Future Batteries 8 (2025) 100124

Contents lists available at ScienceDirect

Future Batteries

journal homepage: www.sciencedirect.com/journal/future-batteries

Predicting remaining useful life of lithium-ion batteries: A review of degradation mechanisms and open-source data availability

Kishan Patel [a,b,*], Vaidehi Gosala [a], Merten Stender [b], Moritz Braun [a], Sören Ehlers [a]

[a] German Aerospace Center (DLR), Institute of Maritime Energy Systems, Geesthacht, Germany
[b] Technische Universität Berlin, Berlin, Germany

ARTICLE INFO

Keywords:
Lithium-ion battery
State of health
Remaining useful life
Open-source datasets

ABSTRACT

Lithium-ion batteries are vital for large-scale industries, especially in transport and renewable energy applications, due to their high energy density, extended cycle life, and low self-discharge rate as compared to other battery types. With increasing demand for sustainable and energy-efficient solutions, it is critical to study, understand, and improve the performance of batteries. Monitoring degradation is important, but acquiring real-time sensor data over lengthy periods is challenging due to the longer life cycles of batteries. This makes data collection costly and time-consuming. Cell-based open-source datasets provide a viable alternative, allowing researchers to estimate the degradation of battery cells without the requirement for constant, real-time testing. Furthermore, estimating degradation factors is crucial for forecasting Remaining Useful Life and extending battery lifespan. Methods such as adaptive filtering techniques, machine learning approaches, etc., have demonstrated reliable solutions in simulating battery degradation. This paper reviews the battery cell degradation mechanisms, followed by the prediction of battery health parameters and relevant degradation modelling approaches for individual cells. The purpose of this review is to provide a structured analysis of how different modelling methods capture degradation behavior, to identify their strengths and limitations, and to clarify how they can be applied for battery health prediction. It also highlights the importance of datasets required for developing predictive models and summarizes open-source datasets based on the chemistry, cycling process, and their key features.

1. Introduction

The majority of modern electronics are powered by batteries, and with the rapid increase in the generation of renewable energy, storage has become increasingly critical [1]. The energy market is undergoing significant shifts across various sectors as the production of electricity from variable sources continues to rise quickly. The rapidly growing demand for energy storage is due to problems in the energy markets of developing countries and shifts in transportation methodologies. [2].

Lithium-ion batteries (LIBs) have emerged as an essential component of modern energy storage systems, transforming many industries and paving the way for a more sustainable future. LIBs have gained popularity over other battery technologies in a variety of applications due to their extended lifespan, high potential density, lightweight design, and low self-discharge. Such applications include airplanes, electric vehicles (EVs), satellites, maritime systems, cellphones, laptops, and other electronic devices [3]. The significance of LIBs resides not only in their immediate applications but also in their ability to force a major transition towards renewable energy sources. As the world moves away from fossil fuels, LIBs will play an important role in storing energy provided by intermittent renewable sources, such as solar and wind. In the field of transportation, LIBs have emerged as a game-changer. Their use in electric vehicles has resulted in much lower greenhouse gas emissions, providing a greener and more environmentally friendly alternative to conventional internal combustion engines. Additionally, LIBs are increasingly offering fast charging and longer driving ranges, solving critical issues about EV practicality and convenience.

In addition, LIBs show potential as a vital component of microgrid systems, allowing distant settlements and islands to switch to renewable energy sources while reducing their reliance on expensive and polluting diesel generators [4]. LIBs enable a more stable and resilient energy supply by storing extra energy during periods of low demand and discharging it when needed, thereby improving energy security and supporting sustainable development. LIBs, which are prevalent in modern

* Corresponding author at: German Aerospace Center (DLR) – Institute of Maritime Energy Systems, Geesthacht, Germany.
E-mail address: kishan.patel@dlr.de (K. Patel).

https://doi.org/10.1016/j.fub.2025.100124
Received 30 September 2025; Received in revised form 5 November 2025; Accepted 24 November 2025
Available online 25 November 2025
2950-2640/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

technology, degrade over time, reducing their performance and durability. This degradation occurs via a variety of chemical and mechanical processes. Chemically, the electrolyte can degrade through reactions with ions and the electrons at the anode surface, leading to the formation of the SEI layer [5]. Although the SEI layer is crucial for protecting the electrolyte, its progressive growth during aging can raise the internal resistance and reduce the capacity retention properties. Moreover, during the charging phase, lithium plating can cause dendrite formation, resulting in short circuits and capacity loss, and making the cathode material structurally weak, which might result in decreased capacity and voltage instability [6]. These inadvertent side reactions between electrodes and electrolytes, such as SEI layer formation and lithium plating, also contribute to capacity loss. Due to the repetitive process of charging and discharging, the electrode material is subjected to constant expansion and compression, which results in mechanical stresses, crack formation, and delamination [7]. These mechanical stresses weaken the component and result in low electrical contact, which in turn accelerates the capacity degradation of the battery. Apart from these mechanical causes, temperature and humidity also accelerate degradation processes, creating safety issues and limiting cycle life [8]. Extensive research exists on the impact of temperature; comparatively less focus has been given to humidity and saline environments, despite their influence on battery performance and safety [8]. Understanding and research on these degradation mechanisms is critical to improving battery performance and the safety aspect. Furthermore, modelling the degradation mechanisms of LIBs is important for predicting the Remaining Useful Life (RUL) [9]. By creating mathematical models and algorithms based on state of health (SOH) estimations from the Battery Management System (BMS), we can predict battery life, apply proactive maintenance techniques, and provide condition-based maintenance. These predictive capabilities assure optimum asset utilization while minimizing downtime. Furthermore, understanding SOH and the underlying degradation mechanisms would enable the design of safety measures and thermal management systems to prevent catastrophic failures and assure safe battery use. In this context, extensive studies exist that focus on the management of charging and discharging, assessment of RUL, and assessment of the performance degradation of LIBs [9–11].

Lu et al. [12] reviewed the main issues of battery degradation. They studied the degradation over the entire life cycle of the battery, starting with reviewing batteries' internal aging mechanisms and also focused on factors such as design, production method, and applications. Kabir et al. [5] did an extensive review emphasizing degradation causes as well as mechanisms. They concluded the research, mentioning that the improved understanding of the mechanisms was vital in developing aging models and, therefore, helping the predictions of the cell EOL more accurately. Elmahallawy et al. [13] gave a brief review of the EV batteries' health status from the point of view of operational safety. They studied various battery modelling techniques, such as physics models and data-driven models, where they also focused on the ML algorithms. Jin et al. [14] gave an overview of ML methods that could be used for the RUL prediction of the LIBs. They provided a summary and classification of different RUL estimation methods that had been proposed in the last few years. After putting forward the methods, they discussed some ML models and tried to compare the accuracy of the methods against each other. They concluded their study by claiming that a Deep Neural Network (DNN) is more viable for the RUL estimations due to the strong learning ability and understanding of the datasets. These studies lacked fundamental information about degradation mechanisms and modelling methods, and focused on lab-scale operating conditions. In the last decade, machine learning, as well as hybrid data-driven models and new cell chemistries, were implemented to improve SOH and improve lifetime predictions, and there was a need to address them. Focusing on and updating the information on new cell chemistries, their dataset availability, and the feasibility of implementing newer degradation models is important.

This review paper provides an overview of LIB degradation mechanisms, their impact on cell performance, and why a few battery health states, such as State of Charge (SOC), SOH, and RUL, need to be studied. The paper also explains different modelling techniques that can be implemented to predict the health states, and dives deep into data-driven modelling techniques, including adaptive filters, machine learning, and models such as Bayesian and Time series approaches, highlighting how they perform and capture different degradation effects, and help in predicting health states of the battery. Given the difficulties, such as lack of experimental facilities, safety concerns, and costs of long-term testing, this review also emphasizes the availability of open-source battery cell datasets. A few widely used degradation datasets from various institutes for modelling purposes are described in detail. Other datasets are summarized in a table stating Important parameters that would enable efficient selection for modelling purposes. The summary table covers diverse cell types, chemistries, and cycling processes, and provides useful information for developing and validating data-driven predictive models.

The paper is structured into 4 main chapters, starting with explaining degradation mechanisms and factors that influence battery life, followed by the significance of battery health parameters and the respective modeling approaches. Lastly, it gives a brief description of available datasets based on the cell cycling process and the key features, which help in easing the selection process for developing the predictive models, providing a clear flow from understanding degradation to selecting methods and datasets for a reliable battery health prediction model.

2. Battery degradation mechanism

At the battery system level, an understanding of degradation mechanisms and appropriate modelling methods is critical, particularly for predicting performance under varying operating conditions and at different stages of life, as well as for ensuring safe operations. The ability to precisely anticipate battery end of life (EOL) or RUL allows the dangers of battery runaway to be reduced [15]. Understanding battery degradation is crucial for cost-effective decarbonization of both transportation and energy systems [16]

LIBs' storage capacity decreases with time and use, while their internal resistance increases as a result of a various degradation mechanisms such as SEI layer growth on the anode, while the internal resistance increases due to lithium plating and active material loss. Three major factors affect degradation: temperature and the load profile as external stressors, and SOC as an internal operating variable that reflects the combined effect of load history and the initial conditions

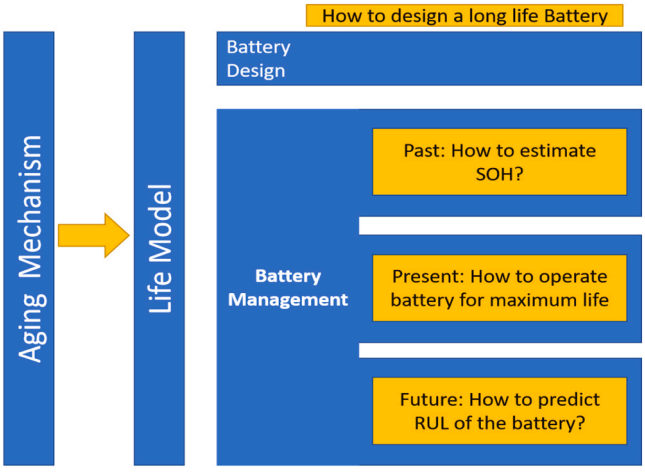

Fig. 1. Battery design and management issues based on ageing mechanisms and life.

[17]. As shown in Fig. 1, based on [18], battery aging and the influence of battery degradation should not only be considered during battery design but also during operation from the perspective of battery management. These models are crucial for estimating battery health influenced by history, optimizing working conditions in the present, and predicting future performance.

Given the battery's nonlinear fading characteristics, the usual extrapolation method cannot reliably anticipate its remaining life. The most apparent signs of battery degradation are capacity and power fade [19,20].

In battery system analyses, the capacity and power performance are essential and relevant factors, which must be correctly predicted by the BMS along with the SOH. Battery life may be divided into two parts: calendar life and cycle life. Calendar life refers to battery degradation caused by storage without cycling, whereas cycle life considers the battery's degradation caused by charge and discharge cycles. Most EV batteries exhibit nonlinear aging characteristics [7,21].

The primary parameters affecting battery life include [6]:

- High temperature, which accelerates internal side reactions.
- Low temperatures cause a rapid reduction of metal ions. They also lead to lithium deposition and can destroy the crystal structure of the active material.
- High SOC or overcharge that can lead to electrolyte decomposition, trigger side reactions between the electrolyte and the cathode, and cause lithium-ion deposition. These effects compromise battery integrity and performance, reducing efficiency and increasing safety risks.
- Low SOC or Over Discharge.
- The anode copper current collector is susceptible to corrosion, which can lead to significant issues in lithium-ion batteries. Corrosion may result in the collapse of the crystal structure of the active material, adversely affecting battery performance and longevity.
- High Charge and Discharge Rate.

Most commercial LIBs have an operating temperature range of 15–35°C and the rate of adverse reactions increases with increasing temperature [7]. Furthermore, if the battery surpasses a specific temperature, it may initiate self-heating, resulting in battery thermal runaway. Moreover, material embrittlement at low temperatures may affect battery life. As a result, ensuring that the battery operates within a proper temperature interval is the key to improving battery life [6].

Battery SOC also has a substantial impact on battery life. A higher value of SOC indicates a higher terminal voltage, implying a lower anode potential and a higher cathode potential. For the graphite anode with a lower potential, the side reaction rate, such as SEI thickening, will be higher, resulting in a faster aging rate, and in the event of aberrant charging, such as overcharging or low temperature charging, the anode potential may be too low and reach the lithium deposition potential, causing the side reaction of lithium deposition to accelerate battery aging [5]. Meanwhile, the cathode with a greater potential would undergo electrolyte oxidation and breakdown. Lower SOC indicates a higher anode potential and a lower cathode potential, which generally improves battery life. However, if the battery's SOC is too low, corrosion of the anode copper current collector and cathode active material structure would significantly reduce battery life [22]. Fig. 2 shows how different parameters change from the first to the last test cycle during the discharging phase from the NASA battery dataset (B0005). As the cycle number increases, the temperature of the battery rises rapidly and in a shorter time period, indicating increased internal resistance and heat generation, which contribute to performance loss and signs of degradation of electrodes and electrolytes.

