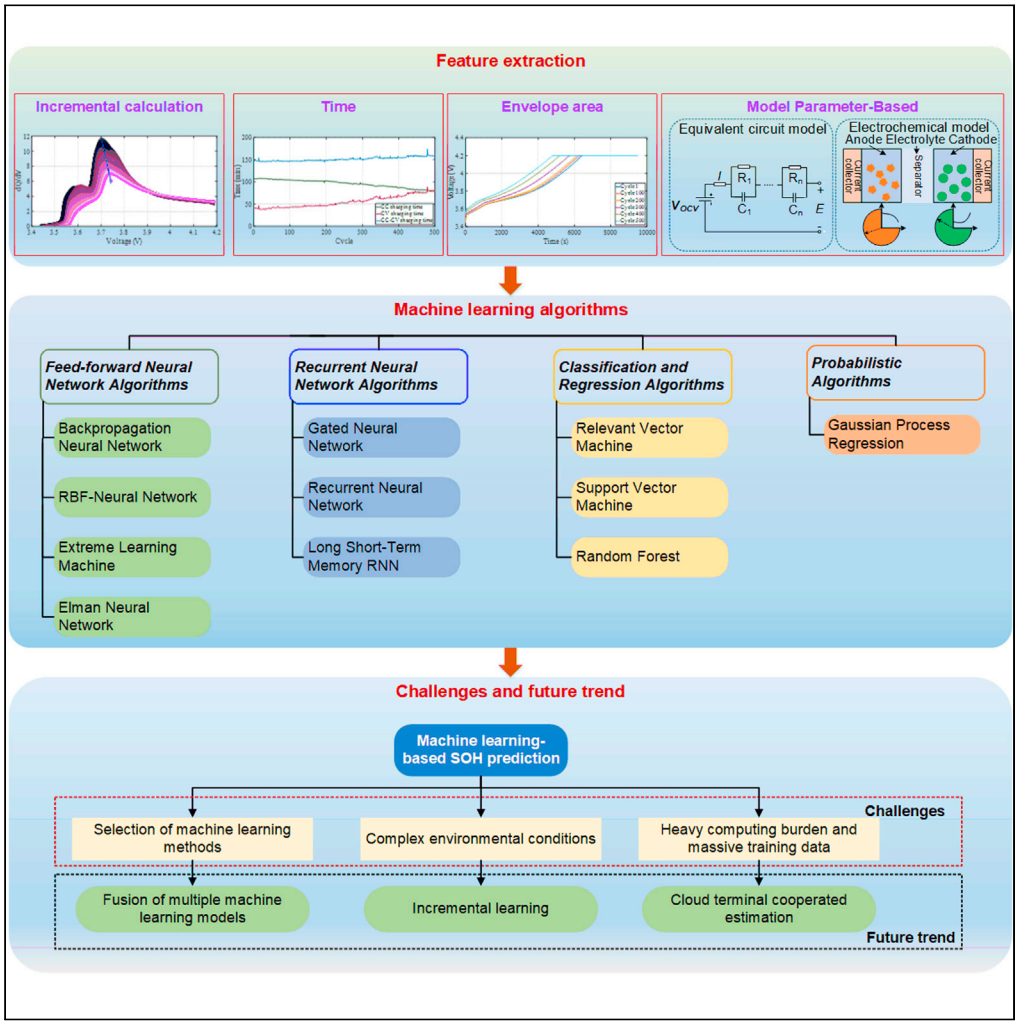
iScience

CellPress
OPEN ACCESS

Article

State of health prediction of lithium-ion batteries based on machine learning: Advances and perspectives



Xing Shu, Shiquan Shen, Jiangwei Shen, Yuanjian Zhang, Guang Li, Zheng Chen, Yonggang Liu

chen@kust.edu.cn (Z.C.)
andyliuyg@cqu.edu.cn (Y.L.)

Highlights

A full review is given for state of health estimation with limitations discussed

Existing health feature extraction methods are comprehensively surveyed

Machine learning based State of Health estimation is elaborately compared

Shu et al., iScience *24*, 103265
November 19, 2021
https://doi.org/10.1016/j.isci.2021.103265

iScience

CellPress
OPEN ACCESS

Article

State of health prediction of lithium-ion batteries based on machine learning: Advances and perspectives

Xing Shu,[1] Shiquan Shen,[1] Jiangwei Shen,[1] Yuanjian Zhang,[2] Guang Li,[3] Zheng Chen,[1,3,5,*] and Yonggang Liu[4,*]

SUMMARY

Accurate state of health (SOH) prediction is significant to guarantee operation safety and avoid latent failures of lithium-ion batteries. With the development of communication and artificial intelligence technologies, a body of researches have been performed toward precise and reliable SOH prediction method based on machine learning (ML) techniques. In this paper, the conception of SOH is defined, and the state-of-the-art prediction methods are classified based on their primary implementation procedure. As an essential step in ML-based SOH algorithms, the health feature extraction methods reported in the literature are comprehensively surveyed. Next, an exhausted comparison is conducted to elaborate the development of ML-based SOH prediction techniques. Not only their advantages and disadvantages of the application in SOH prediction are reviewed but also their accuracy and execution process are fully discussed. Finally, pivotal challenges and corresponding research directions are provided for more reliable and high-fidelity SOH prediction.

INTRODUCTION

To cope with global warming and climate change, electric vehicles (EVs) have been considered as a favorable alternative to conventional fuel vehicles (Guo et al., 2021a). Although the performance of EVs has been promoted significantly, their safety and user acceptance highly depend on the performance of energy storage systems (Zhang et al., 2021a). Currently, lithium-ion batteries dominate the energy storage solutions in EVs (Hosen et al., 2021). However, the performance of lithium-ion batteries degrades with operation, thanks to the deterioration of their electrochemical constituents, showcasing internal resistance increase as well as capacity and power attenuation (Tian et al., 2021). Battery aging can be affected by multiple coupled influences such as electrochemical reactions, mechanical stress, and operation temperature (Lee et al., 2021). To guarantee operation safety of lithium-ion batteries in their lifespan, comprehensive life state knowledge accounts for a critical role in high-performance management in EV applications (Meng et al., 2021). State of health (SOH), usually specified by the percentage of present full charge/discharge capacity over the rated value, has been widely exploited to evaluate the physical degradation level of lithium-ion batteries (Bian et al., 2021a). Usually, lithium-ion batteries are deemed as end of life (EOL) and become unqualified for EV applications when SOH drops below 80% or the internal resistance doubles (Jia et al., 2021).

As an inner state quantity that cannot be directly measured, SOH prediction methods have been widely investigated, and a body of state-of-the-art techniques has emerged (Tian et al., 2019). Even so, SOH prediction is still a bottleneck task, due to its multi-coupling characteristics of temperature, charging rate, and depth of discharge (DOD) (Shu et al., 2020d). It is, therefore, imperative to elaborate these methods and summarize these pros and cons for performance promotion in wider engineering applications (Xu et al., 2021b). Recent developments in communication and artificial intelligence (AI) technologies inspire the substantial exploitation of machine learning (ML)-based methods in SOH prediction (Pan et al., 2020). Owing to its easy application and independence on the detailed knowledge in terms of degradation mechanism, ML-based methods are widely investigated for SOH estimation promotion (Goud et al., 2021). Li et al. reviews the ML algorithms including Gaussian process regression (GPR), neural network (NN), relevance vector machine (RVM), autoregressive model (AM), and support vector machine (SVM) with respect to their viability and efficiency in SOH estimation based on real-world data (Li et al., 2019d). Ng et al. compare equivalent circuit models (ECMs) and physics-based models and outline an exhaustive survey for the application of ML techniques in battery state prediction (Ng et al., 2020). However, the ML algorithms

[1]Faculty of Transportation Engineering, Kunming University of Science and Technology, Kunming 650500, China

[2]Sir William Wright Technology Center, Queen's University Belfast, Belfast BT9 5BS, UK

[3]School of Engineering and Materials Science, Queen Mary University of London, London E1 4NS, UK

[4]State Key Laboratory of Mechanical Transmissions & College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China

[5]Lead contact

*Correspondence: chen@kust.edu.cn (Z.C.), andyliuyg@cqu.edu.cn (Y.L.)

https://doi.org/10.1016/j.isci.2021.103265

Table 1. An overview of the SOH review papers

Ref.	Content	Drawbacks
Ungurean et al., 2017	The prediction methods of SOH and RUL are introduced and evaluated.	ML-based algorithm framework and implementation factors are not discussed.
Meng and Li, 2019	Battery health management progress is expounded in a chronological order.	Only a few ML methods are discussed for SOH estimation.
Tian et al., 2020a	General review on SOH estimation methods is conducted.	The execution process of ML algorithm is not described extensively.
Xiong et al., 2018a	The algorithm and execution process of direct measurement and model-based SOH estimation methods are delivered.	ML-based algorithms are not introduced comprehensively.
Lipu et al., 2018	The prediction method of SOH and RUL are comprehensively introduced.	Some of the mentioned algorithms can be summarized as the same type method.
Li et al., 2019d	GPR, NN, RVM, AM, and SVM algorithms used for SOH estimation are reviewed.	Some ML methods are still not fully involved.
Ng et al., 2020	The promise of ML techniques used for SOH and SOC estimation are showcased.	Health feature generation process is neglected.
This study	1) The health features introduced in the literature are systematically surveyed. 2) The configuration, execution, and mathematical calculation procedure of predicting SOH based on commonly used ML algorithms are classified and elaborated. 3) The key challenges and corresponding future research trends are discussed in depth.	

elaborated in the existing literature are not comprehensive, and their efficiency and application potential are still not fully discussed. On the other hand, a critical step when leveraging ML-based methods for SOH estimation is to select proper feature variables, which, to the authors' knowledge, are seldom discussed in previous review works. Some noteworthy SOH review literature and their limitations are compared in Table 1.

To bridge these investigation gaps, a detailed review of health feature generation and ML prediction methods utilized in SOH prediction for lithium-ion batteries are reviewed in this study. By introducing the up-to-date achievements in SOH estimation, an exhausted survey in terms of direct calibration, filter-based, and ML-based methods is provided. Particularly, the ML-based method is investigated in depth, and their configuration, execution procedure, and mathematical calculations are discussed with detail. Moreover, health features, as one of the pivotal parameters during implementing ML algorithms, are systematically summarized, and their efficiency, feasibility, and viability are compared. Furthermore, the application limitation in the literature is discussed, and the future research trend is suggested from different perspectives. This review will help automotive engineers and experts better grasp and understand the existing SOH prediction methods and developing trend. In addition, the review will supply some insights that will contribute to the development of next-generation intelligent battery management system (BMS).

The remainder of this article is structured into five sections: Overview of SOH prediction methods section introduces the conventional definition of SOH and discusses the merits and drawbacks of existing SOH prediction methods. Study of feature extraction section provides a comparative insight description of feature generation methods. Machine learning-based SOH prediction section explains the fundamental steps to predict SOH with ML-based methods. The challenges and research directions in terms of the implementation of ML-based methods for SOH prediction are presented in Challenges and future trend of machine learning methods section. Finally, Conclusions section outlines the conclusions of this review study.

