment, as well as limitations of the ML and DL models, should be explicitly and fully addressed.
- Assessment and mitigation of overfitting (item 7): To be considered positive in this criterion, it was not sufficient to implement techniques to reduce overfitting, studies are required to perform a critical and objective assessment of overfitting in their models. This means not only applying techniques to reduce overfitting but also empirically evaluating whether these techniques are effective.
- Discussion of reproducibility of results in different environments and with different data (item 13): To qualify as positive on this item, studies had to include a comprehensive and concrete discussion of the ability of the proposed models to generalize to varied charging and discharging protocols, different battery chemistries, or data obtained from real, practical environments. Studies that simply used various datasets without this in-depth discussion were not considered positive.

Table A1. Comparison of our review with other relevant reviews in the field.

Aspect	Our Review	[65]	[5]	[66]	[67]	[6]	[68]
Main Focus	ML and DL for RUL	SOH and RUL	Hybrid techniques for SOH/RUL	ML for SOC, RUL, and inflection pt.	DL for SOC, SOH, and RUL	RUL for LIB storage systems	Datasets, BMS, and RUL
Methodology	PRISMA	Traditional review	Traditional review	Traditional review	Traditional review	Traditional review	Traditional review
Quality Analysis	Yes, with adapted D&B scale	No	No	No	No	No	No
Clustering Techniques	Yes	No	No	No	No	No	No
Reproducibility	Explicitly evaluated	Not evaluated	Not evaluated	Not evaluated	Not evaluated	Not evaluated	Not evaluated
Overfitting Challenges	Discussed in detail	Not discussed	Mentioned	Not discussed	Discussed	Mentioned	Not discussed
Model Generalization	Discussed in detail	Mentioned	Mentioned	Not discussed	Discussed	Not discussed	Mentioned
Practical Implementation Guide	Yes	No	Limited	No	Limited	Limited	Limited
Implementation Factors Analysis	Yes	No	Limited	No	Yes	Yes	Yes

Table A2. Results of the D&B scale for each item. "No" includes direct negative responses and those items where a response could not be determined.

Category	Item	Yes (%)	No (%)
1	1	100	0
	2	92	8
	3	96	4
	4	97	3
	5	9	91
	6	81	19
	7	3	97
	8	88	12
	9	97	3
	10	94	6
2	11	67	33
	12	74	26
	13	6	94
3	14	94	6
	15	100	0
	16	100	0
	17	94	6
	18	93	7
	19	47	53
	20	78	22
4	21	2	98
	22	7	93
	23	0	100
	24	7	93
	25	0	100
5	26	7	93

Table A3. Summary of clustering results in selected studies. Details the different clustering configurations used.

Configuration	Cluster	% of Studies	No. of Studies	Main Characteristics	Rules
With all the columns	0	17.98%	16	Number of Samples: 0.47 Score: 0.09 Date: 0.07	Number of Samples $>$ 66.0
	1	70.79%	63	Number of Samples: 0.40 Number of Samples is NaN: 0.12 Score: 0.10	Number of Samples $\leq$ 66.0 Number of Samples is NaN $\leq$ 0.5
	2	11.24%	10	Number of Samples: 0.32 Number of Samples is NaN: 0.31 Score: 0.08	Number of Samples $\leq$ 66.0 Number of Samples is NaN $>$ 0.5
No number of samples	0	59.55%	53	Date: 0.56 Score: 0.09	No interpretable
	1	40.45%	36	Date: 0.62 Score: 0.07	No interpretable
No sample number or date	0	70.79%	63	Score: 0.57	Score $>$ 14.5
	1	29.21%	26	Score: 0.65	Score $\leq$ 14.5
No sample number, date or score	0	12.36%	11	Source: 0.53	Reproducibility $>$ 1.5 Source $\leq$ 1.5
	1	22.47%	20	Reproducibility: 0.62 Source: 0.03	Reproducibility $\leq$ 1.5
	2	65.17%	58	Reproducibility: 0.34 Source: 0.27	Reproducibility $>$ 1.5 Source $>$ 1.5
No sample number, date, score, reproducibility or source	0	67.42%	60	Methodology: 0.36 Neural Networks: 0.33	Methodology $\leq$ 1.5
	1	32.58%	29	Methodology: 0.42 Neural Networks: 0.30	Methodology $>$ 1.5