The battery current has a clear impact on battery life. On the one hand, the current running through the battery generates Joule heat, influencing the battery temperature. Large charge and discharge rates, in particular, can produce a significant temperature increase, reducing battery life [24]. Battery degradation is an important topic to investigate since it has a significant and considerable impact on battery performance, lifespan, and safety. Understanding degradation methods enables the design of more lasting and efficient energy storage systems and also aids in predicting battery life and maintenance procedures, which would result in greater dependability and cost-effectiveness. Lastly, recognizing degradation pathways is critical for minimizing the environmental impact of batteries. Different degradation pathways create distinct reaction products, such as lithium salts from electrolyte breakdown or transition metal oxides from cathode degradation, which have

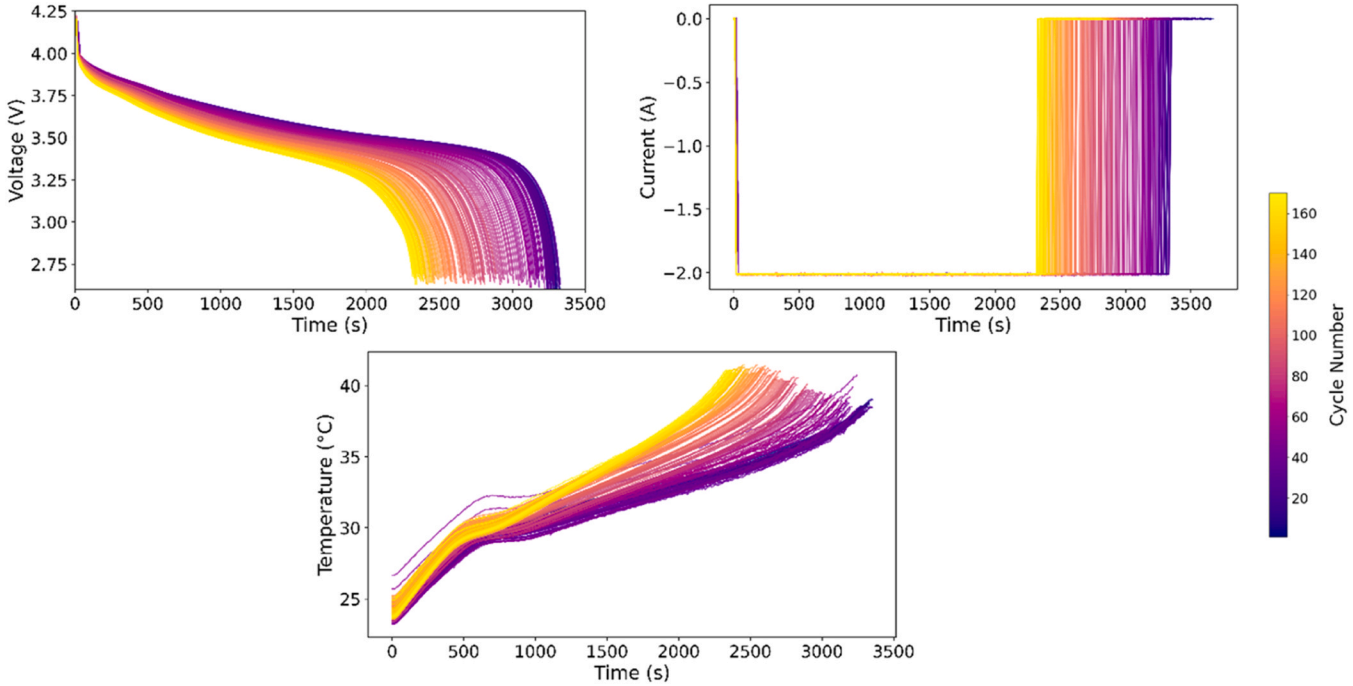

Fig. 2. Comparison of discharge voltage, current, and temperature loss of the first and last cycle of a battery cell from the NASA dataset (B0005) [23].

implications for disposal and recycling techniques, as well as EOL safety.

The above paragraph explains the internal degradation mechanisms that control LIB aging. These processes, both physical and chemical can be observed through measurable datasets based on electrochemical responses. A direct connection between degradation modes and their diagnostic signatures are fundamental for understanding the degradation mechanisms via the datasets. Table 1 below summarizes these connections by providing a systematic mapping.

3. Prediction and modelling of battery health parameters

In the process of studying degradation, understanding battery parameters is crucial. These parameters facilitate understanding and streamlining of the degradation study using predefined terminologies that can be evaluated based on sensor data. A battery is a non-linear system with numerous complications; therefore, the BMS relies on a variety of parameters. Precision monitoring and estimation for these BMS parameters are computationally intensive. Unquestionably, for advanced and futuristic BMS, exact monitoring and prediction are essential for some key terminology, which includes SOC, SOH and RUL. SOC fluctuates rapidly with LIB usage and can be accurately determined through short-term electrical measurements, but it provides no information about how the battery's capacity, power capability, or safety margins have evolved. In contrast, SOH reflects the cumulative effects of aging mechanisms such as loss of active material, lithium inventory depletion, and impedance growth, which directly dictate both performance sustainability and RUL. Consequently, while SOC helps manage daily operations, SOH determines the operational lifespan and safety envelope of the system. For this reason, SOH estimation is considered more critical in advanced BMS and is thus explained in more detail in this study.

3.1. SOC estimation

SOC is a fundamental parameter in BMS that represents the real-time charge level of the cell. It indicates how much energy remains available for use. Accurate and reliable SOC estimation is vital for safe operations and extending the lifespan of LIBs. SOC is commonly defined as the ratio of remaining charge to its nominal capacity and expressed as a percentage [27].

$$SOC(t) = \left(\frac{Q_{rem(t)}}{Q_n}\right) \times 100\% \quad \text{Eq. 1}$$

where $Q_{rem(t)}$ is the remaining charge (Ah) at time t and Q_n is the nominal or rated capacity of the cell.

Although in practical applications, SOC is generally calculated by integrating the current over time as shown in the coulomb-counting equation [28].

$$SOC(t) = SOC(t_0) - \left(\frac{1}{C_n}\right)\int_{t_0}^{t} I(\tau)d\tau \quad \text{Eq. 2}$$

where $SOC(t)$ is the initial SOC, $I(\tau)$ is the current and C_n is the nominal capacity.

3.1.1. Methods used to model SOC

The most-used methods for estimating the SOC can be classified as,

Coulomb counting (CC):

CC also known as ampere hour counting, is widely used to estimate SOC and is calculated from Eq. 2 The reason for its wide adoption is due to its simplicity, low computational demand and compatibility with real-time embedment with BMS. Mohammadi et al. [29] proposed an improved version of CC algorithm where they incorporated uncertainty propagation for the current readings and the integration error. They used a 12 V 100Ah lithium pack and tested it over a 10-year projection. The model achieved a maximum deviation below 0.3 % demonstrating that uncertainty modelling can improve the CC model performance. Zhu et al. [30] developed a non-destructive differentiation-based CC model for the NMC 18650 cell. Their method combined OCV with empirical tuning factors for parameters such as temperature, discharge rate, etc. The experiments gave a maximum error of approximately. 3.5 % over the traditional OCV-CC models. Lee et al. [31] upgraded the framework through an Enhanced CC model that updates the coulombic efficiency after each charge-discharge cycle thereby predicting the SOC at each step. They used a 3 kW energy storage module and achieved MAE of 0.05 % and validated the results with real-time BMS. Huang et al. [32] proposed a fault-diagnosis layer for the CC algorithm to detect the initial SOC values. They used a 50Ah pouch cell for the research. Lastly, Movassagh et al. [33] performed a theoretical analysis of CC error propagation. They derived the parameters from the statistical models of current noise and capacity.

Electrochemical Models:

Electrochemical models offer a physics-based route for predicting the internal states of LIBs. Hosseininasab et al. [34] developed a fractional order reduced model derived from the Pseudo 2D model to estimate the cell resistance and achieved high accuracy with minimal parameters and computational cost. They also performed experimental validation on an aged NMC battery with a nominal capacity of 27 Ah under dynamic current and temperature profiles. Xu et al. [35] proposed a hybrid minimalist electrochemical–equivalent-circuit framework. The circuit model only handled dynamic voltage response while the minimalist electrochemical model quantified lithium inventory loss. They used NASA's battery data and achieved errors of about 2 %. Lastly, Wu et al. [36] further simplified the Pseudo 2D model using Pade approximation and averaging the volume to design a proportional integral and differential observer. The model was capable of jointly estimating SOC and anode potential whilst maintaining SOC error below 2 %.

A summary Table 2 comparing CC models and Electrochemical

Table 1
Mapping of major internal degradation mechanisms to their corresponding externally measurable diagnostic signatures [15,24–26].

Internal Degradation Mechanisms	Cause	External noticeable signature	Measurement method
Loss of lithium inventory	Continuous electrolyte reduction and isolation of Li+ ions	Shift of peaks in Incremental Current Analysis (ICA)/Differential Voltage Analysis (DVA) because of higher voltage as well as monotonic capacity reduction	ICA, DVA, Open Circuit Voltage (OCV) curve
Loss of active material	Electrode delamination, fracture under cycling	Reduction of ICA/DVA peaks and formation of new low-magnitude peaks	ICA/DVA, X-ray imaging
SEI layer growth	Electrolyte reduction at the anode during storage and cycling	Increased overpotential and polarization, columbic efficiency reduction	Electrochemical Impedance Spectroscopy (EIS), Voltage profile analysis
Lithium plating	Li metal deposition on graphite or overcharging	Formation of a new ICA peak, self-heating of the cell	ICA, DVA, OCV, EIS
Electrolyte decomposition	Oxidation under high potential difference or temperature change	Increase in internal pressure of the cell, impedance increase	Pressure mapping, EIS

Table 2
Summary Table comparing Coulomb Counting and Electrochemical Model.

Aspect	Coulomb-Counting (CC) Methods	Electrochemical-Model (EM) Methods
Fundamental Principle	Integrates measured current over time to estimate charge variation; SOC is derived from current, time, and nominal capacity.	Solves simplified electrochemical equations (from P2D or SPM) describing Li-ion transport, reaction kinetics, and potential distribution.
Typical Datasets	Structured laboratory or BMS test profiles with precise current and voltage measurements (e.g., 100 Ah pack, 18650 cells, NEDC cycles).	Physics-based experiments or simulation-validated data (NASA RW profiles, dynamic cycling, COMSOL co-simulation).
Model Complexity	Low; algebraic integration or simple correction factors (temperature, aging, coulombic efficiency).	Moderate to high; reduced-order fractional, minimalist, or simplified P2D models requiring parameter identification.
Computational Demand	Very low; ideal for embedded real-time BMS.	Moderate; feasible only after model-order reduction or hybridization.
Accuracy based on cited references	0.3 – 3 % SOC error; SOH MAE ≈ 0.06 % in enhanced forms.	Typically, < 2 % SOC error; SOH or internal-state estimation errors < 0.1 %.
Strengths	Simple implementation, low sensor and compute requirements, easily deployed in existing BMS hardware.	Physically interpretable, captures degradation and internal electrochemical dynamics, enables safety-oriented state observation.
Limitations	Sensitive to current-sensor bias, cumulative drift, and capacity uncertainty; poor under irregular load or missing calibration.	Requires accurate parameterization, often assumes isothermal or uniform conditions; higher computational cost.

models is presented below.

3.2. SOH and RUL estimation

SOH and RUL are inherently correlated as SOH quantifies the overall condition of the battery relative to its initial and nominal state, while the RUL projects the time or number of cycles remaining until it reaches its end-of-life threshold.

SOH is the ability of the battery to store and deliver energy compared to the performance when the battery was new [37]. SOH can be expressed as the ratio of current usable capacity to nominal capacity at the beginning of life and is given by,

$$SOH(t) = \left(\frac{C_{actual(t)}}{C_{nominal}}\right) \times 100\% \quad \text{Eq. 3}$$

Where, $C_{actual(t)}$ is the measured capacity at time t and $C_{nominal}$ is the rated capacity.

The prognostic element of Prognostic Health Management (PHM) is commonly defined as estimating a system's RUL based on the information provided. Estimating the RUL has been regarded as one of the most important components of PHM along with SOH prediction. RUL represents the expected number of cycles or operating time remaining before the battery reaches its EOL criterion (Sbarufatti). RUL can be predicted using,

$$RUL(t) = f(SOH(t), operating\ conditions) \quad \text{Eq. 4}$$

The calculation of the RUL of components, subsystems, and systems has typically relied on data analysis that includes failure mode signatures gathered throughout a system's life. Reliability-based RUL calculation is based on analyzing the asset's failure characteristics and determining what can be done to mitigate the impact of the failure. Statistical analysis frequently ignores or assumes only one failure mechanism, which is not the case in many complicated engineering systems, leading to component replacement at a predetermined time. In contrast, a prognostic approach seeks to anticipate failure based on individual component state estimation, using either failure physics, data, or fusion degradation models.