OVERVIEW OF SOH PREDICTION METHODS

Conventional definition of SOH

The SOH indicates the percentage of current energy/power output capability over the nominal value (Liu and Xu, 2020). Like most mechanical/electrical systems, the performance of lithium-ion battery indispensably degrades with operation. To be specific, the battery capacity declines and its resistance increases

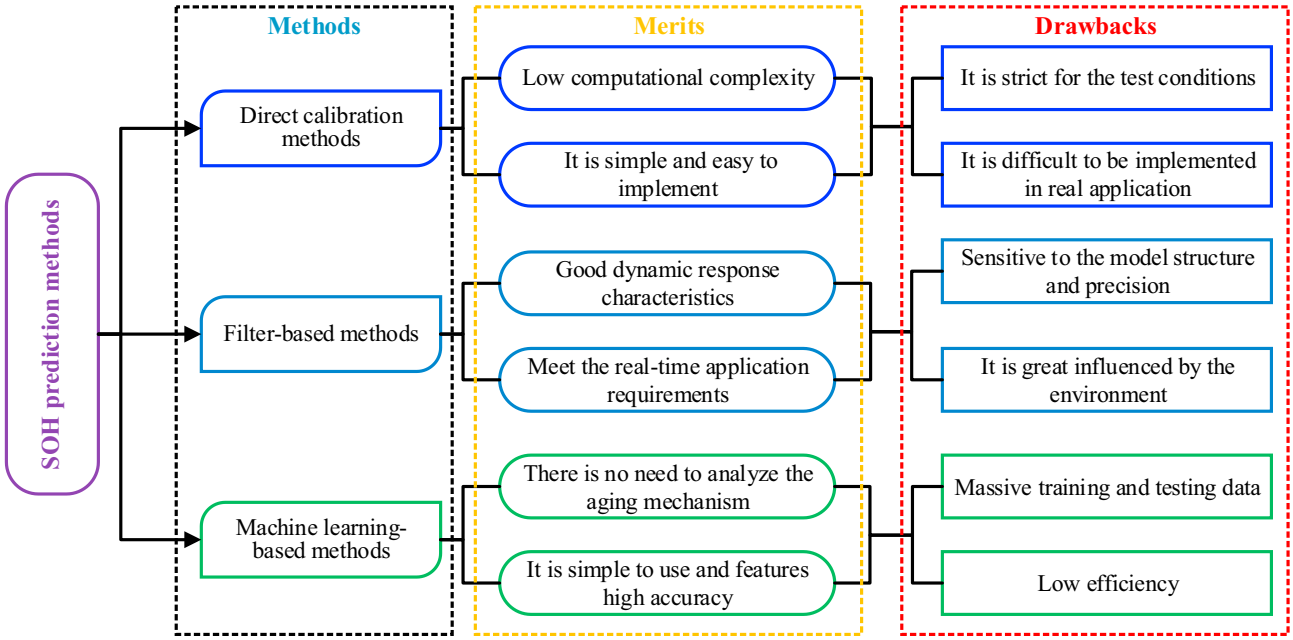


Figure 1. Classification and comparison of battery SOH prediction methods

when the battery ages, and therefore, characterizing SOH variation can be suggested with the expression of capacity and resistance change. In some applications (such as hybrid EVs), where the power capacity is more important than the energy amount, the internal resistance is often considered as an SOH metric, and it can be formulated (Vidal et al., 2020) as

$$SOH = \frac{R_{EOL} - R_{Present}}{R_{EOL} - R_{New}} \quad \text{(Equation 1)}$$

where R_{EOL}, $R_{Present}$, and R_{New}, respectively, represent the battery resistance at EOL, present, and fresh states. In this measurement, the batteries are perceived at EOL when their internal resistance becomes doubled, i.e., $R_{EOL} = 2R_{New}$ (Farmann et al., 2015). By contrast, for application scenarios (e.g., EVs), where the available energy state plays more important role, the capacity is mostly regarded for SOH measure, which can be defined as

$$SOH = \frac{Q_{max}}{Q_n} \times 100\% \quad \text{(Equation 2)}$$

where Q_{max} and Q_n denote the maximum discharge capacity at the current cycle and the nominal capacity when the battery is fresh. Note that the internal resistance can be gauged by electrochemical impedance spectroscopy (EIS) (Koseoglou et al., 2021), calculated by hybrid pulse power characterization (HPPC) test (Shu et al., 2020a), or estimated by advanced identification algorithms (such as genetic algorithm [GA] [Chen et al., 2019c] and recursive least square [RLS] [Shu et al., 2020c]). The internal resistance can be easily interfered by experimental temperature and state of charge (SOC). However, the capacity can be directly accumulated by ampere-hour (Ah) integration and thus is deemed as a more practical quantity to describe SOH, although it is contingent upon operation temperature and current. The existing definition of SOH estimation, consequently, is mainly represented by capacity variation unless specified.

Available methods of SOH prediction

For the purpose of tracking these deterioration behaviors, battery SOH perdition approaches are simply sorted into direct calibration, filter-based, and ML-based methods, as shown in Figure 1.

Direct calibration methods

Two direct calibration methods for ascertaining SOH include measuring the capacity and internal resistance. A complete discharge after full charge is needed to measure the capacity. As well known, Ah

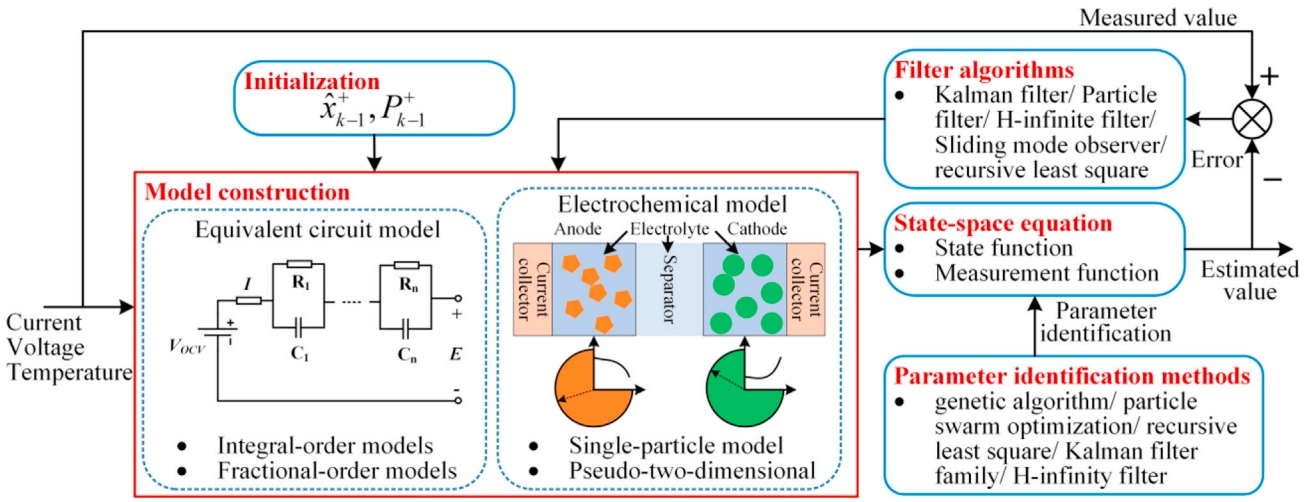


Figure 2. Generic operation framework for SOH prediction using filter-based methods

integration method is one of the most representative methods. After the battery is fully charged by constant current–constant voltage (CC–CV) protocol, the Ah integration method is hired to measure the discharged capacity (Xiao et al., 2020). Then, SOH can be calculated by (Eqution 2). On the other hand, internal resistance can also represent aging state, and it delineates the voltage drop when the load current is conducted. To measure internal resistance, pulse discharge and EIS are two mainstream methods (Zhang et al., 2020b). In the pulse discharging method, a high-rate pulse current excitation is imposed on the test battery, and the Ohm's law is employed to calculate the internal resistance as

$$R_0 = \frac{\Delta U_0}{\Delta I_0} \quad \text{(Equation 3)}$$

where ΔU_0 and ΔI_0 indicate the pulse voltage change and current variation. In frequency domain, EIS can be leveraged to measure impedance over a wide frequency range. The different frequency sinusoidal currents are stimulated, the corresponding voltages are measured, and the impedance spectrum is calculated (Islam and Park, 2020). These methods based on direct calibration are simple and easy to implement. Moreover, it exhibits low computational complexity. Nonetheless, these methods are strict for test conditions and intractable to be applied in practice.

Filter-based methods

Filter-based parameter estimation has also been widely exploited to assess the aging state of lithium-ion batteries with the help of advanced filter techniques. Usually, ECM (Song et al., 2021) and electrochemical model (EM) (Wu et al., 2021) are leveraged to mimic the dynamic and static characteristics of lithium-ion batteries, and then filters/observers with closed-loop control and feedback functionality are tailored to estimate the aging state (Naseri et al., 2020). The filter-based methods have been widely accepted for battery SOH estimation, and the generic operation framework for SOH prediction using filter-based methods is sketched in Figure 2.

To simulate the electric behaviors of lithium-ion batteries, the ECM model is proposed due to its simple structure and easy parameter identification. Currently, the ECM can be divided into integral-order models (IOMs) and fractional-order models (FOMs). The simplest IOM is the internal resistance model, which consists of an open circuit voltage (OCV) source and a resistor connected in series (Johnson, 2002). It is simple but does not account for the polarization and diffusion effect occurring inside of the battery. Yann Liaw et al. develop the first-order ECM, which is efficient in mimicking the charging and discharging behavior of lithium-ion batteries by employing one resistor-capacitor (RC) network (Yann Liaw et al., 2004). The first-order ECM can model the voltage response under different current loads, and the RC network can filter signals by passing fixed frequency information and suppressing the remaining frequency scopes (Sarrafan et al., 2020). Also, more complex ECMs, such as the second-order RC-model-based ECM, Partnership for a New Generation of Vehicles (PNGV) model (Johnson, 2002), and general nonlinear (GNL) model, are widely developed to characterize the subtle variation of battery charge and discharge behavior (Hossain et al.,

2020). More RC networks may describe more complex behaviors occurring inside of batteries, while undoubtedly complicating the model calculation and aggravating the computation burden. In practice, the first- and second-order IOM are preferred. Nevertheless, IOMs usually lack physical and chemical meanings, which encumbers the further investigation of inner mechanisms of lithium-ion batteries. One improvement of IOMs is FOM, which appends constant phase element (CPE) derived through fitting EIS to substitute the ideal capacitor in IOM (Xiong et al., 2019). Hu et al. exploit an FOM to analyze the degradation mechanism and estimate the battery SOH (Hu et al., 2018). Regarding parameter identification of ECMs, GA (Jiang et al., 2021), particle swarm optimization (PSO) (Bian et al., 2021b), RLS (Ouyang et al., 2020), and their extensions (He et al., 2020b) can be deemed as the desired solutions.

Kalman filter (KF) is one of the most widely used estimation methods and has attracted considerable attentions for model parameter and state estimation (Meng et al., 2019b). The feedback and correction mechanism are employed in KF to eliminate the noise and approach the real value continuously. Various KF-based extensions have been advanced to account for more complicated and higher nonlinear SOC estimation, and these solution representatives include extended KF (EKF) (Beelen et al., 2021), adaptive EKF (AEKF) (Shrivastava et al., 2021), unscented KF (UKF) (Marelli and Corno, 2021), adaptive UKF (AUKF) (Li et al., 2020b), adaptive sigma-point KF (ASPKF) (Sun et al., 2021), cubature KF (CKF) (Peng et al., 2019), adaptive CKF (ACKF) (Li et al., 2021c), square root CKF (SRCKF) (Shu et al., 2021a), etc. Usually, two essential steps are involved in KF-based estimation, in which the state prediction is firstly implemented to estimate the current output, and then the estimation is updated and corrected to achieve more authentic output. To reduce the computation lost, Yan et al. introduce the Lebesgue sampling method integrating EKF to predict SOH (Yan et al., 2019). However, KF-based algorithms assume that the statistical characteristics of noise are known beforehand, and in addition, they are sensitive to model errors. To tackle these shortcomings, H-infinity filter (HIF) is proposed by introducing the H-infinite norm and boundary condition. Yu et al. applies the RLS for parameter identification of the first-order ECM, and then the adaptive HIF (AHIF) is advanced for capacity estimation (Yu et al., 2019). Because SOC shows strong correlation with operating conditions such as various temperature and different SOH, SOH and SOC are usually jointly predicted to improve the accuracy and reliability by means of an interaction manner. Feng et al. hire the first-order ECM to delineate the dynamics, and the cascade terminal sliding-mode observers (SMOs) are raised to synchronously determine the model parameters, SOC and SOH of lithium-ion batteries (Feng et al., 2020b). Ouyang et al. employs a fractional-order model and UKF to estimate SOC, and the SOH is predicted by the Gaussian linear model with six commonly used OCV models (Ouyang et al., 2021). Considering the slow time-varying characteristics of SOH, multi-scale estimation is proposed in recent literature. On the basis of the first-order ECM, Song et al. exercises EKF with dual time scales to identify model parameters, SOC and SOH, and a Cramer-Rao bound analysis is addressed to quantify the effect of the current outline on prediction accuracy (Song et al., 2019). Although ECMs show less computational complexity, their accuracy needs to be further promoted. Moreover, the parameters of ECMs cannot provide the physical information of battery, such as ion concentration, current density, etc. Consequently, EMs, taking into account the motion of lithium ions, electrochemical kinetics, and material attributes, have been progressively developed (Bi and Choe, 2020).