References

1. IAE. *Global EV Outlook 2024*; IEA: Paris, France, 2024.
2. Dubarry, M.; Devie, A.; McKenzie, K. Durability and reliability of electric vehicle batteries under electric utility grid operations: Bidirectional charging impact analysis. *J. Power Sources* **2017**, *358*, 39–49. [CrossRef]
3. Cusenza, M.A.; Guarino, F.; Longo, S.; Ferraro, M.; Cellura, M. Energy and environmental benefits of circular economy strategies: The case study of reusing used batteries from electric vehicles. *J. Energy Storage* **2019**, *25*, 100845. [CrossRef]
4. Surucu, O.; Gadsden, S.A.; Yawney, J. Condition Monitoring using Machine Learning: A Review of Theory, Applications, and Recent Advances. *Expert Syst. Appl.* **2023**, *221*, 119738. [CrossRef]
5. Alsuwian, T.; Ansari, S.; Zainuri, M.A.A.M.; Ayob, A.; Hussain, A.; Lipu, M.H.; Alhawari, A.R.; Almawgani, A.; Almasabi, S.; Hindi, A.T. A Review of Expert Hybrid and Co-Estimation Techniques for SOH and RUL Estimation in Battery Management System with Electric Vehicle Application. *Expert Syst. Appl.* **2024**, *246*, 123123. [CrossRef]
6. Ansari, S.; Ayob, A.; Lipu, M.H.; Hussain, A.; Saad, M.H.M. Remaining useful life prediction for lithium-ion battery storage system: A comprehensive review of methods, key factors, issues and future outlook. *Energy Rep.* **2022**, *8*, 12153–12185. [CrossRef]
7. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Int. J. Surg.* **2021**, *88*, 105906. [CrossRef] [PubMed]
8. Downs, S.H.; Black, N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J. Epidemiol. Community Health* **1998**, *52*, 377–384. [CrossRef] [PubMed]
9. Haddaway, N.R.; Page, M.J.; Pritchard, C.C.; McGuinness, L.A. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Syst. Rev.* **2022**, *18*, e1230. [CrossRef] [PubMed]
10. Thomas, F.; Mahdi, L.; Lemaire, J.; Santos, D.M. Technological advances and market developments of solid-state batteries: A review. *Materials* **2024**, *17*, 239. [CrossRef]
11. Cai, X.; Yue, Y.; Yi, Z.; Liu, J.; Sheng, Y.; Lu, Y. Challenges and industrial perspectives on the development of sodium ion batteries. *Nano Energy* **2024**, *129*, 110052. [CrossRef]
12. Miao, Y.; Liu, L.; Zhang, Y.; Tan, Q.; Li, J. An overview of global power lithium-ion batteries and associated critical metal recycling. *J. Hazard. Mater.* **2022**, *425*, 127900. [CrossRef] [PubMed]
13. Etxandi-Santolaya, M.; Casals, L.C.; Montes, T.; Corchero, C. Are electric vehicle batteries being underused? A review of current practices and sources of circularity. *J. Environ. Manag.* **2023**, *338*, 117814. [CrossRef]
14. Ali, H.; Khan, H.A.; Pecht, M.G. Circular economy of Li Batteries: Technologies and trends. *J. Energy Storage* **2021**, *40*, 102690. [CrossRef]
15. Bertolini, A.; Martins, M.S.; Vieira, S.M.; Sousa, J.M. Power output optimization of electric vehicles smart charging hubs using deep reinforcement learning. *Expert Syst. Appl.* **2022**, *201*, 116995. [CrossRef]
16. Birkl, C.R.; Roberts, M.R.; McTurk, E.; Bruce, P.G.; Howey, D.A. Degradation diagnostics for lithium ion cells. *J. Power Sources* **2017**, *341*, 373–386. [CrossRef]
17. Elmahallawy, M.; Elfouly, T.; Alouani, A.; Massoud, A.M. A comprehensive review of lithium-ion batteries modeling, and state of health and remaining useful lifetime prediction. *IEEE Access* **2022**, *10*, 119040–119070. [CrossRef]
18. Dos Reis, G.; Strange, C.; Yadav, M.; Li, S. Lithium-ion battery data and where to find it. *Energy AI* **2021**, *5*, 100081. [CrossRef]
19. Saha, B.; Goebel, K. Battery data set. In *NASA AMES Prognostics Data Repository*; NASA Ames Research Center: Moffett Field, CA, USA, 2007.
20. Battery Research Group. Battery Data. 2013. Available online: https://calce.umd.edu/battery-research-overview (accessed on 30 December 2024).
21. Severson, K.A.; Attia, P.M.; Jin, N.; Perkins, N.; Jiang, B.; Yang, Z.; Chen, M.H.; Aykol, M.; Herring, P.K.; Fraggedakis, D.; et al. Data-driven prediction of battery cycle life before capacity degradation. *Nat. Energy* **2019**, *4*, 383–391. [CrossRef]
22. Howey, D. Oxford Battery Team Data and Code. 2021 Available online: http://howey.eng.ox.ac.uk/data-and-code (accessed on 1 December 2024).
23. Eaty, N.D.K.M.; Bagade, P. Digital twin for electric vehicle battery management with incremental learning. *Expert Syst. Appl.* **2023**, *229*, 120444. [CrossRef]
24. Couture, J.; Lin, X. Image- and health indicator-based transfer learning hybridization for battery RUL prediction. *Eng. Appl. Artif. Intell.* **2022**, *114*, 105120. [CrossRef]
25. Lee, J.; Sun, H.; Liu, Y.; Li, X. A machine learning framework for remaining useful lifetime prediction of li-ion batteries using diverse neural networks. *Energy AI* **2024**, *15*, 100319. [CrossRef]
26. Chen, X.; Chen, Z.; Zhang, M.; Wang, Z.; Liu, M.; Fu, M.; Wang, P. A remaining useful life estimation method based on long short-term memory and federated learning for electric vehicles in smart cities. *PeerJ Comput. Sci.* **2023**, *9*, e1652. [CrossRef]