The SOH and RUL prognostics methods are classified into four categories based on their essential principles and procedures: physics model-based techniques, statistics model-based techniques, AI techniques, and hybrid techniques. Fig. 3 depicts the several categories of prediction methodologies. In general, the prognostic approach consists of three steps: feature selection, health evaluation, detection, and prediction triggering [38]. Several types of sensors can be used to collect data from the monitored device. Using this data, specific criteria can be used to determine the real fitness status. A degradation replica is typically linked to failure time by combining historical data and the failure occurrence. The degradation process and RUL are strongly related. The first step in RUL prediction is to study the degradation procedure using SOH.

3.2.1. Data-driven methods used to model SOH and RUL

Data-driven models are advantageous for RUL and SOH estimations due to their adaptability and usage of real-world sensor data, allowing them to capture patterns and trends from the datasets without needing detailed domain-specific parameter values. Unlike physics-based models, they do not need extensive strong knowledge of the entire system, making them comparatively easier to implement. They also possess the ability to continuously learn from new sensor data, allowing the models to adapt to varying conditions. This ability improves their accuracy over time. However, they require an extensive number of initial datasets for training, but the subsequent benefits in terms of reliable predictions and ease of implementation make them a better option for practical understanding and industries. A few commonly used data-driven RUL estimation along with SOH estimation methods are shown in Fig. 4 below and are explained in the next subchapter.

3.2.1.1. Adaptive filter methods. Parameter estimation approaches based on similar circuit models were utilized as well to measure Li-ion battery degradation. As a classic technique, the adaptive filter algorithm is often used to predict a battery's SOH and RUL. The following are some adaptive filter approaches used in the RUL prognosis of Li-ion cells.

A.1 Kalman filter (KF): The KF is an adaptive method that is broadly used for parameter estimation. It employs a sequence of measurements taken over time to estimate the more exact output variables. When

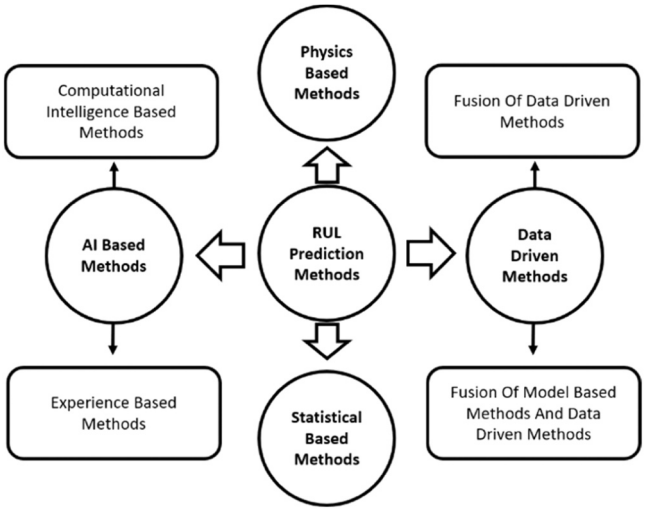

Fig. 3. RUL Prediction Approaches.

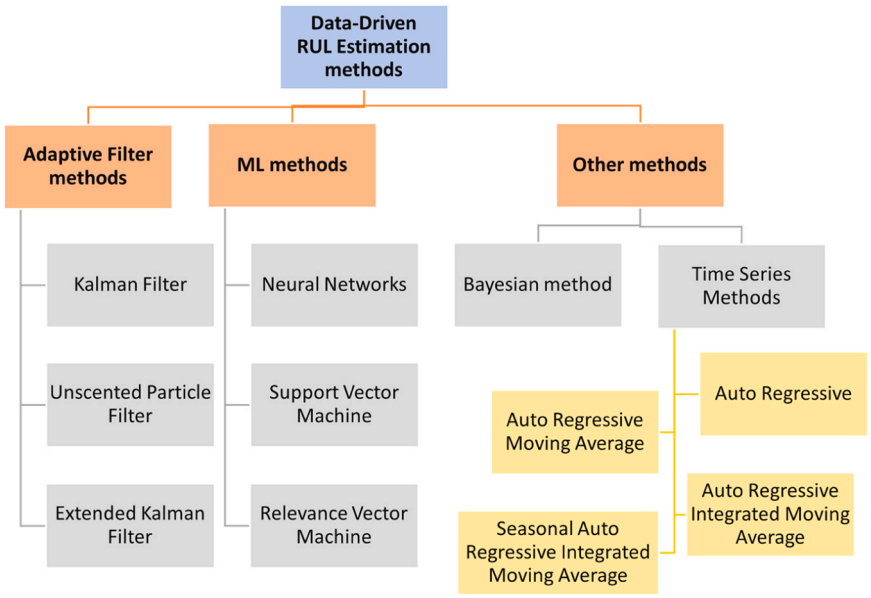

Fig. 4. Data-Driven RUL Estimation Methods.

utilizing a KF, two procedures must be followed. First, a prediction state is required, during which the filter estimates the current output variable. In the second stage, the estimation is revised to produce a more accurate result, increasing the estimation's confidence. It necessitates the use of recursive equations, which may be used when the system's discrete model is known in state space form. This happens because the current state of the system is the result of the effect of all the inputs in already past states. The general equations are as follows.

Firstly, the method starts with a time update state known as the prediction state equation, followed by Kalman gain calculation, and finally the updated state estimation.

$$\boldsymbol{x}_{pred} = \boldsymbol{F} \bullet \boldsymbol{x}_{prev} + \boldsymbol{B} \bullet \boldsymbol{u} \quad \text{Eq. 5}$$

$$K = P_{\text{pred}} \bullet H \bullet \left(H \bullet P_{\text{pred}} \bullet H' + R\right)^{-1} \quad \text{Eq. 6}$$

$$\boldsymbol{x}_{upd} = \boldsymbol{x}_{pred} + \boldsymbol{K} \bullet \left(\boldsymbol{z} - \boldsymbol{H} \bullet \boldsymbol{x}_{pred}\right) \quad \text{Eq. 7}$$

Here, x_{pred} represents the predicted state estimate, x_{prev} is the previous state estimate, F is the state transition model, B & u are the control input models, K is the Kalman gain factor, which determines the prediction adjustment, P_{pred} is the predicted covariance estimate, H shows the measurement model and finally z is the observation (measurement).

Ahwiadi et al. [39] proposed an enhanced Kalman filter framework for SOH and RUL prediction to overcome sample degeneracy and lack of measurements during prognosis. They integrated the model with an evolving fuzzy predictor to adapt to the posterior distribution and forecast degradation when data is not available anymore. They implemented this logic on the NASA dataset and achieved very low RMSE values. They concluded that KF models with add-ons improve both SOH and RUL estimations.

A.2 Extended Kalman Filter (EKF): The KF is only relevant to linear systems, which do not typically include battery models. Because of this constraint, modifications and extensions have been created within the KF. The EKF is a nonlinear variation of the KF. The EKF is commonly utilized in the development of battery models. Many publications utilize the EKF to estimate the SOC [28,40,41] and have shown several advantages, and hence prove to be a good model for battery parameter estimations. EKF models also demonstrated that the acquired results are quite accurate, the technique to implement is simpler compared to other nonlinear estimation methods, such as UKF or even NNs, allowing it to be performed in a real-world application with ease, and the produced model exhibits a linear association with the cell's dynamics. The filter is based on a basic concept consisting of a voltage source and internal resistance. Jiang et. al [42] offer an improved closed-loop estimator based on the EKF. The suggested model has been validated using experimental findings derived from various circumstances. It was proved that the augmented model reduces estimation error by roughly half when compared to an estimator that ignores the hysteresis effect. Meng et. al [27] conceptualized a methodology by combining the EIS Internal Impedance Approach with the EKF to estimate SOH values. The general equations for EKFs are as follows:

$$\boldsymbol{x}_{pred} = \boldsymbol{f}\left(\boldsymbol{x}_{prev}, \boldsymbol{u}\right) \quad \text{Eq. 8}$$

$$\boldsymbol{K} = \boldsymbol{P}_{pred} \bullet \boldsymbol{H}' \bullet \left(\boldsymbol{H} \bullet \boldsymbol{P}_{pred} \bullet \boldsymbol{H}' + \boldsymbol{R}\right)^{-1} \quad \text{Eq. 9}$$

$$\boldsymbol{x}_{upd} = \boldsymbol{x}_{pred} + \boldsymbol{K} \bullet \left(\boldsymbol{z} - \boldsymbol{h}\left(\boldsymbol{x}_{pred}\right)\right) \quad \text{Eq. 10}$$

Where, $f(x_{prev}, \quad u)$ is the nonlinear state transition function, H is the Jacobian matrix of the partial derivatives and $h(x_{pred})$ represents the nonlinear measurement function that relates the predicted state to the measurements.

A.3 Unscented Kalman Filter (UKF): The UKF is an algorithm that employs a sequence of observations over time to provide the most accurate results. It anticipates that the results from several unknown variables will be more exact than those based on a single measurement. It is also used to calculate the SOC, capacity, and internal resistance. Alexprabu et al. [43] propose a novel BMS combining a lossless charge-balancer, a MIMO-Bi-LSTM unit for per-cell SOH estimation, and a UK-ANFI network optimized by Grey Wolf Optimizer for simultaneous SOH and RUL prediction. They tested their model on an EV battery pack. The approach achieved SOH and RUL RMSEs of approximately 0.8 % and 1.8 % respectively, concluding that the integrated balancing and UKF prediction framework significantly enhances lifetime prognosis while lowering prediction latency. Zhu et al. [44] proposed a hybrid framework with a three-dimensional UKF to estimate the SOH of lithium-ion batteries by tracking SOC via UKF and internal resistance. The experimental validation demonstrated markedly improved convergence speed and accuracy with SOH estimation errors consistently below 3 % in cycling tests. This enables selective estimation of parameters, reducing computational effort and time.

The UKF model is represented as follows:

$$\boldsymbol{x}_{pred} = \sum \left(Wm_i \bullet \boldsymbol{X}_{pred_i} \right) \quad \text{Eq. 11}$$

$$\boldsymbol{P}_{pred} = \sum \left(Wc_i \bullet \left(\boldsymbol{X}_{pred_i} - \boldsymbol{x}_{pred} \right) \bullet \left(\boldsymbol{X}_{pred_i} - \boldsymbol{x}_{pred} \right)' + \boldsymbol{Q} \right) \quad \text{Eq. 12}$$

$$\boldsymbol{K} = \boldsymbol{P}_{xz} \bullet \boldsymbol{P}_{zz}^{-1} \quad \text{Eq. 13}$$

$$\boldsymbol{x}_{upd} = \boldsymbol{x}_{pred} + \boldsymbol{K} \bullet \left(\boldsymbol{z} - \boldsymbol{z}_{pred} \right) \quad \text{Eq. 14}$$

$$\boldsymbol{P}_{upd} = \boldsymbol{P}_{pred} - \boldsymbol{K} \bullet \boldsymbol{P}_{zz} \bullet \boldsymbol{K}' \quad \text{Eq. 15}$$

Where, X_{pred_i} is the Sigma points passed through the nonlinear state transition function, Wm_i and Wc_i is the Weights for the mean and covariance, respectively. P_{xz} Cross-covariance between state and measurement, P_{zz} is the measurement's covariance and z is the measurement.

In summary, the KF is a recursive algorithm that estimates a system's internal state using noisy observations. In battery RUL prediction, KF assumes a linear connection between state and measurements, making it appropriate for basic deterioration models. However, real-world battery depletion is frequently nonlinear, which reduces its accuracy. The EKF solves this problem by linearizing the nonlinear dynamics around the current estimate with a first-order Taylor expansion. This enables it to handle nonlinear battery models better than KF, although its performance suffers when the system displays severe nonlinearity. The UKF, on the other hand, employs a collection of deterministic sample points (sigma points) to better capture the system's nonlinearities while avoiding linearization. This improves UKF's accuracy and robustness for complicated, nonlinear battery deterioration models, resulting in better RUL prediction under these situations [45].

3.2.1.2. Machine learning methods. One of the primary advantages of data-driven models (DDMs), clubbed with Machine Learning approaches for Li-ion battery degradation model development and parameter estimation, is their ability to achieve high accuracy by learning battery behavior based on monitored data. Thus, they do not require battery chemical modeling or knowledge. DDMs are also used to simulate the interaction of battery health, performance, and environmental conditions during operation [46]. DDMs use historical data to estimate the SOH and try to model the relationship between degradation, health indicators, and SOH. These methods include ML approaches such as an artificial neural network (ANN), support vector machine (SVM), and Relevance Vector Machine (RVM), as well as other intelligent algorithms, to extrapolate the estimated SOH and map the relationship between battery degradation, health indicators, and battery SOH using a historical database [47]. The SOH is estimated using machine learning approaches with features sensitive to battery degradation. While all DDMs necessitate data collection and analysis during battery operation, ML approaches have the advantage of learning complex patterns in the data, eliminating the need for extensive predefined battery behavior tests and simulations in most cases, allowing for greater adaptability to different features and battery types.

Fig. 5 shows the workflow for battery degradation and SOH prediction models using ML techniques inspired by the work of Rauf et. al [48].