EMs are developed based on the operation principle of batteries to delineate the internal electrochemical dynamics, and they are usually comprised of a suite of nonlinear partial differential equations (Liu et al., 2021c). The implementation principle and procedure based on EMs with filter algorithms to predict SOH are similar as those based on ECMs. Nevertheless, EMs are complex to solve and usually unfavorable for online application and state estimation. Currently, some researches are focused on state estimation of lithium-ion batteries by establishing a simplified pseudo-two-dimensional (P2D) EM (Feng et al., 2020a). Meanwhile, the single-particle model (SPM) with less computation intensity and acceptable precision has also been widely exploited in SOH estimation (Zhang et al., 2020a). Liu et al. firstly derives a nonlinear state-space function from the simplified P2D model, and then the particle filter (PF) is formulated to identify the average lithium concentration and estimate SOC and SOH (Liu et al., 2020a). Khodadadi Sadabadi et al. designs an enhanced SPM to depict the variation of parameters coupled with battery deterioration, and SOH is calibrated based on the number of mole cyclable lithium and the identified internal resistance (Khodadadi Sadabadi et al., 2021).

The simplified battery model combined with advanced filtering method features favorable dynamic characteristics, which can meet the requirement of online application. However, the filter-based approaches

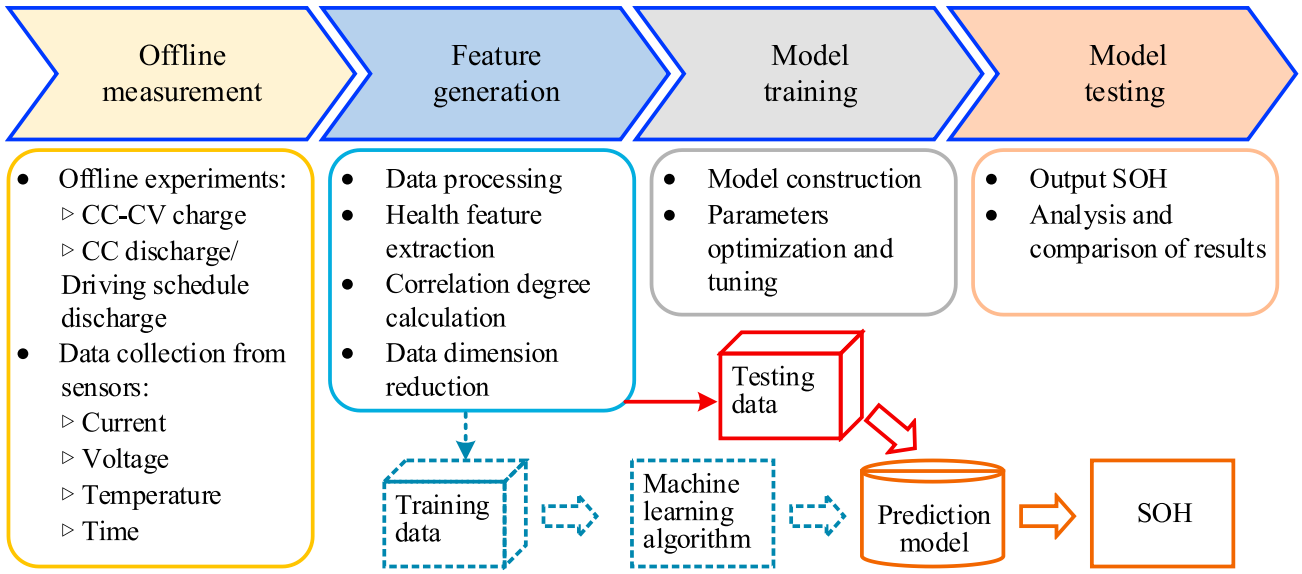


Figure 3. Generic implementation flowchart for SOH prediction using ML algorithms

heavily rely on the physical or electrochemical modeling performance of degradation behavior when portraying the descent trajectory of the battery capacity. The model is usually a set of algebraic and differential equations or empirical equations, which is more suitable for specific application scenarios. Moreover, the authenticity immensely depends on the structure and precision of the built model.

Machine-learning-based methods

With the rapid development of processing capability, data storage, and communication technologies, SOH prediction based on ML algorithms has received widespread attention (Gou et al., 2021). Figure 3 elucidates the generic implementation procedure when exploiting ML algorithms to estimate SOH, and as can be seen, it normally includes four phases: offline measurement, feature generation, model training, and model validation. In the first step, the original data, including current, voltage, temperature, and time, are measured and recorded by means of offline experiments. Next, a set of health features with k dimensions are extracted from the original measure data with d dimensions to represent the key variations during degradation; that said, $k<d$. The third step is to train the ML model with the inputs of health features and output of SOH, and the model parameters are also optimized and tuned in this step. The final step is to validate the constructed model and predict SOH. ML algorithms are efficient in SOH prediction, as they do not need to interpret the implicit electrochemical reactions occurring inside of lithium-ion batteries. Meanwhile, the prediction method of SOH by ML algorithms does not require developers to understand the material characteristics and electrochemical characteristics of battery. With the development of advanced ML algorithms, the estimation accuracy, generalization, learning performance, and convergence speed have been substantially promoted.

Next, an overview of feature extraction and ML algorithms will be provided and discussed.

STUDY OF FEATURE EXTRACTION

Feature extraction is a pivotal step to implement ML algorithms, and it can highly affect the prediction performance of SOH. More relevant and practical feature input will certainly generate more accurate predictions. Here, different feature variables for model training will be summarized, and the feature variables are excavated from four perspectives, including incremental calculation, time, envelope area, and model parameter, as demonstrated in Figure 4. The detailed procedure of these health features is expounded below.

Incremental calculation-based health feature

Incremental capacity (IC) curve is an effective manner to quantify battery capacity loss and is calculated by differentiating the voltage based on capacity variation (Schaltz et al., 2021). Thus, the relationship between capacity and voltage can be calculated as

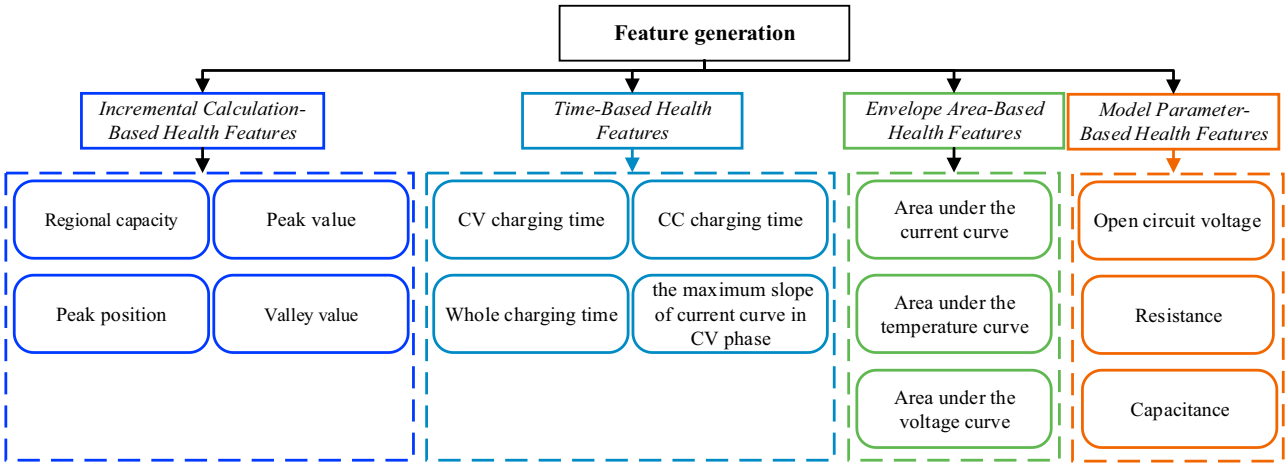


Figure 4. Summary of different health features

$$\begin{cases} Q = \int I dt \\ V = f(Q), \quad Q = f^{-1}(V) \end{cases} \qquad \text{(Equation 4)}$$

where Q, I, and V, respectively, denote the charging capacity, current, and voltage; $f(Q)$ is the mapping function from Q to V; f^{-1} means the inverse of f. The derivative of f^{-1} to V can be formulated as

$$G = \left(f^{-1}\right)' = \frac{dQ}{dV} = \frac{Idt}{dV} = I\frac{dt}{dV} \qquad \text{(Equation 5)}$$

To acquire IC curves, the voltage interval is usually set to 1 mV, and the IC curve for one new cell is plotted in Figure 5A. Owing to the sampling error and environmental interference, the original IC curves are polluted by noise, thus hindering the extraction of health features. To eliminate noise, filtering technologies, such as KF (Tang et al., 2018), moving average technique (Li et al., 2020c), Savitzky-Golay filter (Merla et al., 2016), and Gaussian filter (Zhang et al., 2019a), are usually employed. Hu et al. classifies health features into measurement features and calculation features and evaluates the effects of four noise reduction methods, including Gaussian filtering, wavelet transformation, differential filtering, and moving average filtering during data processing (Hu et al., 2021). The comparison results show that Gaussian filtering provides better elimination performance of the noise before health feature generation. Figure 5B demonstrates the IC curves filtered by KF for a nickel cobalt manganese (NCM) battery, where the variation from black to purple represents the attenuation degree of SOH. All the peaks gradually decrease and move toward higher voltage area. Furthermore, peak I disappears after 300 cycles, and only peak II exists all the time. In addition, different health features are distinguished from IC curves to characterize the aging variation of lithium-ion batteries. He et al. employs an improved Lorentzian function to fit the IC curve and sketches the voltage plateau of lithium-ion batteries (such as lithium-iron phosphate batteries), then the peak height, voltage, and area are extracted to calibrate the SOH by a linear model (He et al., 2020a). To improve the resilience against noises, Tang et al. proposes the Luenberger–Gaussian-moving-average filter to flatten the IC curves, and the observed IC peaks are applied to establish the SOH prediction model (Tang et al., 2020). Li et al. investigates the influence of different filtering methods, different charging rates, and different filtering windows to finally determine an efficient IC/differential voltage (DV) curve (Li et al., 2018a).

The published literature for SOH prediction based on IC curve is tabulated in Table 2. As can be observed, IC curve highly depends on charging current rate. When the battery is charged with a higher rate current, the peak values of IC curves will be easily overwhelmed by the overpotential. To quickly and efficiently unearth the most valued entropy-related information during charging process, differential temperature (DT) method is introduced to distinguish the effective health features (Shibagaki et al., 2018). To now, there exist two different methods to obtain the DT curve. Wang et al. acquires the DT curve through differentiating the measured temperature to voltage during the galvanostatic charge, which is called DTV and calculated according to (Equation 6) (Wang et al., 2021). On the other hand, the DT curve is roughly inferred using finite difference over I sampling time, defined as DTt and expressed in (Equation 7).

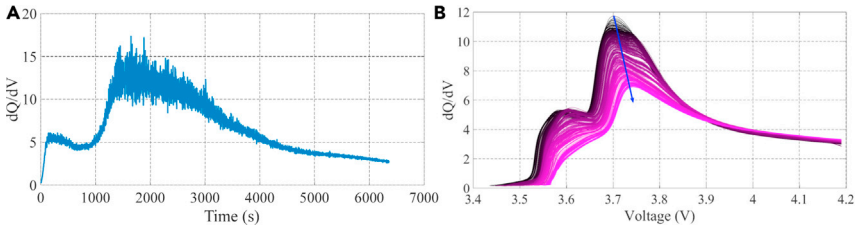


Figure 5. The evolution of the IC curves

(A) The original IC curves.

(B) The IC curves after smoothed.

$$DTV = \frac{dT}{dt} \Big/ \frac{dV}{dt} = \frac{dT}{dV} \quad \text{(Equation 6)}$$

$$DTt(k) = \frac{dT}{dt}(k) \approx \frac{T(k) - T(k-l)}{l} \quad \text{(Equation 7)}$$

where T means the surface temperature of lithium-ion batteries. Tian et al. finds that the DTt profiles during constant current (CC) charging process are highly correlated to SOH, and then partial DTt profiles are leveraged to construct a map with SOH based on SVM (Tian et al., 2020b).

Time-based health feature

Although IC analysis can account for the aging behavior from internal mechanism, it is imperative to differentiate and filter the data when obtaining IC or DT curves, increasing the difficulty of data processing. Figure 6 depicts the CC–CV charging profiles at different cycles, where the charging time for CC process decreases gradually. Inversely, the charging time for the constant voltage (CV) charging process increases distinctly. Thus, the charging time during respective CC and CV process can be abstracted as health features, as illustrated in Figure 7. Usually, lithium-ion batteries will not be fully charged or charged from empty in practice. Obtaining health features, therefore, from the partial charging curves more comply with actual situation. By this manner, the time asked to change from a low voltage to high value is considered and applied as the health feature (Tan and Zhao, 2020), and the specific calculation is formulated as

$$t_i = t_{i,V_2} - t_{i,V_1} \quad \text{(Equation 8)}$$

where t_i represents the charging duration from V_1 to V_2 at the ith cycle and t_{i,V_2} and t_{i,V_1} denote the time to reach V_2 and V_1 at the ith cycle. Richardson et al. divides the prespecified CC voltage for a duration into multiple segments, and the time charging from the beginning voltage to the homogeneous value is taken as the health feature (Richardson et al., 2018). Furthermore, the influences of duration, the lowest voltage, and the number of segments are compared. Considering that different voltage ranges will highly impact the accuracy of SOH prediction, Meng et al. exploits the nondominated sorting GA II (NSGA-II) and grid search techniques to adaptively choose the optimal multiple and single voltage ranges (Meng et al., 2019a).

