International Journal of Science and Engineering Applications
Volume 15-Issue 12, 13 – 16, 2026, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1512.1004

Research on State of Health Prediction of Power Batteries for New Energy Vehicles

Hailan Zhao
School of Automobiles and Low-Altitude Unamnned Aerial Vehicles
Zibo Polytechnic University
Zibo, Shandong, China

Abstract: The state of health (SOH) of power batteries is a core parameter that characterizes driving-range attenuation, power decline and safety risks of new-energy vehicles. Accurate SOH prediction is of great significance for vehicle energy management, warranty judgment and echelon utilization of retired batteries. This paper defines an SOH characterization system dominated by capacity retention rate and internal resistance growth rate, and analyzes aging mechanisms including solid-electrolyte interphase (SEI) film thickening, active-lithium loss and lithium plating, as well as the coupled influences of temperature, state-of-charge window and charge-discharge rate on degradation rates. Existing prediction methods are categorized into model-based, data-driven and mechanism-data fusion approaches, and their accuracy, computing-power requirements and engineering implementability are compared. Furthermore, a vehicle-level prediction framework of "on-board feature extraction - cloud-side model training - cloud-edge collaborative updating" is proposed. Key challenges such as working-condition generalization, label acquisition, battery inconsistency and computing-power constraints are discussed. The analysis reveals that the engineering bottleneck for SOH prediction has shifted from algorithm accuracy to data closed-loop performance and generalization capability. Constructing mechanism-constrained data-driven models calibrated with a small volume of real-vehicle calibration data is a feasible approach to improving practical performance.

Keywords: new-energy vehicles; power batteries; state of health; data-driven; cloud-edge collaboration

1. INTRODUCTION

Power batteries represent the highest-cost component with the most irreversible performance degradation in new-energy vehicles. With the continuously growing ownership of new-energy vehicles in China and the arrival of the warranty-expiry stage for the first batch of vehicles, the actual health level of batteries not only directly determines vehicle driving-range performance and safety margin, but also serves as a prerequisite for used-car residual-value evaluation, battery-swapping operation and echelon utilization of retired batteries. Nevertheless, SOH is an internal state quantity that cannot be directly read by sensors and can only be indirectly inferred from external observable variables such as voltage, current and temperature, which brings inherent ill-posedness and uncertainty to its prediction.

Early-stage SOH evaluation adopted offline tests, namely measuring available capacity through complete full-charge-full-discharge calibration or testing DC internal resistance with dedicated equipment. Such methods deliver reliable results yet suffer from long testing duration and require equipment shutdown, thus failing to support on-board vehicle management. Afterwards, research focus gradually shifted toward online estimation. One technical route centers on electrochemical or equivalent-circuit models, realizing joint estimation of capacity and internal resistance through parameter identification and state observers[4]. Another route, represented by machine learning, extracts health features from charging curves and incremental-capacity curves to establish the mapping relationship between features and SOH[1].

In recent years, benefiting from expanding internet-of-vehicles datasets and advances in deep learning, data-driven methods have achieved high accuracy on public aging datasets from NASA, CALCE, the University of Maryland and other institutions[1]. However, laboratory accuracy does not equal engineering practicability. Obvious discrepancies exist between real-vehicle data distribution and experimental working conditions; capacity labels are difficult to obtain; vehicle controller computing and storage resources are limited; meanwhile, model outputs must satisfy traceable and interpretable safety requirements. These constraints constitute the major obstacles for translating academic indicators into mass-production functions.

Based on the above understanding, this paper avoids pursuing accuracy competition among individual algorithms. From the perspective of engineering implementation, it sorts out the connotation, aging mechanisms and mainstream methods of SOH, establishes a prediction framework covering vehicle-end and cloud-end modules, and summarizes key issues in practical deployment, so as to provide references for the design of power-battery health-management functions.