To forecast RUL, ML approaches frequently use estimated or measured SOH information, such as capacity values, as inputs [49]. An accurate and reliable machine learning-based approach to exact battery degradation modeling and RUL prediction is critical for advanced battery management [50]. The ultimate purpose of battery health management is to anticipate a battery's RUL and identify potential unforeseen circumstances caused by battery aging.

Neural networks (NNs) are one of the industry-leading machine learning approach that achieves high levels of accuracy. NNs are frequently utilized in self-learning and adaptation because they are not dependent on the electrochemical situations that occur within the battery. Neural networks are used to map the relationship between distinctive parameters and the lifetime of Li-ion battery degradation. NNs, including architectures such as Feed Forward (FF), Recurrent (R), and Convolutional Neural Networks (CNN), have a powerful algorithm that properly calculates SOH/SOC/RUL across a wide range of battery states, dynamic loads, and temperatures [51]. Their strength lies in modelling complex nonlinear relationships, adapting to diverse data, and effectively handling multiple input features to provide reliable predictions under varying operating conditions.

Recent studies [52–55] have investigated various forms of neural networks for RUL estimation. including FFNNs, RNNs, and CNNs. NNs have demonstrated a broad range of applications in battery degradation modeling and SOH estimation. Two forms of NNs, RNN and FFNN, have been primarily used for SOH estimation. FFNN and RNN are potential approaches for representing input-output correlations in battery aging data. The battery degradation process often consists of multiple cycles, and the degradation information between these cycles is highly dependent and interrelated. Thus, deriving these dependencies and correlations is essential to ensure reliable estimates.

RNNs are used to process sequential data in artificial intelligence applications and are one of the most promising methods for predicting battery health. The SOH/RUL estimation is based on a gradual battery degradation mechanism that uses dynamic battery data. As a result, employing an RNN to estimate SOH/RUL is an inherent technique. The RNN's key inputs are often the associative memory function, voltage, current, temperature, and time-delayed voltage and current. The RNNs are trained and tested using cell temperature, current, SOC fluctuation, previous time step capacity, and resistance. Teixeira et al. [56] developed a RNN framework using a Gated Recurrent Unit (GRU) model to predict SOH. They implemented the model on LCO cells of 5000mAh which were subjected to various cycling conditions ranging from 1 to 3 A until end of first life. The GRU achieved great prediction accuracy and was effective in capturing the nonlinear degradation pattern. They also proved that data data-driven RNN model can reliably estimate SOH even from a limited dataset.

Long short-term memory (LSTM) is a subset of RNN architecture designed to address the RNN's long-term reliance. Unlike FFNNs, LSTMs include feedback connections, while RNNs lack their input, forget, and

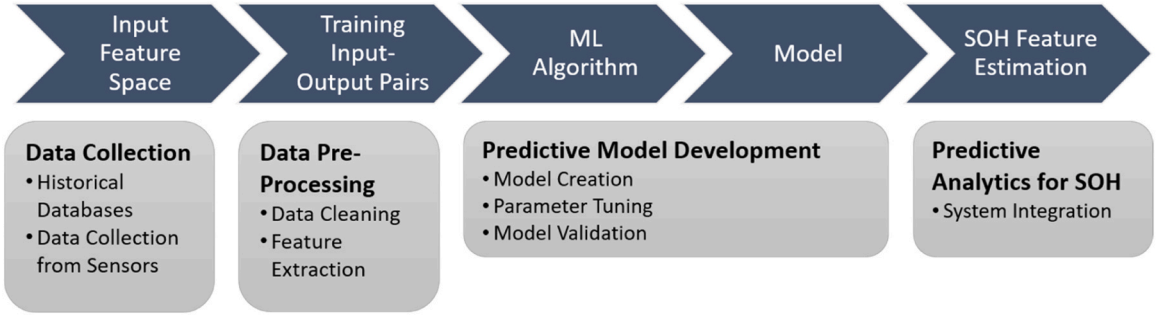

Fig. 5. Workflow for battery degradation and SOH prediction models using ML techniques.

output gates. LSTM has been used in a variety of studies to estimate and predict SOH. All Neural Network approaches have the advantage of being able to quickly adjust to nonlinear battery data, and they do not require physics-based models of batteries. They must, however, be trained over a significant number of cycles. Chinomona et al. [54] used RNN-LSTM to study battery degradation and compute RUL based on aging parameters collected from voltage, current, and temperature. Chen et al. [55] employed LSTM networks to create a prediction model that estimated RUL accurately with an RMSE of less than 4 %. Bharath et al. [57] presented a SOH estimation method using a cascaded LSTM-RNN model. The model was trained on multiple battery parameters, such as voltage, current, temperature, etc., which were collected under diverse real-time conditions. Unlike other models requiring continuous data input, their model estimates SOH once per discharge-charge cycle, which helps in reducing memory and computational load on the BMS.

3.2.1.3. Bayesian method. Bayesian linear regression is a statistical analysis method that treats a linear regression model's parameters as random variables with prior distributions rather than fixed but unknown values [58]. Bayesian linear regression comprises four key phases and parts: the likelihood function, posterior distribution, prediction, and prior distribution. The prior distribution is a previous estimate on the parameters recorded in a previous probability distribution that occurred before any data observation and is often expressed as,

$$p(\boldsymbol{\theta}|\boldsymbol{X},\ \boldsymbol{y}) \propto p(\boldsymbol{y} \mid \boldsymbol{X},\ \boldsymbol{\theta})\, p(\boldsymbol{\theta}) \qquad \text{Eq. 16}$$

Where, $p(\theta|\ X,y)$ is the posterior distribution of the parameters θ given the data X and responses y, $(y \mid X,\ \theta)$ is the likelihood function. The likelihood function represents the chance of observing the data given each of the provided parameters. To obtain posterior parameters, the Bayes' theorem updates the prior using the likelihood function. The posterior distribution captures the uncertainty in parameter estimates following data observation. The posterior predictive distribution is used to provide predictions for fresh input values by integrating the parameters across it.

In cases of continuous discharge, the Bayesian technique can be used to approximate the RUL in the absence of accurate operating parameter values. A study comparing several methodologies demonstrates that prognostic results are more accurate and robust than Support Vector Machine models. Dong et. al [59] proposes a probabilistic method for health prediction and degradation of battery modeling based on charging process data that uses a dynamic Bayesian network.

For a recognized data sample, Bayesian learning may be expressed as,

$$p(\boldsymbol{\theta} \mid D_{new}, D_{old}) \propto p(D_{new} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid D_{old}) \qquad \text{Eq. 17}$$

Where, D is the data and θ is the model parameter. In his latest work, Dong et al. [60] proposed a co-estimation approach for SOH and SOC prediction using Bayesian inference. This co-estimation approach included a Fractional-Order Model, a Bayesian Optimization algorithm, and a Gaussian-sum PF. They were able to effectively optimize the battery characteristics based on their physical relevance by conducting 20 tests before adopting the co-estimation technique to determine a suitable numerical range for the parameters. The dataset from CALCE [61] was utilized for their co-estimation scheme. Through their model, it was proved that the Bayesian Optimization algorithm-based parameter identification process had faster convergence speed, resulting in improved model flexibility and time efficiency for complex objective functions. The results from this method showed that monitoring the predicted SOC achieved an average RMSE of 1.84 %. In another research by Hu et al. [62], a temperature-dependent SOH estimating framework was established using a sparse Bayesian Predictive Modeling technique. An SVM method was also implemented to compare the computational complexity and performance. For model training, validation, and verification, experimental datasets from LIB cells evaluated at 10°C, 22 °C, and 35°C were used. The model showed that the predicted SOH values were accurate and had an average error of less than 1.2 % at each temperature value.

3.2.1.4. Time series methods. Time series data forecasting can be tricky because various statistical techniques and prediction methodologies may produce different results, making it difficult to choose which model to implement. In general, time series models may be categorized by looking at trends, seasonality, and the impact of outside factors [63]. The models discussed include ARMA, ARIMA, SARIMA, ARIMAX, and SARIMAX.

The autoregressive (AR) and moving average (MA) models are combined in the ARMA approach. The deterministic portion, known as the AR, is calculated by regressing the value from its prior p values. Comparably, MA describes the stationary series using residuals or disturbances. The moving average order q, the error term, affects the time series values. It is usually applied to data that exhibits neither seasonality nor a trend. Eq. 18 depicts the AR portion of the model with order p.

$$x_t = c + \Sigma(\phi_i \cdot x_{t-i}) + \epsilon_t,\ for\ i = 1\ to\ p \qquad \text{Eq. 18}$$

The value of the time series at t is denoted by x_t, the lag is represented by x_{t-i}, the autocorrelation coefficient of the time series data at point p is represented by ϕ_i, i, and c are constants. The residual, or the white noise, or the error, is denoted by ϵ_t.

Wei et al. [64] proposed a joint SOC-SOH estimation technique using an AR data-driven model. Based on the SOC-SOH conditions, they identified the battery parameters and extracted the features from the charging process of the battery. These features were used as the inputs and established a feature-SOH mapping, resulting in accurate predictions of SOH. Their model was later validated via the University of Michigan Battery Laboratory and McMaster University Hamilton datasets.

The ARIMA model is a method for time series forecasting that does not account for the effects of seasonality or external factors on the data. ARIMA or autoregressive integrated moving average models are loaded with time series data for estimating future points in the series or to characterize the data better than ARMA [65]. It is a method that allows both AR with parameter p and MA with parameter q. It explicitly includes a preprocessing step with parameter d (The degree of differencing to make the series stationary) in the formulation of the model, which indicates the number of transformations required to make the data stationary. In simple terms, an ARIMA model is just an ARMA model applied to a modified time series. The ARIMA model general form is expressed as,

$$\Delta^d x_t = c + \Sigma(\phi_i \cdot x_{t-i}) + \Sigma\left(\theta_j \cdot \epsilon_{t-j}\right) + \epsilon_t \qquad \text{Eq. 19}$$

here, $\Delta^d x_t$ is the differenced series of x_t (the integrated part) applied d times to make the series stationary, c is the intercept term, $\Sigma(\phi_i \cdot\ \ x_{t-i})$ represent the summation of past values (p) weighted by the coefficients ϕ_i which is the AR part and $\Sigma\left(\theta_j \cdot\ \ \epsilon_{t-j}\right)$ is the summation of q past error terms, weighted by the coefficients θ_j (the MA part). As the ARIMA model excludes external variables, sometimes depending on the dataset, it is difficult to assess the impact of external characteristics and their link to dependent features, such as capacity degradation. To address this, an extended version, ARIMAX can be implemented and is represented by Eq. 20.

$$\Delta^d x_t = c + \Sigma(\phi_i \cdot x_{t-i}) + \Sigma\left(\theta_j \cdot \epsilon_{t-j}\right) + \Sigma(\beta_k \cdot Z_{t-k}) + \epsilon_t \qquad \text{Eq. 20}$$

Where the additional terms compared to the previous equation are the summation of k past values of exogenous variables Z at time $t-k$ denoted by $\Sigma(\beta_k \cdot\ \ Z_{t-k})$. It also considers the effect of external factors weighted by the coefficients β_k.

Shen et al. [66] implemented the ARIMA model to predict the

lifecycle of second-use EV batteries. Their study utilized historical performance data of LIBs. This statistical approach was chosen due to its effectiveness in handling time series data. They tried different training-testing ratios for the dataset at two different temperatures (25°C and 50°C) and compared the predicted values with the experimental values using RMSE, and achieved an error of less than 3 %.

The Seasonal-ARIMA (SARIMA) model contains seasonal factors in its formulation [84]. It is more successful since it considers seasonal variations in data. SARIMA parameters include *p, d, and q*, as well as *P, D, Q,* and *S*. The nonseasonal parameters are *p, d*, and *q*, whereas the seasonal parameters are *P, Q, D*. SARIMA equations are given as:

$$\Delta^d x_t = c + \Sigma(\phi_i \cdot x_{t-i}) + \Sigma(\theta_j \cdot \epsilon_{t-j}) + \Sigma(\Phi_m \cdot x_{t-m}) + \Sigma(\Theta_m \cdot \epsilon_{t-m}) + \epsilon_t \quad \text{Eq. 21}$$

The two new terms, $\Sigma(\Phi_m \cdot x_{t-m})$ and $\Sigma(\Theta_m \cdot \epsilon_{t-m})$ represent the summation of *P* seasonal past values, *Q,* which is the seasonal past error at seasonal lags m weighted by coefficients Φ_m (AR Part) and Θ_m (Seasonal MA part) respectively. The SARIMA model combines seasonal and non-seasonal AR, differencing, and moving average components, allowing it to effectively describe time series data with seasonal patterns as well as underlying trends.