Analogously, the duration when the battery current varies in the same range during the CV charging mode can also be employed as a health feature, as exhibited in Figure 8, where the duration monotonously increases with respect to cycling operation. Then, the SOH predictor is fabricated via the cycle-varying partial charging duration (Deng et al., 2019). Because the battery temperature can be also gauged directly, the time interval of isothermal increment can also be extracted to reflect the battery aging state (Tagade et al., 2020). The health feature generation method based on time slot has the advantages of easy calculation. However, the charging or discharging process needs to cover the feature extraction range. In addition, this method is highly sensitive to the current, and when the charging current changes during the cycle, the built model may not be applicable.

Envelope-area-based health feature

By observing the evolution of battery voltage, current, and temperature during the charging process, it can be found that as battery ages, the envelope area of charging curve (including voltage and temperature) gradually decreases. During the ith cycle of CC charging process, the envelope area ranging from V_1 to V_2 can be calculated as

Table 2. Health features generated from IC curves

Curve	Phase	Current rate	Health features	Ref.
dQ/dV	Charging	1C/0.75C	Regional capacity	Tang et al., 2018
dQ/dV	Charging	1/2C	Position	Li et al., 2018a
dT/dV	Charging	0.75C	Peak position, peak and valley values	Wang et al., 2021
dQ/dV	Discharging	Load profiles	Peak value	Jiang et al., 2018
dQ/dV	Charging	0.05C/0.25C/0.5C/0.75C/1C	Peak value and position	Wang et al., 2017
dQ/dV	Discharging	Load profiles	Peak value	Tian et al., 2018

$$s_{V,i} = \int_{t_{V_1}}^{t_{V_2}} V(t)dt \quad \text{(Equation 9)}$$

Similarly, during the ith cycle of CV charging process, the envelope area from I_1 to I_2 is determined by

$$s_{I,i} = \int_{t_{I_1}}^{t_{I_2}} I(t)dt \quad \text{(Equation 10)}$$

Note that the envelope area represented by (Equation 10) is equivalent to the charging capacity from I_1 to I_2. Deng et al. applies a voltage division method to equally divide the discharging voltage into N_p sections, and the accumulated capacity in each segment as well as the capacity difference of two cycles are calculated. Then, the standard deviations of the discharge capacity curve and capacity difference are extracted as the health feature (Deng et al., 2021). Xiong et al. chooses the charging capacity within a commonly used voltage range of 3.6 V–4.2 V to reflect the battery healthy state. The experimental data test with 1 C and 2 C current at 25°C and 40°C are collected to validate the variability of the proposed health features (Xiong et al., 2018b). Hu et al. repeatedly implements a five-stage CC charging and 1C CC discharging strategy in the battery pack until the pack life reaches its EOL. The charging capacities, the slopes of corresponding voltage curves, voltage drop at the current switching point, and peak voltage at the current switching point are generated as the health features (Hu et al., 2020). Nonetheless, these feature generation methods need to charge or discharge the battery for a quite long time, and the accuracy and the sampling frequency of measures have great influence on the extracted characteristic parameters.

Model-parameter-based health feature

In addition to the aforementioned methods, there exist some health features generated from battery models that have been reported in the literature. The resistance, polarization, and capacitance resistances can be detected or identified based on the measures with the help of battery model and related algorithms. Pan et al. identifies the parameters of the Thevenin model via the RLS algorithm, and the experimental results show that the resistance and polarization resistances increase almost linearly with capacity attenuation (Pan et al., 2018). What is more, this linear relationship does not depend on current loading. Hence, theses parameters can be considered as health features to quantify the degradation state. Bian et al. builds a high-order polynomial equation to describe the relationship between OCV and SOC, and an OCV-based model is constructed and verified for SOH prediction (Bian et al., 2020). Tan et al. hires the Spearman method to compare the correlation between SOH and ohmic resistance, polarization resistance, and capacitance as well as electrochemical polarization at different SOC. Finally, these parameters at SOC values of 60%, 70%, and 80% are selected as the inputs of the built SVM model to estimate SOH (Tan et al., 2021). Since these health features are obtained by constituting the battery model and parameter identification, they are not easily affected by the change of cycle current and showcase preferable adaptability. However, the model construction and parameter identification will increase the computational burden, and the model accuracy and identification algorithm will also influence the extraction of health features. In a nutshell, each feature generation method has its advantages and shortcoming, and a complementary technique should be enrolled to reveal the underlying mechanism of battery aging. To facilitate the reasonable health feature selection, these feature generation methods are summarized in Table 3.

MACHINE-LEARNING-BASED SOH PREDICTION

To now, significant progress has been achieved in predicting battery SOH based on ML. According to the configuration, execution procedure, and mathematical calculation, this review study groups these ML

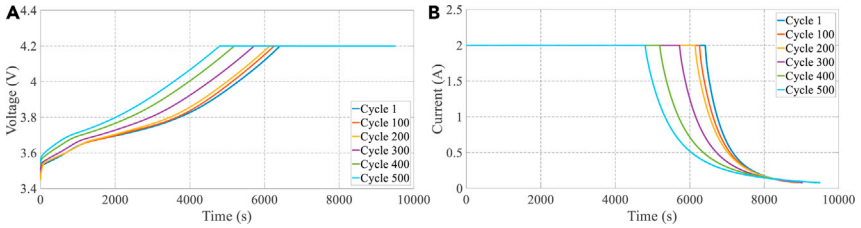


Figure 6. The CC–CV charging voltage profiles

(A) CC–CV voltage curves.

B) CC–CV current curves.

methods into four major types: feedforward NN algorithms, recurrent NN algorithms, classification and regression algorithms, as well as probabilistic algorithms, as shown in Figure 9, and the main peculiarities of these ML algorithms are compared and summarized in Table 4.

Feedforward neural network algorithms

Backpropagation neural network

Backpropagation (BP) NN is firstly proposed by Rumelhard and McClelland in 1986 (Rumelhart et al., 1986). It can be regarded as a typical multilayer feedforward NN, and its topology mainly includes an input layer, a few hidden layers, and an output layer (Houlian and Gongbo, 2018). The error back propagation algorithm is mostly widely accepted in guided learning algorithms. If the error between the predicted value and the expected output does not meet the accuracy requirement, the error will be propagated backward from the output layer, and the weight and threshold will be adjusted, so that the error between the output of the network and the expected output can be gradually converged until the accuracy requirement is met (Chen et al., 2019a). The input and output characteristics in the hidden layer can be expressed as

$$\begin{cases} S_k = \sum_{i=0}^{n} v_{k,i} x_i + \theta_{k,i} \\ z_k = f(S_k) \end{cases} \quad \text{(Equation 11)}$$

where n means the neuron number of the input, $v_{k,i}$ denotes the weight between ith input and neuron, $\theta_{k,i}$ stands for the neuron threshold, and $f(\cdot)$ is the activation function. Usually, Sigmoid function shown in (Equation 12) is selected as the activation function (Dai et al., 2019). Similarly, the input and output characteristics in the output layer can be formulated as

$$f(S) = \frac{1}{1 + e^{-\lambda S}} \quad \text{(Equation 12)}$$

$$\begin{cases} S_j = \sum_{k=0}^{q} \omega_{j,k} z_k + \theta_{j,k} \\ y_j = f(S_j) \end{cases} \quad \text{(Equation 13)}$$

where λ determines the compression of the activation function, and $\omega_{j,k}$ and $\theta_{j,k}$ denote the weight and threshold of the output, respectively. In BP NN, backpropagation refers to the back propagation of error signal and the correction of weight. If some error exists between the output and the labeled value, the error is propagated back from the output layer to the input layer and is distributed to the neurons of each layer, thereby contributing to basis correction of the weight of neurons. The backpropagation operation works through computing the gradient of the loss function with respect to each weight by the chain rule and iterating backward from the last layer to avoid redundant calculations of intermediate terms in the chain rule. This process of signal forward propagation and error back propagation is conducted repeatedly until the error of network output is reduced to an acceptable level or the specified number of the training times is reached. To optimize the initial weight and threshold of traditional BP NN, Yao et al. proposes differential evolution (DE) to reduce the iteration amount of BP NN (Yao et al., 2019). Obviously, the complexity of BP NN is not only related to the number of net layers but also close to the amount of input variables. If only one health feature with strong correlation can be leveraged to sufficiently characterize the SOH degradation, the number of layers can be reduced, and consequently the computational complexity is mitigated.

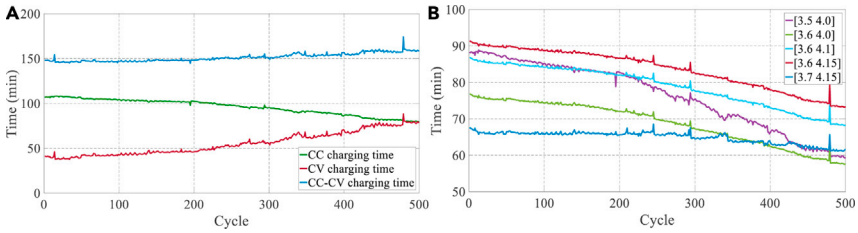


Figure 7. The time asked to change from a low voltage to higher value

(A) CC and CV charging time.

(B) Charging time for different voltage range.

Radial basis function neural network

The formation of radial basis function (RBF) NN is similar to multilayer forward network (Chen et al., 2018a). Compared with BP NN and other algorithms, it features faster convergence, and the activation function for the hidden layer is usually nonlinear; hence, the network can approximate any nonlinear function (Guo et al., 2021c). For the selection of RBF, the Gaussian function in the hidden layer node is usually employed as

$$y_i = \varphi(x, \mu_i, \sigma_i) = \sum_{i=1}^{m} \exp\left(\frac{-\|x - \mu_i\|^2}{2\sigma_i^2}\right) \quad \text{(Equation 14)}$$

where μ denotes the average and σ means the standard deviation (Sbarufatti et al., 2017). The output function of the RBF can be formulated as

$$Y = \left(\sum_{i=1}^{\infty} y_i \cdot w_i\right) + B_2 \quad \text{(Equation 15)}$$

In one study, She et al. exploits the IC analysis and Gaussian filter method to derive the peak values of IC profiles, which are applied to epitomize the aging state of batteries. Then, the RBF NN is exerted to investigate the internal relationship between peak value and aging level based on real-world operation data (She et al., 2019). In another work, Lin et al. develops an adaptive tunable RBF NN for SOH estimation of lithium-ion batteries by integrating the Brownian motion and PF without relying on physics-based models, in which Brownian motion and PF are hired to model the dynamic aging behavior of lithium-ion batteries and identify the model parameters in real time, respectively (Lin et al., 2021).