27. Li, H.; Fu, L.; Zhang, Y. A novel hybrid data-driven method based on uncertainty quantification to predict the remaining useful life of lithium battery. *J. Energy Storage* **2022**, *52*, 104984. [CrossRef]
28. Attia, P.M.; Grover, A.; Jin, N.; Severson, K.A.; Markov, T.M.; Liao, Y.H.; Chen, M.H.; Cheong, B.; Perkins, N.; Yang, Z.; et al. Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature* **2020**, *578*, 397–402. [CrossRef] [PubMed]
29. Bosello, M. UNIBO Powertools Dataset. 2021. Available online: https://cris.unibo.it/handle/11585/849218 (accessed on 30 December 2024).
30. Bole, B.; Kulkarni, C.; Daigle, M. Randomized battery usage data set. *NASA AMES Progn. Data Repos.* **2014**, *70*.
31. Ma, G.; Xu, S.; Jiang, B.; Cheng, C.; Yang, X.; Shen, Y.; Yang, T.; Huang, Y.; Ding, H.; Yuan, Y. Real-time personalized health status prediction of lithium-ion batteries using deep transfer learning. *Energy Environ. Sci.* **2022**, *15*, 4083–4094. [CrossRef]
32. Hu, P.; Ma, G.; Zhang, Y.; Cheng, C.; Zhou, B.; Yuan, Y. State of health estimation for lithium-ion batteries with dynamic time warping and deep kernel learning model. In Proceedings of the 2020 European Control Conference (ECC), St. Petersburg, Russia, 12–15 May 2020; pp. 602–607.
33. Najera-Flores, D.A.; Hu, Z.; Chadha, M.; Todd, M.D. A Physics-Constrained Bayesian neural network for battery remaining useful life prediction. *Appl. Math. Model.* **2023**, *122*, 42–59. [CrossRef]
34. Wang, S.; Fan, Y.; Jin, S.; Takyi-Aninakwa, P.; Fernandez, C. Improved anti-noise adaptive long short-term memory neural network modeling for the robust remaining useful life prediction of lithium-ion batteries. *Reliab. Eng. Syst. Saf.* **2023**, *230*, 108920. [CrossRef]
35. Ma, B.; Yang, S.; Zhang, L.; Wang, W.; Chen, S.; Yang, X.; Xie, H.; Yu, H.; Wang, H.; Liu, X. Remaining useful life and state of health prediction for lithium batteries based on differential thermal voltammetry and a deep-learning model. *J. Power Sources* **2022**, *548*, 232030. [CrossRef]
36. Wang, F.K.; Amogne, Z.E.; Chou, J.H.; Tseng, C. Online remaining useful life prediction of lithium-ion batteries using bidirectional long short-term memory with attention mechanism. *Energy* **2022**, *254*, 124344. [CrossRef]
37. Couture, J.; Lin, X. Novel Image-Based Rapid RUL Prediction for Li-Ion Batteries Using a Capsule Network and Transfer Learning. *IEEE Trans. Transp. Electrif.* **2023**, *9*, 958–967. [CrossRef]
38. Ye, L.H.; Chen, S.J.; Shi, Y.F.; Peng, D.H.; Shi, A.P. Remaining useful life prediction of lithium-ion battery based on chaotic particle swarm optimization and particle filter. *Int. J. Electrochem. Sci.* **2023**, *18*, 100122. [CrossRef]
39. Ansari, S.; Ayob, A.; Lipu, M.S.H.; Hussain, A.; Saad, M.H.M. Particle swarm optimized data-driven model for remaining useful life prediction of lithium-ion batteries by systematic sampling. *J. Energy Storage* **2022**, *56*, 106050. [CrossRef]