Lastly, by incorporating seasonality into SARIMAX *((p, q, d), (P, D, Q, S), r)* model, it provides a deeper understanding of predictable patterns or fluctuations that occur at specific intervals within a time series, into the ARIMAX *(p, q, d, r)* model. SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Inputs) models are widely applied in statistical analysis and demonstrate excellent forecasting performance, and are represented as,

$$\Delta^d x_t = c + \Sigma(\phi_i \cdot x_{t-i}) + \Sigma(\theta_j \cdot \epsilon_{t-j}) + \Sigma(\Phi_m \cdot x_{t-m}) + \Sigma(\Theta_m \cdot \epsilon_{t-m}) + \Sigma(\beta_k \cdot Z_{t-k}) + \epsilon_t \quad \text{Eq. 22}$$

Where the additional term $\Sigma(\beta_k \cdot Z_{t-k})$ is the summation of *k* past values of exogenous variables *Z* at time $t - k$, weighted by the coefficients β_k (the effect of external factors). Hu et al. [67] proposed a novel SARIMA prediction model that consisted of a periodic parameter optimization to fit the nonlinear characteristics of the dataset. It included maximum likelihood estimation and the Akaike information criterion to filter the parameters. To establish model accuracy, extensive sequence testing for tuning parameters was simulated with the help of autocorrelation and partial autocorrelation functions. The model was then tested under four operating conditions, including high and low charge-discharge rates. Their model achieved a maximum prediction error of 4.62 %, showing a 2.25 % improvement in RUL predictions.

To summarize, AR models evaluate time series data by regressing the present value against previous values, thereby capturing internal dependencies. The ARIMA model extends AR by using differencing to achieve stationarity, allowing for the modeling of non-stationary time series. ARIMAX enhances this paradigm by incorporating exogenous factors that can alter the target time series, allowing for the inclusion of external features. SARIMA addresses seasonal trends by incorporating seasonal differencing and seasonal components into the ARIMA framework, allowing it to be used with data that has periodic swings. Finally, SARIMAX combines SARIMA's properties with exogenous variables to allow for thorough modeling of seasonal time series while accounting for external impacts.

Fig. 6 shows the different time series models and their particular features. Together, these models provide a powerful toolset for evaluating and predicting complicated time series data, with each model adapted to the precise properties of the data under consideration.

To evaluate the effectiveness of different modelling approaches for battery health estimation, several recent studies were reviewed. Table 3 summarizes their work, highlighting the estimation method categories, datasets used, and their reported accuracies.

3.3. Summary of health state estimation methods

While Table 3 compares estimation accuracy for a particular dataset with varying estimation methods across a few recent studies, accuracy alone does not determine a model's suitability for deployment within a practical BMS. For real-time applications, particularly in automotive and maritime systems, factors such as model interpretability, computational efficiency, and data dependency play equally critical roles. To address these aspects, Table 4 provides a systematic benchmark of mainstream health state estimation model categories, evaluating their physical interpretability, complexity, data requirements, and computational feasibility for embedded BMS implementation.

4. Open-source battery cell datasets

Numerous data sets have been produced throughout the years as a result of increased study into battery technology, particularly Li-ion batteries. These data sets are available online, and most are free to use for research purposes. These battery data sets are based on a variety of battery models and can be used to estimate a variety of battery states, SOH, SOC, RUL etc., under a variety of conditions [75]. In this chapter, a few commonly used degradation datasets are explained, and a summary table consisting of degradation as well as performance datasets based on cell chemistry and other parameters is presented.

4.1. Fast-charging optimization dataset from MIT and Stanford University

In their combined research [50], the authors implemented data-driven models that accurately forecast the cycle life of industrial lithium ferrous phosphate (LFP)/graphite cells. A dataset of 124 cells

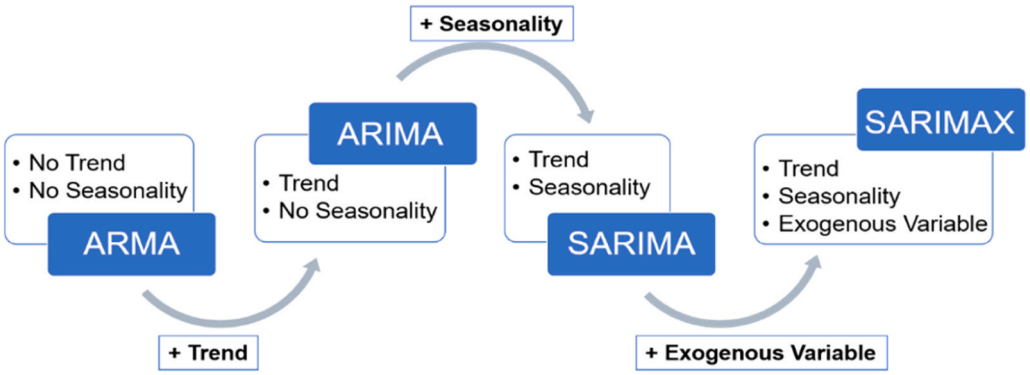

Fig. 6. Time Series Models and Features.

Table 3
Comparison of battery state estimation methods, datasets, and performance grouped based on estimation method and common dataset.

Dataset Used	Estimation Method	Reported Accuracy	Methodology	References
NASA Battery Aging Dataset	Adaptive Filter	RMSE ≈ 0.0653–3.5 %	Developed an adaptive dual Kalman filter to jointly estimate battery health states.	[68]
		RMSE ≈ 0.03–0.07 Ah	Proposed an improved particle filter optimized by a genetic algorithm to estimate the SOH across normal degradation and capacity-regeneration stages of lithium-ion batteries	[69]
	Machine Learning	RMSE ≈ 0.147–0.724 Ah	Conducted a comparative analysis of nine machine-learning and data-driven algorithms for RUL and SOH estimation of LIB, benchmarking error rates and processing times.	[70]
		RMSE ≈ 0.741–1.4 Ah	performed a comparative study of machine-learning techniques for SOH estimation using SVR, FNN, CNN, LSTM models	[71]
	Bayesian + ML	RMSE ≈ 0.8 %	Integrates Bayesian inference with ML for robust estimation.	[72]
	Hybrid	RMSE ≈ 2 %	Joint estimation framework for SOH and RUL in LIB by extracting health-indicators, decomposing them via variational mode decomposition (VMD) and using a model-integration scheme combining feature processing with an optimized support vector machine variant.	[73]
	Time-Series Model	RMSE ≈ 0.01–0.0.35 Ah	Developed AR, ARIMA, SARIMA models for a comparative study	[74]

Table 4
Systematic benchmarking of mainstream health state estimation model categories.

Model Category	Physical Interpretability	Modeling Complexity	Dependency on Dataset Size / Quality	Computational Overhead for Real-Time BMS	Remarks / Typical Use Case
Physics-Based Models	Very High; parameters correspond directly to electrochemical processes and failure mechanisms.	High; requires detailed electrochemical equations, parameter identification, and calibration.	Low; can function with minimal historical data.	High; numerically intensive; difficult for embedded BMS implementation.	Best suited for laboratory diagnostics and mechanism understanding.
Statistical / Empirical Models	Moderate; interpretable trend coefficients but lack physical linkage.	Low to Moderate; simple regression or curve-fitting equations.	Moderate; relies on historical performance data.	Low; lightweight implementation possible.	Suitable for fleet-level degradation trend monitoring.
Adaptive Filter Methods (KF, EKF, UKF)	Moderate to High; uses simplified equivalent circuit parameters tied to physical states.	Moderate; recursive formulations; extensions increase complexity.	Low to Moderate; small calibration datasets suffice.	Moderate; feasible for real-time BMS with proper tuning.	Ideal for on-board SOC-SOH estimation in EVs.
Machine Learning Models (ANN, RNN, LSTM, CNN)	Low; acts as a black-box; limited physical transparency.	High, deep architectures and hyperparameter tuning are required.	Very High; performance heavily depends on large, high-quality datasets.	Moderate to High; inference faster than training but still memory-intensive.	Suitable for predictive analytics and pattern recognition in large datasets.
Bayesian / Probabilistic Models	Moderate; probabilistic inference allows uncertainty interpretation.	Moderate to High; requires distribution modelling and parameter sampling.	Moderate; can combine small data with prior knowledge.	Moderate; acceptable for online updating if simplified.	Effective when uncertainty quantification is critical.
Time-Series Models (ARIMA, SARIMA, ARIMAX)	Low to Moderate; statistically interpretable but lacks physical linkage.	Low; simple to implement and tune.	Moderate; dependent on data stationarity and noise.	Low; minimal computational cost; ideal for embedded systems.	Useful for short-term cycle prediction and second-life analysis.
Hybrid Physics- ML Models	High; combines physical insights with data-driven adaptability.	High; requires coupling of models and cross-validation.	Moderate to High; depends on both prior knowledge and data volume.	Moderate; optimized variants suitable for BMS deployment.	Emerging optimal trade-off between accuracy, interpretability, and real-time capability.

with cycle lifetimes that vary from 150 to 2300 was created using 72 distinct fast-charging settings. Their feature-based models were able to obtain prediction errors of 9.1 % utilizing data from the first 100 cycles, indicating the ability to forecast behavior far into the future. Besides, using data from the first 5 cycles, labeling into low- and high-lifetime groups was accomplished with a misclassification test error of 4.9 %. These findings demonstrated the usefulness of integrating data production and data-driven modeling.

The study systematically varied charging circumstances to record a wide variety of cycle lives, from about 150–2300 cycles. Despite limiting the enclosure temperature, cell temperatures fluctuated by up to 10 °C every cycle due to the substantial heat generated throughout charge and discharge. The collection included roughly 96,700 cycles, making it one of the biggest publicly available dataset.

Data can be accessed at: https://data.matr.io

4.2. Fixed current and arbitrary used profile battery degradation dataset

A type of nominally identical high-energy 18650 LIBs manufactured by LISHEN was employed as the experimental subject. The rated capacity was 2.4 Ah, the nominal voltage was 3.7 V, and the lower and upper cut-off voltages were 3.0 V and 4.2 V, respectively. A total of 77 batteries were cycled for degradation tests. During stage 1, 20 initial cycles were applied for simulating the battery application to observe the initial prediction of battery RUL. Each preliminary cycle consisted of 0.5 C constant-current-constant voltage (CC-CV) charging and 2 C constant current discharging. In stage 2, to examine the degradation properties under varying operating conditions, the 77 batteries were classified into two groups, I and II, for supplementary cyclic degradation experiments. Each working condition included a charge load at a random constant charge current (randomly every 5 cycles) that followed a uniform distribution among three distinct selections of 1 C, 2 C, and 3 C, and a discharge load profile at a stated discharge current of 3 C [76].

Data can be accessed at FCA Battery Degradation Dataset

4.3. Deep discharge aging dataset from NASA

NASA published two high-throughput capability battery degradation datasets available on its website, totaling 62 cells. The first of these datasets, 'Battery Data Set', includes information about 34 Li-ion 18650 cells having a nominal capacity of 2 Ah. This set of data was the first publicly accessible battery dataset, and it had a tremendous influence on the field, offering insight into its value. Cells were cycled at a variety of ambient temperatures (4°C, 24°C, and 43°C), charged using a standard CC-CV methodology, and discharged using various methods. The dataset comprises readings of terminal current, voltage, and cell temperature during the cycle, as well as discharge capacity measurements between cycles and EIS impedance measurements.

NASA's second dataset, the 'Randomized Battery Usage Data Set' [23], includes data for 28 LCO (lithium cobalt oxide) 18650 cells that have a nominal capacity of around 2.2 Ah. The dataset is made up of seven distinct sets of four cells, each cycled at a specific ambient temperature (room temperature, 40°C); for five of these groups, the cells were CCCV charged and then discharged using currents drawn at random from the group's discharge distribution table [23,24,77].

Data can be accessed at Li-ion Battery Aging Datasets | NASA Open Data Portal

4.4. Pulse cycling, capacity fade, storage aging dataset from centre for advanced life cycle engineering (CALCE)

The CALCE battery group conducted extensive cycling testing on a wide range of LCO/graphite cells. The CALCE dataset contains data on 15 LCO prismatic CS2 cells categorized by experimental circumstances as 'Type-1′ to 'Type-6′. 'Type-1′ and 'Type-2′ accompany one document [78], and 'Type-3′ to 'Type-6′ another [79]. Type-1′ has four 0.9 Ah cells, 'Type-2′ has four 1.1 Ah cells, and 'Type-3′ through 'Type-6′ each have one to two 1.1 Ah cells.

The degradation data was logged until the batteries had at least reached their EOL, or 80 % SOH, having fewer than 200 cycles collected data for the 'Type-1′ batteries and roughly 800 cycles for the remaining cells. CALCE tested the second set of cells, which were twelve LCO prismatic CX2 cells having an approved capacity rating of 1.35 Ah. Which, like the CS2 cells, are classified as 'Type-1′ to 'Type-6′. 'Type-1′ and 'Type-2′ (a total of four cells each) have been cycled identically as 'Type-1′ of CS2 cells [80]. The other four groups each feature a single cell that has been cycled using a variety of charge/discharge techniques; one of the cells was cycled at different temperatures (25 °C, 35 °C, 45 °C, 55 °C).

In further battery studies [81], the researchers evaluated the influence of varying depths of discharge, or DOD, and discharging current stressors on the ageing of pouch cells. Battery group tested around 16 LCO 1.5 Ah pouch cells in a 'semi-temperature controlled' room (25 ± 2°C). It contains cycler voltage, current, and charging and discharging capacity data for 400–800 'equivalent cycles' [61].