Extreme learning machine

Extreme learning machine (ELM) features faster training speed than traditional learning algorithms while insuring the estimation accuracy (Huang et al., 2006). The input weight and bias in the ELM model are set randomly to reduce the time of parameter optimization (Lipu et al., 2019). By supposing that n samples, i.e., (X_i, Y_i), are collected, where $X_i \in R^n$ means the input data and $Y_i \in R^n$ indicates the corresponding output, the output of ELM with m hidden layer neurons can be formulated as

$$t_j = \sum_{i=1}^{m} \beta_i g(\omega_i \cdot X_j + b)_i, \quad j = 1, 2, ..., n \quad \text{(Equation 16)}$$

where $g(\cdot)$ represents the activation function and a typical format is shown in (Equation 17), $\omega_i = [\omega_{i1}, \omega_{i2}, \cdots, \omega_{in}]^T$ and $\beta_i = [\beta_{i1}, \beta_{i2}, \cdots, \beta_{im}]^T$ are the weight vectors, and b_i means the bias of the ith hidden neuron (Yang et al., 2021). By randomly initializing ω_i and b_i, the optimal solution can be formulated in (Equation 18):

$$g(\omega_i \cdot \mathbf{X}_j + b_i) = \frac{1}{1 + e^{(\omega_i \cdot \mathbf{x}_j + b_i)}} \quad \text{(Equation 17)}$$

$$\widehat{\beta} = \mathrm{H}^{+}\mathrm{Y} \quad \text{(Equation 18)}$$

where H^{+} denotes the Moore–Penrose generalized inverse of H, which can be presented as

$$\mathrm{H}(\omega_1, \ldots \omega_m, x_1, \ldots x_n, b_1, \ldots b_m) = \begin{bmatrix} g(\omega_1 \cdot x_1 + b_1) & \cdots & g(\omega_m \cdot x_1 + b_m) \\ \vdots & \cdots & \vdots \\ g(\omega_1 \cdot x_n + b_1) & \cdots & g(\omega_m \cdot x_n + b_m) \end{bmatrix} \quad \text{(Equation 19)}$$

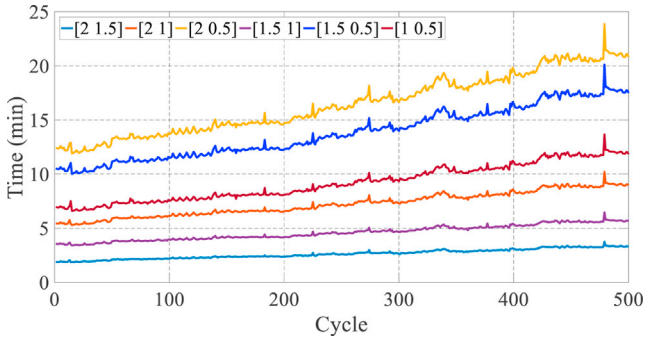


Figure 8. The time interval of equal current decrease during CV charging process

An example of successful application of ELM in predicting SOH is illustrated by Liu et al. (2020b), who leverages the ELM to map the implicit dependency between voltage variance during time interval and healthy state. The ELM can lead to better prediction accuracy with faster calculation speed and the average root-mean-square error (RMSE) of less than 0.5%, compared with the traditional artificial NN (ANN) and SVM. Chen et al. fuses the metabolic mechanism and ELM-based SOH recognition to reveal the latest aging trend in the case of small data samples (Chen et al., 2021a). The simulation results justified that the presented metabolic ELM can attain better regression than the least square SVM (LSSVM) and GPR, and the estimation precision is more accurate than those of other models. Although ELM features faster calculation and better prediction accuracy, it is a batch-based algorithm, indicating that it needs to obtain all the training data during the training phase and then conduct training and testing instead of updating online with the arrival of new data. To overcome this defect, an online sequence ELM (OSELM) is introduced for online learning and parameters update (Liang et al., 2006). The training process of OSELM contains initialization phase and sequential learning phase. In the first stage, by supposing that the number of data to be trained is N_0, and the number of hidden layer neuron nodes is L_0, the output matrix $\mathbf{H}_0$ of the initial hidden layer can be calculated as

$$\mathbf{H}_0 = \begin{bmatrix} g(\omega_1 \cdot x_1 + b_1) & \cdots & g(\omega_m \cdot x_1 + b_{L_0}) \\ \vdots & \cdots & \vdots \\ g(\omega_1 \cdot x_{N_0} + b_1) & \cdots & g(\omega_m \cdot x_{N_0} + b_{L_0}) \end{bmatrix} \quad \text{(Equation 20)}$$

According to the generalized inverse calculation method, the initial output weight β_0 can be yielded as

$$\beta_0 = D_0^{-1} H_0^T Y_0 \quad \text{(Equation 21)}$$

where $D_0 = H_0^T H_0$. In the sequential learning stage, the problem becomes minimizing $\| \begin{bmatrix} H_0 \\ H_1 \end{bmatrix} \beta - \begin{bmatrix} Y_0 \\ Y_1 \end{bmatrix} \|$. Considering initialization of training samples and newly updated data, the output weight β_1 can be calculated:

$$\beta_0 = D_1^{-1} \begin{bmatrix} H_0 \\ H_1 \end{bmatrix}^T \begin{bmatrix} Y_0 \\ Y_1 \end{bmatrix} \quad \text{(Equation 22)}$$

where

$$D_1 = \begin{bmatrix} H_0 \\ H_1 \end{bmatrix}^T \begin{bmatrix} H_0 \\ H_1 \end{bmatrix}_0 = D_0 + H_1^T H_1 \quad \text{(Equation 23)}$$

Thus, the output weight β_1 can be formulated as

$$\beta_1 = \beta_0 + D_1^{-1} H_1^T (Y_1 - H_1 \beta_0) \quad \text{(Equation 24)}$$

To account for the impact of random time-varying discharge current, Xu et al. develops an OSELM based on the beetle antenna search (BAS) algorithm (Xu et al., 2021a). To improve the online learning ability and update mechanism of OSELM, Tian et al. proposes a drift detection method based on the Bernstein inequality algorithm to guide the OSELM and save the learning time. The experimental results show that the improved OSELM can reduce the learning time by up to 88.87%, and the mean absolute error (MAE) can be limited within 1% (Tian and Qin, 2021).

Table 3. Summary and comparison of different health features

Features	Advantages	Disadvantages
Incremental calculation-based	✔ It is easy to explore the mechanism of battery aging ✔ It shows a strong linear relationship with SOH	✔ It is sensitive to current and sampling frequency ✔ It needs difference operation and filtering
Time-based	✔ Easy to acquire ✔ Lower computation complexity	✔ Fixed current ✔ The charging or discharging process needs to cover the feature extraction range
Envelope-area-based	✔ Easy to acquire ✔ Lower computation complexity	✔ It takes a long time to charge or discharge ✔ It is sensitive to current and sampling frequency
Model-based	✔ It is not affected by current load ✔ It has strong adaptability	✔ Higher computation complexity ✔ Highly depend on model accuracy and identification algorithm

Elman neural network

The construction of Elman NN is shown in Figure 10, and a one-step delay mechanism is imbedded in the hidden layer to promote the global stability and time-varying ability of Elman NN (Zhao et al., 2020). The unit of input layer accounts for only the signal transmission, whereas the output layer takes charge of weighting (Li et al., 2019a). The context layer is responsible for memorizing the output value of the hidden layer apriority, which can be considered as a one-step delay operator (Singh et al., 2021). The output of the hidden layer is connected to the input with feedback, so as to achieve dynamic modeling. The nonlinear state space function of Elman NN can be formulated as

$$\begin{cases} x_j(t) = f(\omega_1 x_c(t) + \omega_2 u(t-1)),\ j = 1, 2, \ldots n \\ x_c(t) = x(t-1) \\ y_q(t) = g(\omega_3 x(t)), q = 1, 2, \ldots, r \end{cases} \quad \text{(Equation 25)}$$

where ω_1, ω_2, and ω_3 denote the weight matrix between the context layer and the hidden layer, the input layer, and the hidden layer, as well as the hidden layer and the output layer. $g(\bullet)$ means the activation function in the output layer (Li et al., 2019c). Currently, Elman NN is widely exerted to diagnose the battery SOH and identify the *in situ* battery capacity. Chen et al. conducts the empirical mode decomposition (EMD) to firstly dispose the phenomenon of capacity recovery. Then, the autoregressive moving average (ARMA) model and Elman NN are respectively leveraged by training the subsequent time series data and the residue data to realize SOH prediction (Chen et al., 2019d). Li et al. carries out accelerated aging test to probe into the capacity degradation characteristics and hires Elman NN to perform battery capacity prognosis in real time (Li et al., 2019b).

Recurrent neural network algorithms

Recurrent neural network

For feedforward NN, the input and output dimension is usually fixed and cannot be changed arbitrarily. When processing sequential data, traditional feedforward NN lacks the ability of incorporating historical information (Shu et al., 2021b). To enable that the feedforward NNs can process the sequence data with varying length, recurrent NN (RNN) is advanced to memorize the historical information with different lengths by using neurons with self-feedback (Zhao et al., 2019). At time step t, the information from the previous layer x_t and the previous position h_{t-1} are inputted into the recurrent neuron. As such, the output is determined by not only the present input but also the data at time step $t-1$. The procedure can be formulated as

$$h_t = f(\omega_{hx} x_t + \omega_{hh} h_{t-1} + b_h) \quad \text{(Equation 26)}$$

$$y_t = f(\omega_{yh} h_t + b_y) \quad \text{(Equation 27)}$$

where ω_{hh} means the weight matrix between hidden layer and output layer at adjacent time steps. ω_{hx} stands for the weight matrix between input layer and hidden layer. b_h and b_y represent the bias parameters (Chen et al., 2021b). As a typical time series forecasting algorithm, RNN can be performed in both single and multiple steps. Compared with one-step-ahead forecasting, multi-step-ahead forecasting tasks are

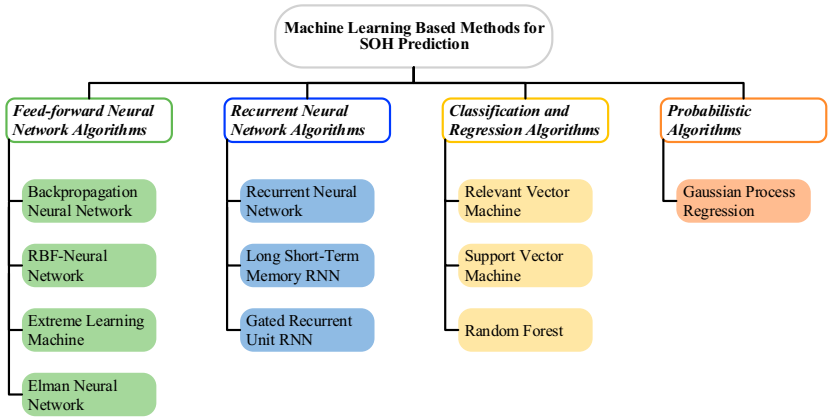


Figure 9. Classification of ML-based SOH prediction algorithms

more intractable, because this kind of forecasting needs to deal with additional complications, such as accumulation of errors, reduced accuracy, and increased uncertainty (Xu et al., 2019). A multi-step-ahead time series prediction task mainly accounts for forecasting the next L values $[y_{t+1}, \cdots, y_{t+L}]$ of a historical time series data, and $L>1$ defines the forecasting horizon. In our study, y_t means the output data at time step t, and it contains a series of output values, $y_t = [y_{t,1}, \cdots, y_{t,L}]$. Eddahech et al. harnesses RNN to predict the battery SOH of lithium-ion batteries with the consideration of EIS measure (Eddahech et al., 2012). Chaoui et al. estimates both SOC and SOH simultaneously by means of the designed RNN (Chaoui and Ibe-Ekeocha, 2017). Although RNN can capture historical information for a certain duration, the input of traditional RNN hidden layer will involve the original data information with time recursion, resulting in the loss of context information. Therefore, the range of context information that can be used in practical applications of traditional RNN structure is limited, causing the problem of gradient vanishing.

Long short-term memory recurrent neural network

To conquer the discussed shortcomings of traditional RNN, LSTM network is put forward, which contains a memory cell in its multi-layer framework (Ren et al., 2021). The difference between LSTM and RNN is that there exists only one state in a single loop unit, and four states are included in one single cycle unit of LSTM, as delineated in Figure 11 (Liu et al., 2021d). Compared with RNN, LSTM keeps a persistent unit state, which is used to decide what information needs to be forgotten or kept (Li et al., 2021a). Accordingly, the computation procedure of LSTM is presented as

$$\begin{cases} f_k = \sigma(b_f + IP_k IW_f + OP_{k-1} OW_f) \\ i_k = \sigma(b_i + IP_k IW_i + OP_{k-1} OW_i) \\ g_k = \tanh(b_g + IP_k IW_g + OP_{k-1} OW_g) \\ c_k = c_{k-1} f_k + g_k i_k \\ O = \sigma(b_O + IP_k IW_O + OP_{k-1} OW_O) \\ OP_k = \tanh(p_k) \cdot O \end{cases} \quad \text{(Equation 28)}$$

where f, i, O, and c, respectively, represent the forget, input, output gates, and memory cell. $b.$ indicates the bias of forget gate. $OW.$ and $IW.$ correspondingly denote the weights for last output and input. IP_k denotes the input of current step, and OP_{k-1} means the output of step $k-1$. p_k is an internal variable of LSTM cells; σ means the activation function, and the sigmoid function is usually selected to restrict the output value between 0 and 1; *tanh* is defined as the hyperbolic function (Tan and Zhao, 2020).