40. Ma, Y.; Yao, M.; Liu, H.; Tang, Z. State of Health estimation and Remaining Useful Life prediction for lithium-ion batteries by Improved Particle Swarm Optimization-Back Propagation Neural Network. *J. Energy Storage* **2022**, *52*, 104750. [CrossRef]
41. Yao, F.; He, W.; Wu, Y.; Ding, F.; Meng, D. Remaining useful life prediction of lithium-ion batteries using a hybrid model. *Energy* **2022**, *248*, 123622. [CrossRef]
42. Hussien, S.A.; BaQais, A.; Al-Gabalawy, M. Estimation of the residual useful life of EV batteries using advanced hybrid learning tools. *Electr. Eng.* **2023**, *106*, 2651–2677. [CrossRef]
43. Duan, W.; Song, S.; Xiao, F.; Chen, Y.; Peng, S.; Song, C. Battery SOH estimation and RUL prediction framework based on variable forgetting factor online sequential extreme learning machine and particle filter. *J. Energy Storage* **2023**, *65*, 107322. [CrossRef]
44. Chen, L.; Zhang, Y.; Zheng, Y.; Li, X.; Zheng, X. Remaining useful life prediction of lithium-ion battery with optimal input sequence selection and error compensation. *Neurocomputing* **2020**, *414*, 245–254. [CrossRef]
45. Cheng, G.; Wang, X.; He, Y. Remaining useful life and state of health prediction for lithium batteries based on empirical mode decomposition and a long and short memory neural network. *Energy* **2021**, *232*, 121022. [CrossRef]
46. Chang, Y.; Fang, H.; Zhang, Y. A new hybrid method for the prediction of the remaining useful life of a lithium-ion battery. *Appl. Energy* **2017**, *206*, 1564–1578. [CrossRef]
47. Wang, Z.; Liu, Y.; Wang, F.; Wang, H.; Su, M. Capacity and remaining useful life prediction for lithium-ion batteries based on sequence decomposition and a deep-learning network. *J. Energy Storage* **2023**, *72*, 108085. [CrossRef]
48. Chen, Z.; Chen, L.; Shen, W.; Xu, K. Remaining Useful Life Prediction of Lithium-Ion Battery via a Sequence Decomposition and Deep Learning Integrated Approach. *IEEE Trans. Veh. Technol.* **2022**, *71*, 1466–1479. [CrossRef]
49. Nguyen, C.D.; Bae, S.J. Equivalent circuit simulated deep network architecture and transfer learning for remaining useful life prediction of lithium-ion batteries. *J. Energy Storage* **2023**, *71*, 108042. [CrossRef]
50. Jafari, S.; Byun, Y.C. Accurate remaining useful life estimation of lithium-ion batteries in electric vehicles based on a measurable feature-based approach with explainable AI. *J. Supercomput.* **2023**, *80*, 4707–4732. [CrossRef]
51. Wei, Y.; Wu, D. Prediction of state of health and remaining useful life of lithium-ion battery using graph convolutional network with dual attention mechanisms. *Reliab. Eng. Syst. Saf.* **2023**, *230*, 108947. [CrossRef]
52. Zhang, Y.; Zhao, M. Cloud-based in-situ battery life prediction and classification using machine learning. *Energy Storage Mater.* **2023**, *57*, 346–359. [CrossRef]