The data can be accessed at CALCE Battery Datasets

4.5. Long-term battery degradation dataset from sandia national labs

The Sandia National Laboratories has completed experimenting with 3 different 18650 cells:

- A123 Systems - LFP (APR18650M1A, 1.1 Ah),
- Panasonic NCA (NCR18650B, 3.2 Ah), and
- LG Chem NMC (18650HG2, 3 Ah)

In total, it includes 86 cells (30 LFP, 24 NCA, and 32 NMC). The cells were cycled at a variety of temperatures (15 °C, 25 °C, and 35 °C) with varying DODs (0–100 %, 20–80 %, and 40–60 %) and discharge currents (0.5 C, 1 C, 2 C, and 3 C); at least two cells from every group were cycled for each combination of temperature, DOD, and discharge current (12 groups), except for the 3 C discharge for the NCA cells. All cells have been charged at a constant rate of 0.5 C. Periodically (about every 3 % capacity loss), measurements using the EIS were made to determine the cell's entire capacity. All of the information is provided in the '.csv' format [82,83].

The data can be accessed at R&D Data Repository – DOE Office of Electricity Energy Storage Program

4.6. Battery degradation data from Oxford University

The dataset's parts are as follows: Parts one, two, and three [84]. One of these is the 'Path dependence battery degradation dataset', which consists of three sections. The three-year experiment [85], running from 2017 to 2020, investigated the 'path dependence' of Li-ion cells by subjecting them to a series of combination load profiles that included defined periods of calendar and cyclic aging.

The study looked at 28 industrial 3 Ah 18650 NCA/graphite cells (NCR18650BD). The dataset is divided into three sections (Sections 1, 2, and 3), with 28 cells divided into 10 groups (9 that consist of three cells and 1 group of one cell), all of which were evaluated at 24°C [86]. The data provided included fundamental metrics such as time, current, voltage, capacity, and temperature. Groups 1–4, each containing three cells, were aged by cycling at half or quarter C rates after about one week of calendar aging per 48 cycles.

'Part 1′ presents the first 18 months of experimental data [25], whereas 'Part 2′ presents months 19–36 [87]. In addition to cell Groups 1–4, Part 2 has Groups 5 and 6, which serve as controls. Group 5 cells are subjected to continuous C/2 cycling, but Group 6 cells are only exposed to calendar degradation (at 90 % SOC). Group 7–10 is presented in the dataset's 'Part 3′ and corresponds to Group 1–4. Every group undergoes cycling with CC-CV profiles, followed by 5 or 10 days of calendar ageing. Reference performance tests (RPT) along with EIS tests are utilized regularly to characterize the cells and distinguish the effects of different storage durations and C-rates on battery degradation.

Data can be accessed here: Oxford Battery Degradation Dataset 1

4.7. EVERLASTING project battery dataset

A recent project [88] 'Electric Vehicle Enhanced Range, Lifetime and Safety Through INGenious battery management' (EVERLASTING) funded by the European Commission, has published some battery-related datasets on the '4TU.ResearchData' website [88,89]. The report investigates 3 datasets for ageing from three perspectives: drive cycle, calendar, and CC-CV ageing at various temperatures.

One of these datasets was used in an experiment called 'Lifecycle ageing' to study the connections between temperature, charge/discharge C-rates, and capacity loss. These experiments were carried out on 28 Li ion 18650 3.5 Ah commercial cells at various temperatures (0°C, 10°C, 25°C, and 45°C), discharge rates (0.5 C, 3 C), and charge rates (0.5 C, 1 C). Two cells were tested for each possible combination of temperature/charge rate as well as temperature/discharge rates (except for 0°C discharge). All 'charge' ('discharge') studies followed a similar discharge (charge) profile. The data is stored separately by temperature (0°C and 10°C) and (25°C and 45°C) [90,91].

Data can be accessed here: EVERLASTING Project

4.8. Summary of open-source battery datasets

Table 5 shows a comprehensive collection of LIB datasets that cover topics such as cathode types (LFP, LCO, NMC, etc., various cycling processes, applications, key features, and details on the institute responsible for hosting and managing them.

To complement the dataset overview, Table 6 summarizes some representative studies that have employed these open-source datasets. This comparison underlines how model accuracy and generalizability are strongly influenced by the dataset choice, preprocessing quality, and

Table 5
Summary of available battery degradation dataset based on cathode material, and cycling process.

Dataset Name	Institute	Cathode	Cycling Process	Intended Application	Key Features	References
Fast-Charging Optimization Dataset	MIT, Stanford University	LFP	150–2300 cycles, 72 distinct fast charging settings	EV	Fast-charging strategies, voltage/current/temp	[92]
Accelerated Cycle Life and Capacity Degradation	University of Maryland	LCO	Testing under various temperatures and current cycles	Capacity fade modeling	Capacity vs cycle degradation under multiple stress profiles	[93]
Fixed Current Profiles & Arbitrary Used Profiles	MIT, Beijing Institute of Technology	LCO	Complete cycle testing under various temperature and current conditions	Capacity fade modeling	Capacity vs cycle degradation under multiple stress profiles	[94]
Deep Discharge Aging Dataset	NASA	LCO	Discharge was carried out at CC level of 2 A until the voltage drops to 2.7, 2.5, 2 V	-	Charge/Discharge, EIS cycles to induce deep discharge aging effects	[95]
Pulse Cycling CS2 Series Dataset	CALCE	LCO	CCCV charge (0.5 C), discharge 0.5–1 C, pulsed load protocols	Consumer electronics/aging studies	Multiple discharge modes, pulsed cycling	[96]
Capacity Fade and Variable DOD Cycling Dataset		LCO	Full CCCV cycles at 0.5 C, partial SOC cycles (20–80 %, 40–60 %) at 0.5 C & 2 C	SOC-range cycling impact	Comparison of partial vs full DOD cycling, Capacity fade	
Storage Aging Dataset		LCO, LFP, NMC	Calendar aging (storage at 0, 50, 100 % SOC) with periodic measurement	Calendar aging & reliability	SOC & temperature effects on calendar fade	
Long-Term Degradation Study	Sandia National Laboratories	NCA, NMC, LFP	Sequential rounds of cycling: capacity checks (0–100 %) at 0.5 C + conditioning cycles (C-rate, temperature, DoD)	Lifetime degradation modeling	Evaluates effects of temperature, DoD & discharge rate across chemistries; capacity fade vs cycle; EIS-based impedance growth data	[97]
NCA Battery Degradation Dataset	University of Oxford	NCA	Drive cycle, characterization	EV / Consumer Electronics	Voltage/current/temp measurements	[98]
Life Cycle Aging Dataset	EVERLASTING project	NMC	Long-term cycling	EV	Emphasis on longevity and degradation	[88,89]
Second Life Battery Diagnostics PulseBat Dataset	N/A	NMC, LFP, LMO	Pulse Charge/Discharge Cycles	EV, second-life	Second-life diagnostics, voltage & temp response	[99]
Large Scale EV Battery Dataset	N/A	—	Real-world driving cycles	EV	Charging records, health & capacity estimation	[100]
Battery Relaxation Dataset	Hawaii Natural Energy Institute (HNEI)	NMC, LFP	Relaxation studies (charge & discharge)	EV / Research	Focus on relaxation phenomena	[101]
Intrinsic Variability Dataset	Hawaii Natural Energy Institute (HNEI)	NMC, LCO	Periodic Reference Performance Tests, then CC-CV charge (0.5 C) and CC discharge (1.5 C) for aging	SOH degradation modeling & variability analysis	Focus on cell-to-cell degradation variability under identical cycling conditions and temp	[26]
Comprehensive Battery Aging Dataset	Karlsruhe Institute of Technology	NMC	Cyclic & calendar aging, driving cycles	EV	3 + billion data points, raw & results data	[102]
Fast Charging Test Dataset	N/A	NMC, LCO	Baseline: CCCV cycling at increasing C-rates; baseline tests to study capacity fade	Fast-charging protocol optimization & aging risk assessment	Multi-mode fast-charge protocols, capacity fade under high C-rate charging	[103]
Lithium-Ion Battery Field Data Iontech Dataset	N/A	LFP	Aging under identical profiles	General	133 million rows of measurements.	[104]
Electrochemical cycling dataset	University College London (UCL)	NMC	CC-CV charge (1.5 A to 4.2 V), CC discharge (4 A to 2.5 V) – 400 full cycles	EV / degradation modelling	Full electrochemical cycles + x-ray CT microstructure	[105]
NMC Cyclic Aging Data	Politechnika Poznanska	NMC	Cycling under varied discharge current, temperature, DoD; capacity vs SOH captured at intervals	SOH degradation modeling, prognostics	Study of effects of current, DoD, temperature on aging and SOH regression models	[106]

experimental context.

5. Discussion

The rapid advancement and extensive usage of LIBs have emphasized the critical need for a strong understanding of their degradation mechanisms to avoid accidents. Due to their high energy density and long cycle life, they make them ideal for a bunch of small and large-scale applications, but performance reliability and loss over time remain a concern. In this discussion chapter, modelling techniques, their broader implications, as well as the challenges and opportunities that remain for future research in this evolving field, are discussed.

Accurate and reliable battery health parameter predictions depend heavily on the understanding of degradation mechanisms in LIBs. The importance of estimating SOH, SOC, and RUL has therefore become important as these parameters directly determine performance, safety, degradation mechanisms and lifecycle costs. The estimation methodologies reflect the multifaceted nature of degradation, where no single model type universally applies across all battery chemistries, use cases, or degradation patterns.

In recent years, research into LIB health prediction and modelling has exponentially accelerated due to the availability of open-source battery cell degradation datasets repositories from institutes such as the Toyota Research Institute, Mendeley Data, NASA, CALCE, Sandia

Table 6
Comparative summary of a few studies using common open-source battery datasets for SOC, SOH and RUL Estimation grouped based on the dataset.

Dataset Used	Model / Method	Reported Accuracy	Remarks	Study / Reference
NASA Battery Aging Dataset	Adaptive dual Kalman filter	RMSE ≈ 0.0653–3.5 %	Robust under sparse data and noise	[68]
	Bayesian inference model	RMSE ≈ 0.8 %	Fast convergence, accurate tracking	[72]
	ML Hybrid	RMSE ≈ 2 %	Effective nonlinear degradation capture	[73]
CALCE Dataset	Bayesian Inference + G-sum PF	RMSE ≈ 1.8 %	Stable under multi-temperature conditions	[60]
Oxford Path-Dependence Dataset	Machine learning Regression Model	RMSE ≈ 2.5 %	A simple and feasible way to estimate the SOH of electric vehicles	[107]
Sandia Dataset	Light Gradient Boosting Machine (LightGBM) and Long Short-Term Memory (LSTM)	RMSE ≈ 2.3–5 %	A novel feature engineering approach with purpose-designed features	[108]
MIT–Stanford Fast-Charging Dataset	Feature-based ML (Elastic Net + RF)	RMSE ≈ 9.1 % test error	Life prediction using the first 100 cycles	[50]
HNEI Relaxation Dataset	Battery-Insight-PSO, using the Extreme Gradient Boosting Regression (XGBoost)	RMSE ≈ 0.251–5.206 %	Accurate SOH mapping under variable loads	[109]

National Laboratories, and Oxford University. Given the cost, safety concerns, and time requirements of experimental testing, the availability of such open-source datasets is particularly important for developing predictive models across different chemistries and cycling processes. These datasets, however, have some limitations, including different test protocols, sensor resolutions, battery chemistries, and testing conditions. These limitations lead to challenges in model generalization across datasets. For example, models developed and trained on the NASA dataset may underperform when tested against datasets from Oxford University or CALCE due to variations in the cycling process or data recording frequencies that are not represented during training, even if they are of the same chemistry and cell type. Transfer learning is a viable option to overcome these limitations by utilizing the training knowledge learned from these open-source datasets to accelerate model adaptation for novel or limited-data batteries through selective retraining or fine-tuning of learned parameters. The transfer learning approach works by capturing the universal degradation features, such as voltage, current and capacity curve learning only for lower layers of the neural networks or for hybrid architectures. When implementing on a new cell, for example, moving from NMC to LFP cells or from laboratory-based testing to actual systems, only the top layers of the model are retrained using a comparatively small number of cycles from the target dataset. This process reduces the dependency of the new dataset which might be unavailable or unreliable due to sensor fault for example, in the process the model still retains the high prediction accuracy. This methodology enhances model generalization, reduces the time needed to develop predictive degradation models for new cell types supporting fast-paced deployment in real-life industrial applications. This directly improves the industrial efficiency by minimizing the requirement of repeated full-cycle age tests which take a lot of time to run.

Moreover, certain datasets may lack metadata or contextual information, which might be crucial for modeling. Despite these challenges, the combined usefulness of these datasets has the potential to serve as benchmarks for developing initial degradation models, assuming that future efforts focus on standardizing data formats with contextual metadata to improve transferability and real-world relevance.