2. DEFINITION OF POWER-BATTERY SOH AND INFLUENCING FACTORS OF AGING

2.1 SOH Characterization System

There exists no unique physical definition for SOH in engineering practice. Quantitative indicators are usually given from three dimensions: energy, power and dynamics. The most widely adopted metric is the capacity retention rate, i.e.

$$SOH_C = C_{aged} / C_{rated} \times 100\% \quad (1)$$

where C_aged denotes the current maximum available capacity and C_rated stands for the rated capacity. The industry generally takes 80% as the end-of-life (EOL) threshold of batteries[2]. Another power-oriented indicator characterizes degradation through the growth of DC internal resistance:

$SOH_R = (R_{EOL} - R_{now})/(R_{EOL} - R_{BOL}) \times 100\%$ (2)

This indicator reflects the large-rate charge-discharge capability and is particularly critical for models featuring fast-charging functions. In addition, constant-current charging duration, voltage-plateau position, peak values of incremental-capacity curves and other parameters can also serve as indirect health characterization metrics. It should be noted that capacity attenuation and internal-resistance growth are asynchronous. Evaluation based merely on capacity retention rate will underestimate power-performance degradation, so at the vehicle level the primary indicators should be determined according to application scenarios, with other indicators serving as constraint conditions[3].

2.2 Aging Mechanisms and Influencing Factors

The aging of lithium-ion batteries arises from the coupling effects of multiple side reactions. First, continuous thickening of the solid-electrolyte interphase (SEI) film on the negative-electrode surface consumes active lithium and electrolyte, which is manifested as capacity fading and rising interfacial impedance. Second, positive-electrode active materials suffer particle fracture, phase transition and transition-metal dissolution during cycling, so that lithium-accommodation sites decrease accordingly. Third, under low-temperature or high-rate charging conditions, lithium ions fail to intercalate into the graphite interlayers of negative electrodes and are reduced on the material surface, generating lithium dendrites. This phenomenon not only consumes active lithium but may also pierce separators and trigger internal short circuits, representing the most safety-hazardous aging form[4].

External factors alter the rates of the above side reactions and further affect degradation trajectories. In terms of temperature, empirical results show that side-reaction rates roughly double with every 10 °C rise in ambient temperature; high-temperature storage causes severer lifetime damage than high-temperature operation, while low-temperature charging greatly increases lithium-plating risks. As for the state of charge, long-term storage at high SOC accelerates electrolyte oxidation and positive-electrode material degradation, whereas deep discharge aggravates current-collector corrosion. In terms of charge-discharge rates, polarization heat and mechanical stress induced by high-rate cycling accelerate structural damage, and frequent power fluctuations exert more adverse impacts than constant-rate cycling. After cell grouping, monomer discrepancies are amplified by the "cask effect", and the weakest cell determines the available capacity and safety boundary of the whole module.

Accordingly, SOH is essentially a function of the "working-condition history" of a battery. Degradation rates of identical-model batteries may differ by several times among different users and geographical regions, which fundamentally explains why laboratory datasets cannot be directly transferred to real-vehicle scenarios.

3. HEALTH-STATE PREDICTION METHODS

3.1 Model-based Methods

Electrochemical models, represented by the P2D (Newman) model, describe lithium-ion transport in solid and liquid phases based on porous-electrode and concentrated-solution theories, and can output internal lithium-concentration and potential distributions. They feature clear physical mechanisms and can interpret the sources of capacity degradation. Nevertheless, they contain numerous parameters requiring specialized tests for calibration, and single-run computation is time-consuming and difficult for on-board deployment. Simplified single-particle models reduce the computational burden by lowering spatial dimensions, yet exhibit obvious accuracy loss under high-rate and low-temperature conditions.

Equivalent-circuit models adopt first-order or second-order RC Thevenin networks, employing ohmic resistance, polarization resistance and polarization capacitance to describe battery dynamic characteristics. Aging is reflected by increasing ohmic and polarization resistance as well as drifting time constants; therefore, these parameters can be treated as health features. Combined with extended Kalman filters, unscented Kalman filters or particle filters, joint estimation of capacity and internal resistance can be realized. Such methods deliver favorable real-time performance and can be easily embedded into existing battery-management-system software frameworks[4]. Their major limitation lies in the strong temperature and SOC dependence of parameters, which requires online identification; estimation accuracy degrades when excitation is insufficient (e.g., long-term low-current operation).

Empirical models fit degradation curves taking cycle number, ampere-hour throughput or equivalent full-cycle count as independent variables. They are concise and easy for engineering implementation, yet show poor adaptability to individuals and working conditions, and are mostly applied to rough life prediction.