53. Wei, M.; Ye, M.; Zhang, C.; Li, Y.; Zhang, J.; Wang, Q. A multi-scale learning approach for remaining useful life prediction of lithium-ion batteries based on variational mode decomposition and Monte Carlo sampling. *Energy* **2023**, *283*, 129086. [CrossRef]
54. Wang, S.; Ma, H.; Zhang, Y.; Li, S.; He, W. Remaining useful life prediction method of lithium-ion batteries is based on variational modal decomposition and deep learning integrated approach. *Energy* **2023**, *282*, 128984. [CrossRef]
55. Ding, P.; Liu, X.; Li, H.; Huang, Z.; Zhang, K.; Shao, L.; Abedinia, O. Useful life prediction based on wavelet packet decomposition and two-dimensional convolutional neural network for lithium-ion batteries. *Renew. Sustain. Energy Rev.* **2021**, *148*, 111287. [CrossRef]
56. Chen, D.; Zhang, W.; Zhang, C.; Sun, B.; Cong, X.W.; Wei, S.; Jiang, J. A novel deep learning-based life prediction method for lithium-ion batteries with strong generalization capability under multiple cycle profiles. *Appl. Energy* **2022**, *327*, 120114. [CrossRef]
57. Zhou, Y.; Wang, S.; Xie, Y.; Shen, X.; Fernandez, C. Remaining useful life prediction and state of health diagnosis for lithium-ion batteries based on improved grey wolf optimization algorithm-deep extreme learning machine algorithm. *Energy* **2023**, *285*, 128761. [CrossRef]
58. Yan, L.; Peng, J.; Gao, D.; Wu, Y.; Liu, Y.; Li, H.; Liu, W.; Huang, Z. A hybrid method with cascaded structure for early-stage remaining useful life prediction of lithium-ion battery. *Energy* **2022**, *243*, 123038. [CrossRef]
59. Zhang, C.; Wang, H.; Wu, L. Life prediction model for lithium-ion battery considering fast-charging protocol. *Energy* **2023**, *263*, 126109. [CrossRef]
60. Guo, Y.; Wang, Y.; Ding, P.; Huang, K. Future degradation trajectory prediction of lithium-ion battery based on a three-step similarity evaluation criterion for battery selection and transfer learning. *J. Energy Storage* **2023**, *72*, 108763. [CrossRef]
61. Hong, J.; Lee, D.; Jeong, E.R.; Yi, Y. Towards the swift prediction of the remaining useful life of lithium-ion batteries with end-to-end deep learning. *Appl. Energy* **2020**, *278*, 115646. [CrossRef]
62. Zhang, F.; Shen, Z.; Xu, M.; Xie, Q.; Fu, Q.; Ma, R. Remaining useful life prediction of lithium-ion batteries based on TCN-DCN fusion model combined with IRRS filtering. *J. Energy Storage* **2023**, *72*, 108586. [CrossRef]
63. Fei, Z.; Zhang, Z.; Yang, F.; Tsui, K.L. A deep attention-assisted and memory-augmented temporal convolutional network based model for rapid lithium-ion battery remaining useful life predictions with limited data. *J. Energy Storage* **2023**, *62*, 106903. [CrossRef]
64. Ardeshiri, R.R.; Liu, M.; Ma, C. Multivariate stacked bidirectional long short term memory for lithium-ion battery health management. *Reliab. Eng. Syst. Saf.* **2022**, *224*, 108481. [CrossRef]
65. Lipu, M.H.; Hannan, M.A.; Hussain, A.; Hoque, M.; Ker, P.J.; Saad, M.H.M.; Ayob, A. A review of state of health and remaining useful life estimation methods for lithium-ion battery in electric vehicles: Challenges and recommendations. *J. Clean. Prod.* **2018**, *205*, 115–133. [CrossRef]
66. Shah, A.; Shah, K.; Shah, C.; Shah, M. State of charge, remaining useful life and knee point estimation based on artificial intelligence and Machine learning in lithium-ion EV batteries: A comprehensive review. *Renew. Energy Focus* **2022**, *42*, 146–164. [CrossRef]
67. Lipu, M.H.; Ansari, S.; Miah, M.S.; Meraj, S.T.; Hasan, K.; Shihavuddin, A.; Hannan, M.; Muttaqi, K.M.; Hussain, A. Deep learning enabled state of charge, state of health and remaining useful life estimation for smart battery management system: Methods, implementations, issues and prospects. *J. Energy Storage* **2022**, *55*, 105752. [CrossRef]
68. Hasib, S.A.; Islam, S.; Chakrabortty, R.K.; Ryan, M.J.; Saha, D.K.; Ahamed, M.H.; Moyeen, S.I.; Das, S.K.; Ali, M.F.; Islam, M.R.; et al. A comprehensive review of available battery datasets, RUL prediction approaches, and advanced battery management. *IEEE Access* **2021**, *9*, 86166–86193. [CrossRef]