Adaptive filters, ML models, statistical filters, and hybrid methods are the most often implemented SOH and RUL modeling approaches to incorporate the non-linear battery degradation behavior. Adaptive filter methods such as Kalman Filters have proven benefits for real-time battery health estimation. However, the EKF models may not perform well in extremely nonlinear or dynamic situations and still rely on the correctness of the underlying battery model. To overcome this shortcoming, ML methods such as NN-based approaches exhibit strong degradation pattern recognition abilities under dynamic and nonlinear conditions. Estimation models demonstrate NNs' advantage in modeling nonlinear degradation but rely on large datasets and may lack physical interpretability. Even though NNs are efficient at detecting complex patterns and adapting to new data, they require extensive training, and model performance increases with increasing training dataset size. On the other hand, time series models generally offer statistically dependable and easily interpretable frameworks for capturing temporal performance and degradation trends. In situations where capturing temporal dependencies is important, time series models have proven to be effective. The weakness of time series models is that they are less adaptable in handling highly nonlinear and large datasets. These disparities highlight the growing need for hybrid approaches that can combine the strengths of both data-driven and physics-based models to improve battery health state prediction accuracy.

Reliable prediction of battery health states and RUL is critical across many industries that rely heavily on LIBs. For example, in EVs, RUL predictions help optimize battery consumption and schedule maintenance proactively. This also helps avoid unexpected failures and reduces overall lifecycle costs of the battery as well as the overall system. Likewise, in the maritime industry, LIBs are increasingly implemented for hybrid propulsion and onboard energy storage.

In addition, upcoming regulatory requirements strengthen the importance of such models. The revised EU Battery Regulation 2023/1542 mandates the implementation of a Digital Battery Passport for EVs, light transport, and industrial batteries over 2 kWh capacity starting in February 2027 [110]. A structured electronic record will be accessible via a QR code, which must include detailed information on battery cell chemistry, carbon footprint, declared EOL, and performance metrics such as SOH, degradation trends, cycle history, and RUL. These regulatory requirements significantly reinforce the case for developing robust predictive models, as they will directly support compliance, transparency, and safe second-life applications.

6. Conclusion

To summarize, LIBs have become vital in almost all applications due to their high energy density and long cycle life. A recurring challenge remains in reliably predicting the degradation, which has a direct influence on system performance and safety. Hence, comprehensive degradation studies are required to guarantee the long-term viability of the LIBs. However, studying the degradation of batteries is a challenging task due to the difficulties associated with the collection of data over lengthy time periods, the cost associated with testing facilities, and the safety aspect of the experiments.

This paper highlights some open-source degradation datasets that are frequently implemented for developing predictive models. This database also provides researchers with a time-effective and scalable way to investigate battery performance and degradation, allowing them to overcome the limitations of prolonged experimental data generation. Important parameters such as RUL must be predicted to anticipate when batteries would no longer satisfy performance specifications through a degradation model. Hence, the accurate estimation of RUL is critical for the effective implementation of maintenance strategies and replacement

plans in large-scale energy systems, where unexpected breakdowns can prove to be costly and have effects on the entire process, which is undesirable. Equally, SOH and SOC estimation remain central to understanding battery status and ensuring safe operation, linking degradation mechanisms to predictive modelling outcomes.

To improve the predictions, advanced data-driven methodologies such as adaptive filters, including KFs, EKFs, and UKFs, provide dynamic modeling of battery health states while accounting for nonlinearities and uncertainties in the degradation process. Machine learning models can evaluate complex degradation patterns and forecast battery performance under a range of operating conditions and especially with large datasets. Recently, researchers have also been exploring hybrid models that can integrate the strengths of adaptive filtering and machine learning together. These hybrid models have shown considerable good results due to their nature of combining physical interpretability with data-driven adaptability. Although developing hybrid models requires a deep technical understanding of the domain. These hybrid models offer the potential to achieve higher accuracy if modelled correctly. Integrating open-source data with new computational tools like adaptive filtering and machine learning can help create more reliable battery degradation models.

Furthermore, the upcoming EU Digital Battery Passport policy makes accurate prediction of SOH and RUL not only a research priority but also a regulatory requirement. This policy implementation will further strengthen the need for precise degradation models for transparent tracking of battery health throughout the lifecycle. By linking mechanistic understanding, predictive modelling, and standardized datasets, these efforts will help enhance the safety and operational efficiency of LIBs in large-scale energy and transport systems.

CRediT authorship contribution statement

Kishan Patel: Writing – review & editing, Writing – original draft, Project administration, Methodology, Conceptualization. **Sören Ehlers:** Writing – review & editing, Supervision. **Moritz Braun:** Writing – review & editing, Supervision. **Merten Stender:** Writing – review & editing, Supervision. **Vaidehi Gosala:** Writing – review & editing, Validation.

Declaration of Competing Interest

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

Acknowledgement

The research leading to these results has been carried out under the framework of the development of the innovation platform by the DLR Institute of Maritime Energy Systems. The project started in January 2025 and is led by the Program Directorate Transport within the German Aerospace Center (DLR), whose support we greatly appreciate.

Data availability

No data was used for the research described in the article.

References

[1] P.J. Hall, E.J. Bain, Energy Policy 36 (12) (2008) 4352–4355, https://doi.org/10.1016/j.enpol.2008.09.037.
[2] A. Stoppato and A. Benato, "The Importance of Energy Storage," in Energy Storage (World Scientific Series in Current Energy Issues), G. M. Crawley, Ed., World Scientific, pp. 1–26.
[3] Z. He, M. Gao, C. Wang, L. Wang, Y. Liu, Energies 6 (8) (2023) 4134–4151, https://doi.org/10.3390/en6084134.
[4] M.A. Giovanniello, X.-Y. Wu, Appl. Energy 345 (2023) 121311, https://doi.org/10.1016/j.apenergy.2023.121311.
[5] M.M. Kabir, D.E. Demirocak, Int J. Energy Res. 41 (14) (2023) 1963–1986, https://doi.org/10.1002/er.3762.
[6] F. Leng, C.M. Tan, and M. Pecht, "Effect of Temperature on the Aging rate of Li Ion Battery Operating above Room Temperature," Scientific reports, early access. doi: 10.1038/srep12967.
[7] S.F. Schuster, et al., J. Energy Storage 1 (2025) 44–53, https://doi.org/10.1016/j.est.2015.05.003.
[8] Z. Wang, Y. Song, Q. Zhao, B. Shi, J. He, J. Li, J. Power Sources 630 (2025) 236185, https://doi.org/10.1016/j.jpowsour.2025.236185.
[9] K. Goebel, B. Saha, A. Saxena, J. Celaya, J. Christophersen, IEEE Instrum. Meas. Mag. 11 (4) (2020) 33–40, https://doi.org/10.1109/MIM.2008.4579269.
[10] G.K. Prasad, C.D. Rahn, J. Power Sources 232 (2013) 79–85, https://doi.org/10.1016/j.jpowsour.2013.01.041.
[11] M.E. Orchard, P. Hevia-Koch, B. Zhang, L. Tang, IEEE Trans. Ind. Electron 60 (11) (2023) 5260–5269, https://doi.org/10.1109/TIE.2012.2224079.
[12] L. Lu, X. Han, J. Li, J. Hua, M. Ouyang, J. Power Sources 226 (2023) 272–288, https://doi.org/10.1016/j.jpowsour.2012.10.060.
[13] M. Elmahallawy, T. Elfouly, A. Alouani, A.M. Massoud, IEEE Access 10 (2022) 119040–119070, https://doi.org/10.1109/ACCESS.2022.3221137.
[14] S. Jin, X. Sui, X. Huang, S. Wang, R. Teodorescu, D.-I. Stroe, Electronics 10 (24) (2021) 3126, https://doi.org/10.3390/electronics10243126.
[15] C. Hendricks, N. Williard, S. Mathew, M. Pecht, J. Power Sources 297 (2023) 113–120, https://doi.org/10.1016/j.jpowsour.2015.07.100.
[16] G.E. Blomgren, J. Electrochem. Soc. 164 (1) (2023) A5019–A5025, https://doi.org/10.1149/2.0251701jes.
[17] T. Rahman, T. Alharbi, Batteries 10 (7) (2024) 220, https://doi.org/10.3390/batteries10070220.
[18] X. Han, et al., eTransportation 1 (2019) 100005, https://doi.org/10.1016/j.etran.2019.100005.
[19] I. Bloom, et al., J. Power Sources 101 (2) (2023) 238–247, https://doi.org/10.1016/S0378-7753(01)00783-2.
[20] R. Wright, et al., J. Power Sources 119-121 (2023) 865–869, https://doi.org/10.1016/S0378-7753(03)00190-3.
[21] T.C. Bach, et al., J. Energy Storage 5 (2023) 212–223, https://doi.org/10.1016/j.est.2016.01.003.
[22] R. Guo, L. Lu, M. Ouyang, and X. Feng, "Mechanism of the entire overdischarge process and overdischarge-induced internal short circuit in lithium-ion batteries," Scientific reports, early access. doi: 10.1038/srep30248.
[23] NASA. "Li-ion Battery Aging Datasets." Accessed: 09-Apr-24. [Online]. Available: https://data.nasa.gov/dataset/Li-ion-Battery-Aging-Datasets/uj5r-zjdb/about_data.
[24] Y. Preger, et al., J. Electrochem. Soc. 167 (12) (2020) 120532, https://doi.org/10.1149/1945-7111/abae37.
[25] T. Raj, Path Dependent Battery Degradation Dataset Part 1. University of Oxford, 2020. [Online]. Available: https://ora.ox.ac.uk/objects/uuid:de62b5d2-6154-426d-bcbb-30253ddb7d1e.
[26] A. Devie, G. Baure, M. Dubarry, Energies 11 (5) (2023) 1031, https://doi.org/10.3390/en11051031.
[27] H. Meng, Y.-F. Li, Renew. Sustain. Energy Rev. 116 (2019) 109405, https://doi.org/10.1016/j.rser.2019.109405.
[28] J. Lee, O. Nam, B.H. Cho, J. Power Sources 174 (1) (2023) 9–15, https://doi.org/10.1016/j.jpowsour.2007.03.072.
[29] F. Mohammadi, J. Energy Storage 48 (2022) 104061, https://doi.org/10.1016/j.est.2022.104061.
[30] Y. Zhu, Y. Xiong, J. Xiao, T. Yi, C. Li, Y. Sun, J. Energy Storage 73 (2023) 108917, https://doi.org/10.1016/j.est.2023.108917.
[31] J. Lee, J. Won, IEEE Access 11 (2023) 15449–15459, https://doi.org/10.1109/ACCESS.2023.3244801.
[32] C.-S. Huang, J. Electr. Eng. Technol. 19 (1) (2024) 433–442, https://doi.org/10.1007/s42835-023-01533-9.
[33] K. Movassagh, A. Raihan, B. Balasingam, K. Pattipati, Energies 14 (14) (2021) 4074, https://doi.org/10.3390/en14144074.
[34] S. Hosseininasab, C. Lin, S. Pischinger, M. Stapelbroek, G. Vagnoni, J. Energy Storage 52 (2022) 104684, https://doi.org/10.1016/j.est.2022.104684.
[35] Z. Xu, J. Wang, P.D. Lund, Y. Zhang, Energy 240 (2022) 122815, https://doi.org/10.1016/j.energy.2021.122815.
[36] L. Wu, H. Pang, Y. Geng, X. Liu, J. Liu, K. Liu, Int J. Energy Res. 46 (9) (2022) 11834–11848, https://doi.org/10.1002/er.7949.
[37] X. Hu, D. Cao, B. Egardt, IEEE/ASME Trans. Mechatron. 23 (1) (2018) 167–178, https://doi.org/10.1109/TMECH.2017.2675920.
[38] D. Siegel, J. Lee, C. Ly, IEEE Conf. Progn. Health Manag. (2011) [Online]. Available: https://www.semanticscholar.org/paper/Methodology-and-framework-for-predicting-rolling-Siegel-Lee/379c77749992e03be291293ef558923fdb25336a.
[39] M. Ahwiadi, W. Wang, Measurement 191 (2022) 110817, https://doi.org/10.1016/j.measurement.2022.110817.
[40] Y. Kim, N.A. Samad, K.-Y. Oh, J.B. Siegel, B.I. Epureanu, A.G. Stefanopoulou, 2016 Am. Control Conf. (ACC) (2016) 1500–1505, https://doi.org/10.1109/acc.2016.7525128.
[41] Z. Chen, Y. Fu, C.C. Mi, IEEE Trans. Veh. Technol. 62 (3) (2020) 1020–1030, https://doi.org/10.1109/TVT.2012.2235474.
[42] J. Jiang, Y. Gao, C. Zhang, W. Zhang, Y. Jiang, J. Electrochem. Soc. 166 (6) (2019) A1070–A1081, https://doi.org/10.1149/2.1051904jes.
[43] A. S.P, Sathiyasekar, JNMES 27 (2) (2024) 86–98, https://doi.org/10.14447/jnmes.v27i2.a02.