On the basis of LSTM framework, Hong et al. acquires a battery degradation model oriented for real work scenario through fitting the relationship between aging factors and different operation conditions (Hong et al., 2021). Li et al. applies the LSTM to construct the capacity prediction network under real-world operating environment, and the voltage and time samples from the raw partial charging process are chosen as the model input (Li et al., 2021b). The model's application potential is validated in an on-board management unit with different scenarios. Liu et al. employs the EMD method to decompose the original capacity information into some intrinsic mode functions (IMFs) along with a residual. Then, the GPR network is employed to match the IMFs, and the LSTM network is developed to predict the residual. Thus, the long-term

Table 4. Comparison of ML algorithms in SOH estimation

Method	Advantage	Disadvantage
BP-NN	● Flexible and simple execution ● Reasonable accuracy	● Lower operating efficiency
RBF-NN	● Good performance ● Global approximation	● Lower operating efficiency ● It is easy to fall into local optimality
ELM	● Less computation ● Faster learning speed	● Sensitive to the number of hidden neurons
Elman-NN	● Adaptation to time-varying characteristics ● Fast approaching speed	● Slow training operation ● It is easy to fall into local optimality
RNN	● Efficient for data with sequential properties	● Gradient disappearance or explosion
LSTM	● Successful tracking under long-term dependency ● Selective storage of data	● Complex training execution ● Difficultly to accelerate training
GRU	● Captures long-term sequential dependencies ● Addresses the gating mechanism in LSTM	● Needs a lot of training data ● It demands a considerable storage tool
RVM	● Better sparsity ● Free from Mercer restraint ● Avoids overfitting and underfitting	● High computational load with large datasets ● Not appropriate for long-term prediction ● Lack of stability
SVM	● Acceptable accuracy in high dimensional systems ● Quick and accurate estimation	● High complex computation ● Lack of sparseness
RF	● Higher robust ● The processing of complex data is very efficient	● Not suitable for high dimensional systems ● Sensitive to the number of random trees
GPR	● Easy covariance to generate uncertainty estimation ● Easy interpretability of features	● Sensitive to the kernel functions ● High computational cost

dependence and indeterminacy caused by capacity degradation can be quantified (Liu et al., 2021b). In addition to the above one-step-ahead prediction, Zhang et al. uses only a small amount of online data to predict the multi-step-ahead capacity based on LSTM RNN, and by this manner accurate RUL can be obtained in advance (Zhang et al., 2018). Li et al. connects the input gate and forget gate of LSTM network by a fixed connection, contributing to simultaneous determination of old information and new data. In addition, an accurate and robust multi-step SOH and RUL estimation algorithm is designed based on the variant LSTM RNN (Li et al., 2020a). However, the nonlinear and nonstationary SOH variation generated by local regenerations and fluctuations deteriorate the prediction performance of SOH. To deal with capacity regeneration, Yu addresses the EMD to decouple the global tendency and different fluctuations of SOH signals, and a multi-step-ahead modeling strategy is developed to predict real degradation tendency of the battery SOH (Yu, 2018). Although LSTM can accurately predict long series of data, its structure is complex, and many intrinsic parameters need to be tuned, making it intractable for online application.

Gated recurrent unit neural network

Gate recurrent unit (GRU) NN is an extended format of RNN based on gate control, and it integrates the input and forgetting gates of LSTM into an update gate (Yang et al., 2019). Compared with LSTM NN, its structure is simpler, and the training and prediction efficiency is higher. Figure 12 exhibits the general configuration of GRU NN. As can be found, the former hidden state h_{t-1} and the current input x_t indirectly influence the current hidden state h_t, similar as that in traditional RNN (Hannan et al., 2021). Inversely, the GRU NN employs "update gate" to determine how much previous information should be passed and "reset gate" to determine how much information should be discarded (Chen et al., 2019b). The mathematical formulation of GRU NN can be presented as

$$\begin{cases} z_t = \sigma\left(W_z x_t + W_z h_{t-1} + b_y\right) \\ r_t = \sigma\left(W_r^x x_t + W_r^h h_{t-1} + b_r\right) \\ \tilde{h}_t = \tanh\left(W_h^x x_t + W_h^h (h_{t-1} \cdot r_t) + b_h\right) \\ h_t = z_t \cdot \tilde{h}_t + (1 - z_t) \cdot h_{t-1} \end{cases} \quad \text{(Equation 29)}$$

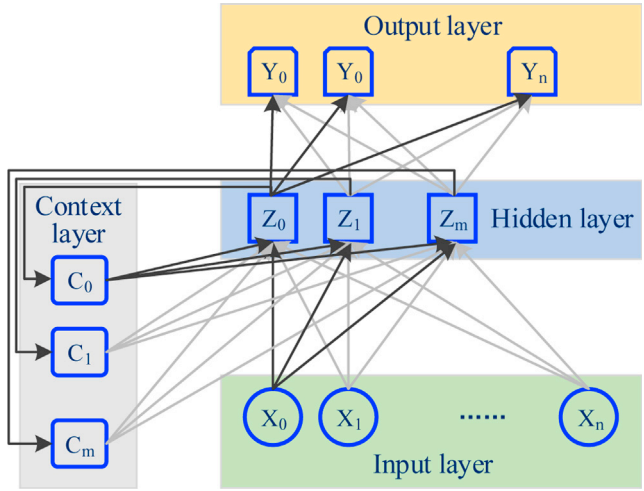


Figure 10. Schematic diagram of Elman NN

Compared with LSTM NN, GRU NN has fewer parameters and leads to faster training speed in the case of large amount of data. Ungurean et al. devises an online SOH determination algorithm based on GRU NN (Ungurean et al., 2020). The comparisons results show that the performances of GRU network are similar with that of LSTM, while about 25% parameters need not to be tuned. Because of the drastic merit in shortening the training time, GRU NN becomes a promising solution when it comes to predict SOH. Fan et al. proposes a hybrid NN, including GRU NN and convolutional NN (CNN), to learn the shared information and time dependencies of charging voltage variation and SOH, and the MAE is restricted within 4.3%, manifesting its effectiveness (Fan et al., 2020). To delineate the uncertainty of battery deterioration and avoid overfitting, Wei et al. establishes a model combining Monte Carlo dropout technique and GRU NN to forecast battery capacity with the input of constant voltage charge duration (Wei et al., 2021).

Classification and regression algorithms

Support vector machine

SVM shows high efficiency in solving nonlinear and high-dimensional model fitting and classification with less samples (Chen et al., 2018b). The SVM algorithm deals with the linear inseparability of sample data based on kernel functions (Roman et al., 2021). It maps the vector to a higher dimensional space, separates the data on both sides of the hyperplane by determining two parallel hyperplanes, and maximizes the distance between the two parallel hyperplanes (Cai et al., 2020). The related nonlinear problem of SVM can be formulated as

$$\begin{aligned} &\min_{w,b} = \frac{1}{2}\omega^T \cdot \omega \\ &s.t. \quad y_i - \omega \cdot x_i - b \le \varepsilon \\ &\qquad \omega \cdot x_i + b - y_i \le \varepsilon \end{aligned} \qquad \text{(Equation 30)}$$

where ε is toleration deviation, the coefficients ω and b can be solved by introducing Lagrange multipliers α_i and α_i^*, and the approximation function can be expressed as

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(x, x_i) + b \qquad \text{(Equation 31)}$$

Current health diagnostic methods primarily employ SVM as a regression technique for continuous values, and the process is known as support vector regression (SVR), which is attained by searching for a minimum margin fit instead of a maximum margin classifier. For instance, Feng et al. identifies the support vectors of SVR according to the charging statistics of fresh batteries (Feng et al., 2019). The implementation of the developed method only needs 15-min charging data, making it applicable for on-board SOH diagnosis. In another example, Guo et al. reconstructs the battery terminal voltage profiles based on a linear interpolation method to eliminate the noise, and avoid extraction of wrong health features from IC curve, then an

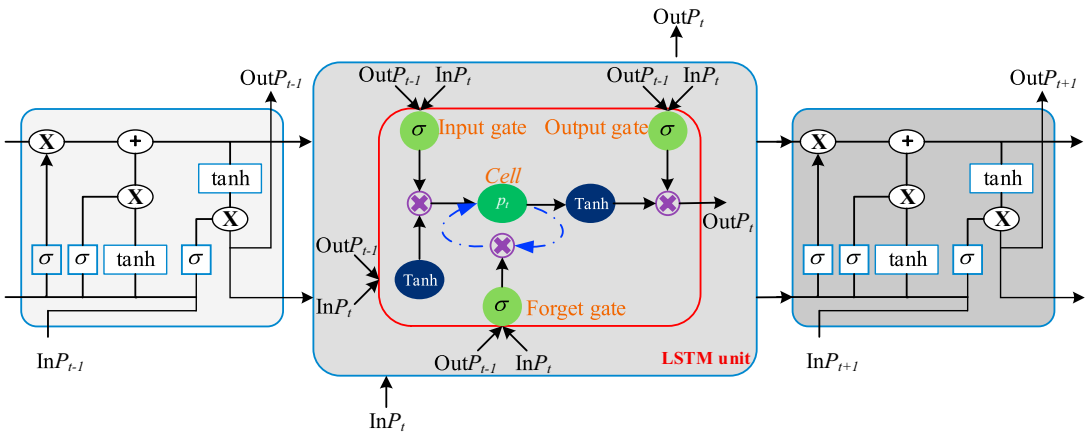


Figure 11. Schematic diagram of LSTM

SVR model is built to map the health feature and the battery SOH (Guo et al., 2021b). To alleviate the problem of intensive calculation when facing with a large amount of sampling data, the least squares SVM (LSSVM) is advanced to convert the quadratic programming problem into a linear programming problem. Yang et al. establishes an improved Thevenin model to explore the affinity between internal parameters and battery state (Yang et al., 2018a). Then, a PSO-LSSVM method is leveraged to supply accurate and reliable SOH prediction, wherein the PSO is harnessed to search the optimal model parameters of LSSVM and improve the fidelity of global optimization. Although the LSSVM algorithm has gained massive progress, it lacks critical criterion in the process of support vector selection, making the algorithm lack of sparsity and robustness. Shu et al. amplifies the quadratic Renyi criterion and applies it to screen support vectors; the comparison results indicate that the proposed framework can control the SOH estimation error within 2%, outperforming that by the traditional NN, SVM, and LSSVM models (Shu et al., 2020b).

Relevant vector machine

Relevant vector machine (RVM) is a sparsely supervised probability learning algorithm based on Bayesian estimation. It incorporates the structure of prior parameters with auto-correlation decision and can remove irrelevant data points, so as to obtain the sparse probability model (Tipping, 2001). As such, the RVM can compensate the shortcomings of SVM and provide a probability explanation for optimal solution (Chang et al., 2017). In RVM, the nonlinear probability model between input and output is presented as

$$\begin{cases} y(x,\omega) = \sum_{i=1}^{N} \omega_i k(x,x_i) + \omega_0 \\ t_i = y(x,\omega) + \varepsilon_i \end{cases} \quad \text{(Equation 32)}$$

Hu et al. considers the RVM as a probability kernel regression algorithm to map the implicit correlation of the battery capacity on healthy features that are generated from the measured charging current and voltage (Hu et al., 2015). To prevent overfitting and improve the generalization of RVM, Guo et al. utilizes principal component analysis (PCA) method to reduce feature dimensionality and combines PSO and RVM to find the optimization hyperparameters during model training. The proposed PSO-based RVM framework can limit all the relative error within 2% during the operating temperature of 24°C–43°C (Guo et al., 2019). Jia et al. constructs the RVM model to predict health state with the multiple inputs of current and voltage sample entropy, and the sudden and unusual change of capacity degradation is preprocessed using the wavelet denoising method (Jia et al., 2021).

Random forest

Random forest (RF) algorithm is a progressive method based on decision tree, and it can be used in both classification and regression (Zhang et al., 2019b). The generation structure of RF is presented in Figure 13, and it works by generating multiple classifiers/models, each of which learns alternatively and makes prediction independent. These predictions are finally combined into a single prediction based on the bagging algorithm, and on this account it performs better than a single classification prediction (Liu et al., 2021a).

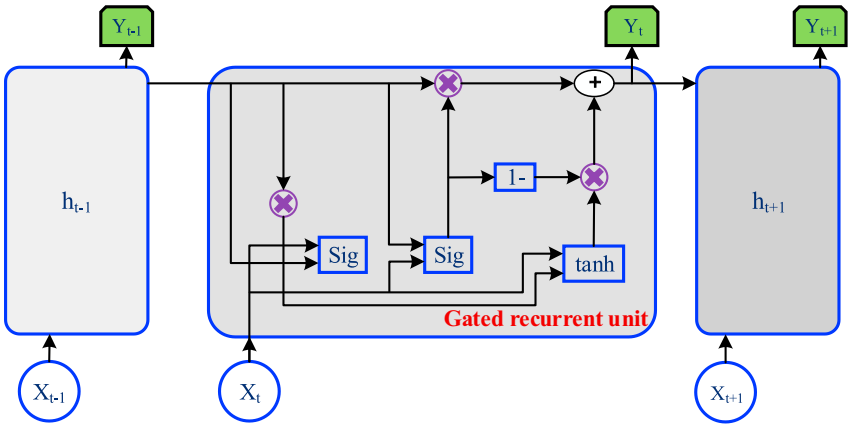


Figure 12. Schematic diagram of GRU unit

For each tree's construction, partial original sample data are selected in the bagging process, and the remaining samples are deemed as out-of-bag (OOB) samples to assess the performance of regression tree. By this manner, the problem of overfitting can be drastically mitigated, and the generalization capability of the RF can be improved by the built-in validation characteristics.