3.2 Data-driven Methods

The general workflow of data-driven methods includes data acquisition and cleaning, health-feature construction, model training and validation, and online inference and updating. Health-feature construction is critical for prediction accuracy. Features are divided into direct features and refined features. Direct features include charging duration, average voltage and cumulative temperature and so on, which are easy to extract yet weakly correlated with aging. Refined features mine degradation information from charging curves, such as the peak height and peak-position offset of incremental-capacity (IC) curves, the valley-value variation of differential-voltage curves, and the charging-time proportion of the constant-voltage phase, all of which present strong correlations with capacity fading. Gu et al.[5] further introduced dynamic-time-warping distance and Wasserstein distance to measure the similarity between initial and current charging-voltage curves as novel health features; combined with the CatBoost model and SHAP attribution analysis, they achieved an average evaluation error below 2.2% and realized quantitative interpretation of feature contributions.

For model selection, support-vector regression, random forests, gradient-boosting trees and Gaussian-process regression fit small-sample scenarios, and Gaussian-process regression can further output prediction variance to evaluate result reliability. Long short-term memory networks, gated recurrent units, temporal convolutional networks and convolution-recurrent hybrid architectures excel at capturing long-term dependencies of capacity sequences and generally achieve higher accuracy with sufficient datasets.

Severson et al.[6] adopted statistical voltage-current-temperature features from early cycles and machine-learning models to predict battery cycle life before obvious capacity decay, proving that early aging trajectories contain

information on long-term degradation. Ng et al.[7] pointed out in their review that data-driven methods obtain excellent accuracy within the covered working-condition range yet possess limited extrapolation capability for out-of-distribution scenarios. This indicates that accuracy gains from merely deepening network layers are often obtained at the cost of larger datasets and higher computing-power consumption, while the black-box nature of such models hinders anomaly attribution and brings obstacles to warranty judgment and safety management.

3.3 Mechanism-Data Fusion Methods

To balance prediction accuracy and physical consistency, fusion methods have become a hot research direction in recent years, with three major technical paths. First, mechanism models generate simulation data to expand training sets for insufficient experimental samples, or transfer-learning techniques adapt laboratory-trained models to real-vehicle domains. Second, physical constraints are embedded into learning procedures: for instance, imposing monotonic-decay constraints on capacity sequences, introducing Coulomb-counting conservation relationships, or taking the correlation between aging rate and temperature and charge-discharge rate as prior knowledge to suppress spurious correlations learned by the models. Third, signal decomposition, filtering and learners are cascaded; for example, variational-mode decomposition separates the trend term and regeneration term of capacity sequences, which are then predicted and reconstructed by long short-term memory networks and Gaussian-process regression respectively[8]. Compared with pure data-driven approaches, fusion methods demonstrate stronger robustness under small-sample and high-noise conditions, but suffer from complex structures and high parameter-tuning costs, so trade-offs are still required for engineering deployment.

4. VEHICLE-ORIENTED PREDICTION FRAMEWORK AND KEY PRACTICAL ISSUES

4.1 Cloud-Edge Collaborative Prediction Framework

For real-vehicle applications, SOH prediction cannot rely merely on individual algorithms; instead, a closed-loop system covering perception, computation and updating should be constructed. This paper proposes a three-layer framework: on-board feature extraction, cloud-side model training and cloud-edge collaborative updating.

At the vehicle-end layer, the battery-management system samples cell voltage, current, temperature and equalization status at fixed frequencies and segments the raw data according to charging procedures. Considering that complete full-charge-full-discharge cycles rarely occur in real vehicles, constant-current-constant-voltage charging segments serve as the analysis objects. Coulomb integration and local capacity estimation are implemented in the constant-current phase, and incremental-capacity-curve features are calculated when the voltage range is sufficiently wide. At the cloud-end layer, multi-vehicle data are aggregated and grouped by vehicle model, battery type and service condition for modeling. Digital twins are adopted for parameter calibration of mechanism models, and candidate models go through offline validation and version management. In the closed-loop layer, models or parameters are delivered to vehicle ends via over-the-air downloads, and vehicle ends feed back inference outputs, confidence levels and hard-case samples to form an iterative closed loop of active learning. This framework moves labeling and training to resource-abundant cloud platforms while preserving vehicle-end real-time performance and data-privacy boundaries[9].