[44] F. Zhu, J. Fu, IEEE Sens. J. 21 (22) (2021) 25449–25456, https://doi.org/10.1109/JSEN.2021.3102990.
[45] C. Huang, Z. Wang, Z. Zhao, L. Wang, C.S. Lai, D. Wang, IEEE Access 6 (2018) 27617–27628, https://doi.org/10.1109/ACCESS.2018.2833858.
[46] S.M. Rezvanizaniani, Z. Liu, Y. Chen, J. Lee, J. Power Sources 256 (2020) 110–124, https://doi.org/10.1016/j.jpowsour.2014.01.085.
[47] H. Chaoui, C.C. Ibe-Ekeocha, IEEE Trans. Veh. Technol. 66 (10) (2020) 8773–8783, https://doi.org/10.1109/TVT.2017.2715333.
[48] H. Rauf, M. Khalid, N. Arshad, Renew. Sustain. Energy Rev. 156 (2022) 111903, https://doi.org/10.1016/j.rser.2021.111903.
[49] Y. Li, et al., Renew. Sustain. Energy Rev. 113 (2019) 109254, https://doi.org/10.1016/j.rser.2019.109254.
[50] K.A. Severson, et al., Nat. Energy 4 (5) (2019) 383–391, https://doi.org/10.1038/s41560-019-0356-8.
[51] M.A. Hannan, M.M. Hoque, A. Hussain, Y. Yusof, P.J. Ker, IEEE Access 6 (2018) 19362–19378, https://doi.org/10.1109/access.2018.2817655.
[52] X. Tang, K. Liu, X. Wang, F. Gao, J. Macro, W.D. Widanage, IEEE Trans. Transp. Electrific 6 (2) (2020) 363–374, https://doi.org/10.1109/TTE.2020.2979547.
[53] Y. Che, Z. Deng, X. Lin, L. Hu, X. Hu, IEEE Trans. Veh. Technol. 70 (2) (2021) 1269–1277, https://doi.org/10.1109/tvt.2021.3055811.
[54] B. Chinomona, C. Chung, L.-K. Chang, W.-C. Su, M.-C. Tsai, IEEE Access 8 (2020) 165419–165431, https://doi.org/10.1109/access.2020.3022505.
[55] Z. Chen, M. Wu, R. Zhao, F. Guretno, R. Yan, X. Li, IEEE Trans. Ind. Electron 68 (3) (2021) 2521–2531, https://doi.org/10.1109/tie.2020.2972443.
[56] R.S.D. Teixeira, R.F. Calili, M.F. Almeida, D.R. Louzada, Batteries 10 (3) (2024) 111, https://doi.org/10.3390/batteries10030111.
[57] Y.K. Bharath, V.P. Anandu, U. Vinatha, S. Sudeep, Energy Storage 6 (8) (2024) e70066, https://doi.org/10.1002/est2.70066.
[58] J. Zhang, C. Wang, J. Li, Y. Xie, L. Mao, Z. Hu, Appl. Energy 351 (2023) 121855, https://doi.org/10.1016/j.apenergy.2023.121855.
[59] G. Dong, W. Han, Y. Wang, IEEE Trans. Ind. Electron 68 (11) (2021) 10949–10958, https://doi.org/10.1109/TIE.2020.3034855.
[60] G. Dong, F. Shen, L. Sun, M. Zhang, J. Wei, IEEE Trans. Instrum. Meas. 74 (2025) 1–12, https://doi.org/10.1109/TIM.2024.3497053.
[61] Battery Data | Center for Advanced Life Cycle Engineering." Accessed: 10-Apr-24. [Online]. Available: https://calce.umd.edu/battery-data.
[62] X. Hu, J. Jiang, D. Cao, B. Egardt, IEEE Trans. Ind. Electron 2 (2020) 1, https://doi.org/10.1109/TIE.2015.2461523.
[63] S. Kim, P.-Y. Lee, M. Lee, J. Kim, W. Na, J. Energy Storage 46 (2022) 103888, https://doi.org/10.1016/j.est.2021.103888 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352152X2101553X.
[64] Z. Wei, X. Sun, Y. Li, W. Liu, C. Liu, H. Lu, Electronics 14 (7) (2025) 1290, https://doi.org/10.3390/electronics14071290.
[65] B. Saha, K. Goebel, J. Christophersen, Trans. Inst. Meas. Control 31 (3-4) (2020) 293–308, https://doi.org/10.1177/0142331208092030.
[66] S. Shen, S. Ci, K. Zhang, X. Liang, 2019 IEEE Globecom Workshops (GC Wkshps) (2019) 1–6, https://doi.org/10.1109/GCWkshps45667.2019.9024477.
[67] Y. Hu, S. Wang, J. Huang, P. Takyi-Aninakwa, X. Chen, Int. J. Electrochem. Sci. 17 (5) (2022) 220555, https://doi.org/10.20964/2022.05.61.
[68] J. Bao, Y. Mao, W. Zhang, Z. Zhang, Y. Zhang, J. Power Sources 656 (2025) 238036, https://doi.org/10.1016/j.jpowsour.2025.238036.
[69] H. Pan, C. Chen, M. Gu, Energies 14 (16) (2021) 5000, https://doi.org/10.3390/en14165000.
[70] M. Gupta, A. Arya, U. Varshney, J. Mittal, A. Tomar, Comparative analysis of RUL and SOH prediction algorithms using machine learning techniques. 4th International Conference on Advancement in Electronics & Communication Engineering (AECE), AECE, GHAZIABAD, India, 2024, pp. 67–72, 10.1109/AECE62803.2024.10911184.
[71] P. Eleftheriadis, M. Gangi, S. Leva, A.V. Rey, E. Groppo, L. Grande, Electr. Power Syst. Res. 235 (2024) 110889, https://doi.org/10.1016/j.epsr.2024.110889.
[72] L. Ye, et al., Proc. Inst. Mech. Eng. Part D J. Automob. Eng. 239 (12) (2025) 5455–5470, https://doi.org/10.1177/09544070241287254.
[73] X. Yu, Z. Ma, J. Wen, Int. J. Electrochem. Sci. 19 (11) (2024) 100842, https://doi.org/10.1016/j.ijoes.2024.100842.
[74] K. Patel, A. Pulickakudy Salin, M. Stender, M. Braun, S. Ehlers, 2025, Lithium-Ion Battery Degradation Forecasting Using Data-Driven Time Series Models," in Volume 7: Blue Economy Symposium; Specialty Symposium on OTEC and Correlate Devices; Smart and Sustainable Maritime Systems; Prof. Sander Calisal Honoring Symposium in Hydrodynamics, Tidal Energy, Ship Design and Education, Vancouver, British Columbia, Canada.
[75] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, J. Lin, Mech. Syst. Signal Process. 104 (2018) 799–834, https://doi.org/10.1016/j.ymssp.2017.11.016.
[76] Jiahuan Lu, "Battery Degradation Dataset (Fixed Current Profiles & Arbitrary Uses Profiles)," 2022, doi: 10.17632/kw34hhw7xg.1.
[77] Randomized battery usage data set, 2014. [Online]. Available: https://scholar.google.com/citations?user=ckhiu5saaaaj&hl=de&oi=sra.
[78] W. He, N. Williard, M. Osterman, M. Pecht, J. Power Sources 196 (23) (2020) 10314–10321, https://doi.org/10.1016/j.jpowsour.2011.08.040.
[79] N. Williard, W. He, M. Osterman, M. Pecht, IJPHM 4 (1) (2020), https://doi.org/10.36001/ijphm.2013.v4i1.1437 [Online]. Available: http://papers.phmsociety.org/index.php/ijphm/article/view/1437.
[80] Y. Xing, E.W. Ma, K.-L. Tsui, M. Pecht, Microelectron. Reliab. 53 (6) (2023) 811–820, https://doi.org/10.1016/j.microrel.2012.12.003.
[81] S. Saxena, C. Hendricks, M. Pecht, J. Power Sources 327 (2020) 394–400, https://doi.org/10.1016/j.jpowsour.2016.07.057 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378775316309247.
[82] Y. Preger, et al., J. Electrochem. Soc. 167 (12) (2020) 120532, https://doi.org/10.1149/1945-7111/abae37 [Online]. Available: https://iopscience.iop.org/article/10.1149/1945-7111/abae3.
[83] "BatteryArchive.org" Accessed: 10-Apr-24. [Online]. Available: https://www.batteryarchive.org/snl_study.html.
[84] Data and code · Battery Intelligence Lab." Accessed: 22-Apr-24. [Online]. Available: https://howey.eng.ox.ac.uk/data-and-code/.
[85] T. Raj, A.A. Wang, C.W. Monroe, D.A. Howey, Batter. Supercaps 3 (12) (2020) 1377–1385, https://doi.org/10.1002/batt.202000160.
[86] T. Raj, Path Dependent Battery Degradation Dataset Part 2. University of Oxford, 2021. [Online]. Available: https://ora.ox.ac.uk/objects/uuid:be3d304e-51fd-4b37-a818-b6fa1ac2ba9d.
[87] T. Raj, Path Dependent Battery Degradation Dataset Part 3. University of Oxford, 2021. [Online]. Available: https://ora.ox.ac.uk/objects/uuid:78f66fa8-deb9-468a-86f3-63983a7391a9.
[88] Everlasting Project. "Home - Everlasting Project." Accessed: 22-Apr-24. [Online]. Available: https://everlasting-project.eu/.
[89] 4TU.ResearchData. "4TU.ResearchData." Accessed: 22-Apr-24. [Online]. Available: https://data.4tu.nl/.
[90] Jaykanth Govindarajan, "Lifecycle ageing tests on commercial 18650 Li ion cell @ 10°C and 0°C," 2021, doi: 10.4121/14377295.V1.
[91] Khiem Trad, "Lifecycle ageing tests on commercial 18650 Li ion cell @ 25°C and 45°C," 2021, doi: 10.4121/13739296.V1.
[92] "Experimental Data Platform." Accessed: Sep. 6, 2025. [Online]. Available: https://data.matr.io/1/.
[93] W. Diao, "Data for: Accelerated Cycle Life Testing and Capacity Degradation Modeling of LiCoO2-graphite Cells," 2019, doi: 10.17632/C35ZBMN7J8.1.
[94] Jiahuan Lu, "Battery Degradation Dataset (Fixed Current Profiles & Arbitrary Uses Profiles)," 2022, doi: 10.17632/KW34HHW7XG.3.
[95] Y.C. Chen, "NASA lithium ion battery dataset," 2024, doi: 10.21227/p4st-2305.
[96] Battery Data | Center for Advanced Life Cycle Engineering." Accessed: Sep. 6, 2025. [Online]. Available: https://calce.umd.edu/battery-data.
[97] Sandia National Laboratories Study Overview." Accessed: Sep. 6, 2025. [Online]. Available: https://www.batteryarchive.org/snl_study.html.
[98] Christoph Birkl, Oxford Battery Degradation Dataset 1. University of Oxford, 2017. [Online]. Available: https://ora.ox.ac.uk/objects/uuid:03ba4b01-cfed-46d3-9b1a-7d4a7bdf6fac.
[99] S. Tao *et al.*, "PulseBat: A field-accessible dataset for second-life battery diagnostics from realistic histories using multidimensional rapid pulse test," 2025, doi: 10.48550/arXiv.2502.16848.
[100] H. He et al., "EVBattery: A Large-Scale Electric Vehicle Dataset for Battery Health and Capacity Estimation," Jan. 2022. [Online]. Available: http://arxiv.org/pdf/2201.12358.
[101] A. Fernando, M. Kuipers, G. Angenendt, K.-P. Kairies, and M. Dubarry, "Relaxation data for commercial NMC and LFP cells," 2023, doi: 10.17632/Y8NSTXMDRG.1.
[102] M. Luh and T. Blank, "Comprehensive battery aging dataset: capacity and impedance fade measurements of a lithium-ion NMC/C-SiO cell," Scientific data, early access. doi: 10.1038/s41597-024-03831-x.
[103] D. Gun, H. Perez, and S. Moura, "Fast Charging Tests," 2015, doi: 10.6078/D1MS3X.
[104] J. Schaeffer, E. Lenz, D. Gulla, M. Bazant, R.D. Braatz, and R. Findeisen, "Lithium-Ion Battery Field Data: 28 LFP battery systems with 8 cells in series, up to 5 years of operation," 2024, doi: 10.5281/zenodo.13715693.
[105] Thomas Heenan *et al.*, "Lithium-ion Battery INR18650 MJ1 Data: 400 Electrochemical Cycles (EIL-015)," 2020, doi: 10.5522/04/12159462.v1.
[106] D. Burzyński, "NMC cell 2600 mAh cyclic aging data," 2021, doi: 10.17632/k6v83s2xdm.1.
[107] R. Gao, Y. Zhang, Z. Lyu, J. Energy Storage 103 (2024) 114309, https://doi.org/10.1016/j.est.2024.114309.
[108] K. Alamin, D.J. Pagliari, Y. Chen, E. Macii, S. Vinco, M. Poncino, 2024, Model-Driven Feature Engineering for Data-Driven Battery SOH Model, in 2024 Design, Automation & Test in Europe Conference & Exhibition (DATE), Valencia, Spain, 2024, pp. 1–6, doi: 10.23919/DATE58400.2024.10546702..
[109] M.F.H. Shiblee, H. Laaksonen, Future Batter. 8 (2025) 100114, https://doi.org/10.1016/j.fub.2025.100114.
[110] A. Pohlmann, M. Popowicz, J.-P. Schöggl, R.J. Baumgartner, Clean. Prod. Lett. 8 (2025) 100090, https://doi.org/10.1016/j.clpl.2024.100090.