The RF is easy to train and robust against outliers, yet the delivered function is often discrete rather than continuous. In an earlier study, Li et al. exerts the RF algorithm to capture the capacity deterioration performance of lithium-ion batteries. The model is examined by different number of trees and is proved more accurate than GPR (Li et al., 2018b). A recent study excavates two healthy indicators from the operation data collected through the driving, charging, and parking missions of a number of EVs. Then, a data-driven SOH predictor based on RF algorithm is developed after incorporating the behaviors of end users and ambient conditions. The experimental results demonstrate that the developed predictor leads to precise SOH prediction with the maximum error of 1.27% (Mawonou et al., 2021). Considering the slow variation and weak predictability of the battery degradation, Zhang et al. designs a hybrid prediction model combining RF, artificial bee colony (ABC), and general regression NN (GRNN) to rank the importance of health features and predict battery degradation state (Zhang et al., 2021b).

Probabilistic algorithm

Gaussian process regression

GPR is a nonparametric model, which can attain state prediction under Bayesian framework (Lyu et al., 2021). Unlike other regression algorithms such as NN, GPR adopts the statistical mechanism to grasp the learning process (Che et al., 2021). Richardson et al. manifests the strong feature abstraction ability of GPR in forecasting battery capacity (Richardson et al., 2017). Zhou et al. constructs an SOH prediction network by using the actual diverse data stemming from different batch batteries, and the GPR method is developed to evaluate the credibility of the estimation results (Zhou et al., 2021). Chehade et al. proposes a collaborative GPR method that integrates the capacity data from all the cells to capture both the correlations among cycles as well as their cross-correlations (Chehade and Hussein, 2020). Yang et al. fabricates a GPR framework to forecast SOH over the entire cycle life, and the time of CC mode duration, the time of CV mode duration, the slope of the curve at the end of CC charging stage, and the vertical slope at the corner of the CC charging curve are considered as the input of GPR model (Yang et al., 2018b).

Integrated learning

Integrated learning aggregates the outputs of diverse learning algorithms to reach a final decision. The main objective of integrated learning is to reduce the risk of choosing a single learning algorithm with poor performance and promote the performance of individual algorithm by employing an intelligent ensemble of different individual algorithms. It has been widely used in financial crises prediction (Kim and Sohn, 2012), land-cover mapping (Kavzoglu and Colkesen, 2013), and image classification (Han and Liu, 2015). Integrated learning combined with ML has been demonstrated to achieve satisfactory performance in RUL prediction. However, the utilization of integrated learning for SOH estimation is still lacked.

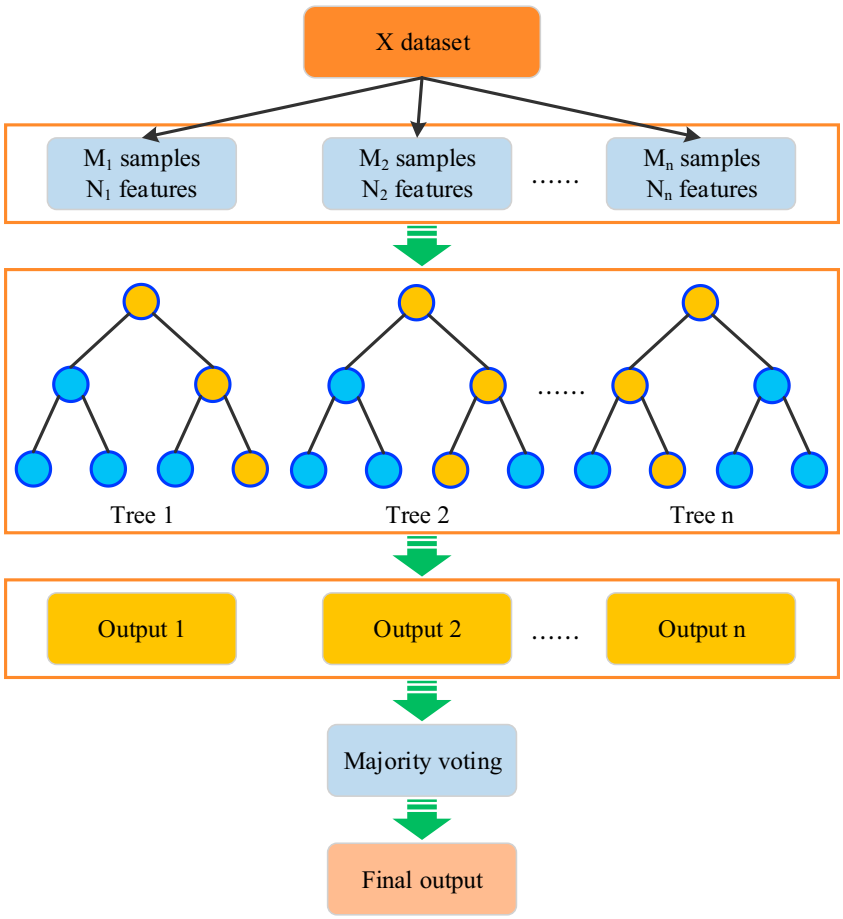


Figure 13. Schematic diagram of RF

Actually, random forest is one of the integrated learning methods, and it works by generating multiple classifiers/models, each of which learns and makes predictions independently (Lee et al., 2020). These predictions are finally integrated into a single prediction based on the bagging algorithm, and on this account, it performs better than individual prediction. In addition to random forest, Shen et al. applies eight deep convolutional NN (DCNN) to pretrain the source model based on 10-year daily cycling data. Then, the learned parameters of the DCNN models are transferred from the source task to the target task via the transfer learning algorithm (Shen et al., 2020). These models are then integrated to build a synthetic learning model for capacity estimation. Li et al. introduces an integrated learning method by producing differential data samples and composing the output of a series of NNs to estimate the SOH of lithium-ion batteries (Li et al., 2019e). Gou et al. designs an integrated model by fusing ELM and random vector functional link NN (RVFLNN) to track the degradation trend, and the nonlinear autoregressive (NAR) algorithm is advanced to reduce the RUL prediction error of each learning model (Gou et al., 2020). In the commonly used integrated learning methods, once the weights of each learner are assigned, they remain unchanged during the prediction process. Nevertheless, the prediction performance of individual learners may fluctuate in different periods. On this regard, Cheng et al. introduces an induced ordered weighted averaging (IOWA) operator to attain time-varying weight allocation of RBF NN, gray model (GM), autoregressive integrated moving average (ARIMA), and SVM. By summing the weighted prognostic results of each member, the integrated estimation of battery capacity is finally achieved (Cheng et al., 2020).

Based on the earlier discussion, the state-of-the-art research output in terms of SOH estimation based on ML methods is expounded. The detailed synopsis of the evaluation performance based on ML methods is tabulated in Table 5, and it can be found that more and more ML-based methods are devoted to SOH

Table 5. The detailed synopsis of the prediction results for the SOH based on ML methods

Method	Ref.	Health feature	Feature sources	Index	Precision
BP NN	Yao et al., 2019	Battery pack charge cutoff total voltage; discharge cutoff total voltage; battery internal resistance	Charging and discharging	MAE	5%
RBF NN	She et al., 2019	The peak values of IC curves	Charging and discharging data	Average RMSE	0.5%
	Lin et al., 2021	Kullback–Leibler distance based on the kernel density estimation and hidden Markov model	Discharging	Average RMSE	0.75%
ELM	Liu et al., 2020b	Voltage variance during equal time interval	Discharging	MAE	2.5%
	Chen et al., 2021a	The increment of ohmic internal resistance and polarized internal resistance	Discharging	MAE	1.93%
	Xu et al., 2021a	Discharge time; CC charging time; charging capacity; the minimum value of IC; the maximum of the temperature change curve	Charging and discharging	MAE	0.4%
	Tian and Qin, 2021	Discharge time, voltage upward appreciation after discharge, variance of a random walk cycle voltage curve	Random discharge data	MAE	1%
Elman NN	Chen et al., 2019d	CC mode duration	CC charging	RMSE	0.95%
	Li et al., 2019b	The known charged capacities in past cycles	Charging	MAE	2.3%
LSTM	Hong et al., 2021	Pack voltage, brake pedal stroke value, motor speed, vehicle speed, current, SOC, average temperature, the maximum temperature, cell voltages	Dynamic driving	RMSE MAE	0.23% 0.13%
	Li et al., 2021b	Voltage and time samples	Charging curves	MAE	4.26%
	Li et al., 2020a	Voltage, temperature, current, time	Discharging	Average RMSE	2.16%
GRU	Ungurean et al., 2020	Battery capacity at cycle t	SOC	MAE	2.91%
	Fan et al., 2020	Voltage, current, temperature	Charging curve	MAE	4.3%
	Wei et al., 2021	Charging voltage saturation time	Charging curve	RMSE	1.41%
SVM	Feng et al., 2019	15-min charging segment	Charging curve	MAE	3%
	Yang et al., 2018a	The parameter identification results of the Thevenin model	Dynamic cycle	RMSE	2%
	Shu et al., 2020b	Duration of voltage ranges	Charging	MAE	2%
RVM	Hu et al., 2015	Initial charge voltage; CC and CV charge capacity, final charge voltage and current	Charging curve	RMSE MAE	0.92% 2.07%
	Guo et al., 2019	Capacity, charge time, temperature and current/ voltage decline at CV/CC phase	Charging curve	RMSE	4.3%
RF	Li et al., 2018b	Charging capacity in a specific voltage range	Charging curve	RMSE	1.3%
	Mawonou et al., 2021	Mileage; discharge energy, histogram of the daily covered distances, histogram of average driving speed per driving event, SOC variation, charging temperature, total number of days	Driving; Charging; Parking	MAE	1.27%
GPR	Yang et al., 2018b	Time of CC mode duration, time of CV mode duration, the slope of the curve at the end of CC charging stage, the vertical slope at the corner of the CC charging curve	Charging curves	RMSE	6%
	Liu et al., 2019	Historical capacity, storage SOC level, temperature	NA	RMSE	1.05%
Integrated learning	Shen et al., 2020	Voltage, current, and charge capacity values	Charging curves	RMSE	1.503%
	Li et al., 2019e	Abscissa of the peak point of the IC curves, the equal voltage drops, the area near the peak point, the ratio of CC mode, the ordinate of the peak point of IC curves, the sample entropy of discharging voltage	Charging and discharging	RMSE	2.84%
	Gou et al., 2020	Duration of equal charging voltage difference	Charging curves	RMSE	0.8%

estimation based on real operation data, and the MAE and RMSE can be respectively restricted within 3% and 2%. Moreover, most of the features in these applications are extracted from the charging process. In actual operations of EVs, lithium-ion batteries are discharged stochastically, and the operation cycle is determined by the driver and road information. It is intractable to encounter an ideal regular discharge scenario and therefore is difficult to extract the health features related to SOH from the discharging operations. On the contrary, when EVs are charged, the batteries can be in a regular and deterministic charging process, thus contributing to online SOH estimation. In short, to apply these ML algorithms based on real-world battery data, it is critical to extract useful health features.

CHALLENGES AND FUTURE TREND OF MACHINE LEARNING METHODS

Although ML-based methods have delivered momentous contributions toward accurate SOH prediction, they also encounter a number of challenges and have long way to achieve high-fidelity SOH estimation. The challenges and suggested resolution perspectives for ML-based methods will be discussed in this section.