4.2 Key Issues in Engineering Deployment

(1) Scarcity of labels. Real vehicles hardly undergo periodic full-charge-full-discharge capacity calibration, so ground-truth values for supervised learning are missing. Feasible countermeasures include adopting capacity estimated from charging segments as weak labels, building benchmark datasets with laboratory calibration vehicles and bench-test data, and applying semi-supervised or self-supervised learning to reduce dependence on annotated data.

(2) Distribution drift and generalization. Training datasets mostly originate from constant-temperature, constant-rate laboratory cycling, while real vehicles operate under complex conditions with varying temperature, charge-discharge rate and depth of discharge, creating obvious distribution discrepancies. Besides transfer learning and domain adaptation, more fundamental solutions include grouped modeling by working-condition category and explicit incorporation of covariates such as temperature and charge-discharge rate into the models, thereby reducing implicit dependence on dataset coverage.

(3) Inconsistency and pack-level estimation. Vehicle applications focus on the available capacity and safety boundary of modules or complete battery packs. Pack-level SOH is not the simple average of individual cells, so corrections combining equalization status, internal-resistance distribution and weakest-cell identification are indispensable.

(4) Computing-power constraints and real-time requirements. Main-control chips of battery-management systems have limited resources, and deep neural networks cannot be directly deployed. Model compression, fixed-point quantization and event-triggered computation should be adopted, with inference activated only during charging segments to lower the computational load.

(5) Interpretability and safety accountability. SOH outputs are applied to warranty judgment, residual-value assessment and thermal-runaway early warning. Models must provide confidence intervals and dominant influencing features and retain traceable versions and data records; otherwise they cannot satisfy the requirements of functional safety and liability identification.

This paper investigates SOH prediction for power batteries of new-energy vehicles and conducts a review from four dimensions: characterization system, aging mechanisms, methodological system and engineering framework. The main conclusions are as follows.

(1) SOH results from the coupling of multiple mechanisms and influencing factors. Capacity and internal-resistance indicators reflect battery aging from different perspectives. For vehicle-level applications, primary indicators should be determined according to service scenarios with auxiliary constraint indicators, since single indicators cannot cover the full battery lifecycle.

(2) Model-based, data-driven and fusion methods each possess applicable boundaries. Data-driven methods achieve favorable accuracy on public datasets; nevertheless, the engineering bottleneck has shifted from algorithm precision to data closed-loop performance, working-condition generalization and computing-power constraints.

(3) Mechanism-constrained data-driven models are recommended as the primary technical route, together with a cloud-edge collaborative prediction framework. Weak labels derived from charging segments and a small volume of real-vehicle calibration data are utilized for model calibration. Meanwhile, constructing public datasets covering real-world working conditions and unified evaluation specifications will lay the foundation for mass-production application of SOH prediction technologies.

5. REFERENCES

[1] LI Z H, SHI Q L, WANG K L, et al. Research status and prospect of state-of-health estimation methods for lithium-ion batteries[J]. Automation of Electric Power Systems, 2024, 48(20): 109-129.

[2] TANG K, LUO B, CHEN D, et al. The state of health estimation of lithium-ion batteries: a review of health indicators, estimation methods, development trends and challenges[J]. World Electric Vehicle Journal, 2025, 16(8): 429.

[3] YAO L, XU S, TANG A, et al. A review of lithium-ion battery state of health estimation and prediction methods[J]. World Electric Vehicle Journal, 2021, 12(3): 113.

[4] XIONG R, LI L, TIAN J. Towards a smarter battery management system: a critical review on battery state of health monitoring methods[J]. Journal of Power Sources, 2018, 405: 18-29.

[5] GU J P, JIANG L, ZHANG X S, et al. State-of-health evaluation and influencing-factor analysis of lithium-ion batteries based on feature extraction[J]. Transactions of China Electrotechnical Society, 2023, 38(19): 5330-5342.

[6] SEVERSON K A, ATTIA P M, JIN N, et al. Data-driven prediction of battery cycle life before capacity degradation[J]. Nature Energy, 2019, 4(5): 383-391.

[7] NG M F, ZHAO J, YAN Q, et al. Predicting the state of charge and health of batteries using data-driven machine learning[J]. Nature Machine Intelligence, 2020, 2(3): 161-170.

[8] HANNAN M A, LIPU M S H, HUSSAIN A, et al. A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: challenges and recommendations[J]. Renewable and Sustainable Energy Reviews, 2017, 78: 834-854.