Selection of machine learning method

As a typical electrochemical system, the performance degradation of lithium-ion battery is related to not only current state of use but also historical operation information. The capacity deterioration of lithium-ion batteries in aging process is usually a time series problem with the number of cycles as the input. To forecast SOH more accurately, it is a reasonable scheme to process the cycle life of batteries based on time series data. GPR, RNN, LSTM, and GRU are essentially time series prediction techniques. Among them, RNN, LSTM, and GRU belong to nonprobabilistic approaches and can furnish only a predicted point in model regression. However, on account of the uncertainties from distinct operation data sources, including measure noise, state estimation error, and load uncertainty, capturing the prediction uncertainty of conditional distribution remains challenging. The GPR is developed in terms of the normal distribution principle and can give probability density function (PDF). It enables to detect the rule by processing the variation of training data, estimate and generate prior distribution using training data, and calculate posterior distribution function based on the Bayesian principle, thereby attaining the estimation of posterior distribution. When GPR model is applied to practical problems, it can output the mean value and give the confidence interval at the same time, so that the effectiveness of the prediction results can be enhanced. On this account, probabilistic estimation algorithms are more preferable for SOH prediction. For model training complexity, it not only depends on the complexity of the designed ML algorithms but also is related to the dimension of inputted health features. When one health feature with strong correlation with SOH is found, a reliable SOH value can be obtained through designing a simple ML algorithm. On the contrary, if the correlation is weak, a complex ML algorithm may not attain the desired SOH estimation, while increasing the computational intensity of BMS. On the other hand, when the input variables are kept the same, the model complexity also affects the training time. For instance, hundreds of epochs are essential to achieve accurate SOH estimation when a traditional RNN is trained and exploited. However, to reach the same accuracy, the number of training epochs required by the LSTM RNN is usually less than 50 (Zhang et al., 2018). Thus, compared with LSTM RNN, more training duration is entailed for traditional RNN. Nonetheless, in contrast to GRU RNN, LSTM RNN includes input gate, output gate, and forgetting gate, whereas GRU RNN includes only reset gate and update gate. The parameters of GRU RNN that need to be adjusted are less than those of LSTM RNN, thus the training expense is cheaper. In addition, GPR provides a flexible framework for probabilistic regression and is widely used to solve the high-dimensional, small-sampling nonlinear regression problems. However, the nonparametric characteristics of GPR directly leads to larger amount of calculation. In the training process, the hyperparameters are generally obtained through the optimal edge likelihood method. Each gradient calculation needs to conduct the inversion operation of the covariance matrix. When dealing with large dataset, the amount of calculation will become a bottleneck to restrict the application of GPR (Chen et al., 2020). In practical applications, the training time of constructing an ML algorithm with acceptable performance may take up to tens of minutes. Nonetheless, the SOH estimation of lithium-ion batteries is forecasted once per cycle at most, and the interval time is much longer than the training time. Hence, it is feasible to conduct online regular training of ML algorithms to attain accurate prediction of SOH. To shorten the model training time, the number of health feature selection can be reduced during the experiment. In addition, some dimensionality reduction methods, such as principal component analysis (PCA) (Fei et al., 2021) and local linear embedding (LLE) algorithm (Hong and Zeng, 2021), may also be an effective manner to reduce the dimension of health features and correspondingly shorten the model training time.

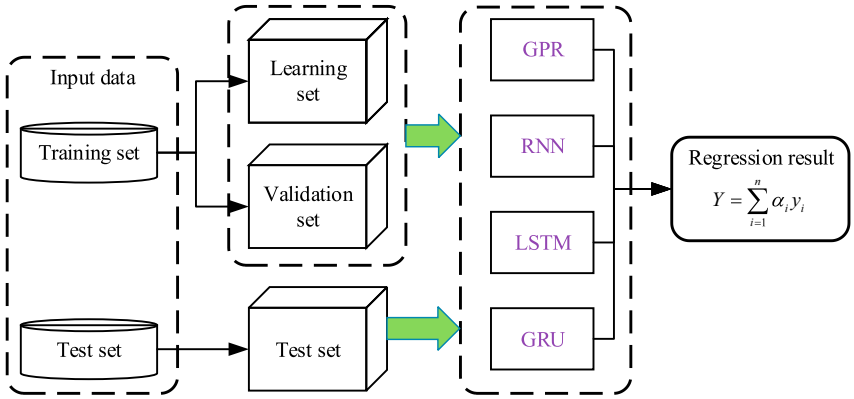


Figure 14. The fusion of multiple ML models

However, each ML algorithm has its pros and cons; how to choose the best forecasting model is a permanent gambit. Fortunately, the fusion of multiple ML models can often improve the overall prediction performance, as described in Figure 14. In short, model fusion can improve the final prediction ability more or less, and different sub-models have different approximation ability on different aspects (Aykol et al., 2021). Thus, the mutually beneficial functionality of each sub-model can contribute to a better estimation of SOH together.

Incremental learning

Current SOH estimation algorithms of lithium-ion batteries mainly consider the effectiveness of estimation under specific temperature and charging and discharging conditions. These experiments are usually conducted under laboratory conditions, which are difficult to simulate under time-varying dynamic operation conditions. To be concrete, the complicated operation conditions and operating environment of EVs lead to wide temperature variation range and large current fluctuation, which can be remedied by two manners: (1) sufficient training data preparation under different environments and (2) incremental learning. In the first way, model training needs to involve as many operation conditions as possible, and the model is often complex and inefficient. However, to improve model training efficiency and real-time applicability of SOH estimation, it is preferred to reduce the training data size and simplify the model complexity. Although, incomplete data selection often leads to low reliability and awkward estimation performance of the preselected algorithm; by contrast, when more feature data are selected, the operation speed will slow down, and the model training will be more time-consuming. Moreover, a large number of data will undoubtedly increase the offline test time and entail more experimental resources. How to trade off the algorithm complexity and SOH estimation performance needs to be further investigated.

Most of these algorithms are in batch learning mode; that said, it is assumed that all training samples can be obtained at one time before training. After learning these samples, the learning process will terminate, and no new knowledge will be imported. However, in practical applications, the training samples are difficult to be sufficiently prepared for one time, and the information reflected by the samples may gradually change over time. If new samples containing more valuable information arrive, the well-trained ML algorithm needs to learn all the data again. Thus, the training process will consume more extra time, and hence the batch learning algorithm may not meet this demand. By contrast, the incremental learning algorithm can update the knowledge and modify and strengthen the previous knowledge gradually. By this manner, the updated knowledge can adapt to the newly arrived data without learning all the data again. It is mainly manifested in two aspects: on one hand, as it does not need to save historical data, it reduces the storage space, and on the other hand, incremental learning can make full use of the historical training results in the current sample training, which significantly reduces the expense of subsequent training (Losing et al., 2018). Thus, the incremental learning is another way to mitigate the problem of data size, and the schematic diagram is depicted in Figure 15. As battery degrades, more experimental data will be collected, including stochastic loading information and fickle ambient temperatures. These historical data can be in turn exploited to train and refine the original model. Through rolling optimization and incremental learning, the training dataset will become more comprehensive, and the adaptability of the SOH estimation will be gradually promoted.

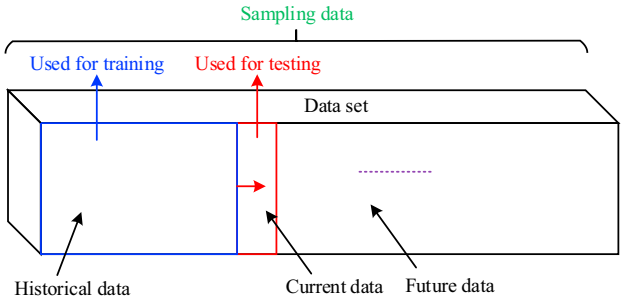


Figure 15. Schematic diagram of incremental learning

Numerous incremental and online algorithms have been spurred by adding the incremental setting to existing batch methods, such as online sequential extreme learning machine (OS-ELM) (Liang et al., 2006), online random forest (ORF) (Pernici and Bimbo, 2014), incremental support vector machines (ISVM) (Losing et al., 2018), self-organizing incremental neural network (SOINN) (Shen and Hasegawa, 2008), and Hoeffding tree (Domingos and Hulten, 2000). These incremental learning algorithms can be considered as potential schemes that can reduce the data size and computation labor of model training for SOH prediction.

Cloud-terminal cooperated estimation

ML methods need to find the high-dimensional nonlinear relationship between input and output and entail considerable calculation and storage room to existing BMSs. With the development of communication technologies, more operation data can be wirelessly transmitted to the cloud monitoring platform. In this context, more complicated and advanced ML methods can be developed in cloud platform due to its strong computing power and timely data acquisition. One typical cloud and on-board BMS-based hierarchical SOH estimation framework is shown in Figure 16, where the measure information, including current, voltage, and temperature, are collected and some immediate management tasks, such as safe control are achieved by on-board BMS. Then, the battery signals are sent to the cloud control system. By this manner, the EV management platform can easily acquire the operation data of lithium-ion batteries and track the use process of the battery in an all-round way, thereby overcoming the limitation of on-board SOH estimation and providing strong computation, massive storage, and viable SOH prediction. In addition, different battery types, different temperatures, and different operating condition data can be obtained to expand the battery operation database through multi-EV monitoring. Through the cloud-terminal cooperative estimation framework, the data or features of other similar batteries can be learned to facilitate model construction and remedy the information loss caused by the limited on-board data. It will break through the resource limitation of traditional embedded hardware terminals and make full use of immense computing capacity and massive storage capability. Thus, the enabling ML-based algorithms can be easily implemented to supply high-quality SOH estimation. To conclude, it can be presumed that online SOH estimation based on cloud-terminal cooperation will become the future direction.

Conclusions

This study furnishes an opportune and comprehensive review for the development of battery SOH prediction methods, with the focus on recent exploitations based on ML technologies. The overview, merits, and defects of SOH estimation approaches, including direct calibration, filter-based, and ML-based methods, are sufficiently compared and analyzed. ML methods are widely leveraged for SOH prediction, as they do not need complex physical model and conduct well in modeling highly nonlinear dynamic systems. As a key step in ML algorithms, health feature generation methods reported in the literature are intuitively classified, and the features of incremental calculation, time, envelope area, and model parameters are discussed. Furthermore, the detailed discussions in terms of the state-of-the-art ML-based SOH prediction algorithms, configuration, and execution process are elaborated. Finally, some challenges and perspectives for future development of ML-based SOH prediction techniques are provided. Manifold discerning and noteworthy research directions are discussed with regard to promoting adaptation, extendibility, accuracy, and reduction of training data. In a nutshell, ML-based SOH prediction methods will become one of the research focuses and will generate a profound influence on shaping future transportation electrification. Moreover, the key conclusion and

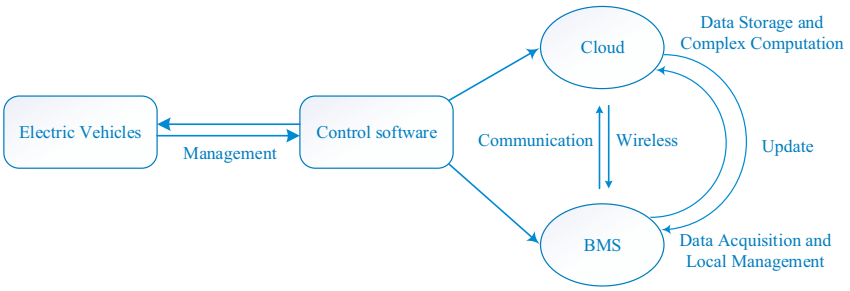


Figure 16. Cloud-terminal cooperate estimation framework

analysis summarized by this review would be momentous for EV applications incurred by the evolution of ML applications and would furnish an explicit suggestion to researchers and manufacturers on the advancement of EVs.

Limitations of the study

In this paper, we mainly focus on the retrospect of SOH prediction method for lithium-ion batteries based on machine learning. However, lithium-ion battery degradation mechanism analysis and SOH prediction from the perspective of electrochemistry under varying temperatures and different current rates are beyond the discussions of this work and require further investigations.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHODS DETAILS
  - Experimental battery parameters
  - Equipment used in the experiment
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

ACKNOWLEDGMENTS

This work was supported in part by the National Key R&D Program of China (No. 2018YFB0104000), the National Natural Science Foundation of China (No. 61763021), and the EU-funded Marie Skłodowska-Curie Individual Fellowships (No. 845102).

AUTHOR CONTRIBUTIONS

X. Shu and Z. Chen conceived the idea. X. Shu and J. Shen carried out the experiments and performed data extraction. X. Shu, S. Shen, and Y. Zhang conducted the literature review and wrote the manuscript. Z. Chen and G. Li revised and polished the manuscript. Z. Chen and Y. Liu reviewed and edited the manuscript. All authors commented on the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 21, 2021
Revised: September 8, 2021
Accepted: October 9, 2021
Published: November 19, 2021

REFERENCES

Aykol, M., Gopal, C.B., Anapolsky, A., Herring, P.K., van Vlijmen, B., Berliner, M.D., Bazant, M.Z., Braatz, R.D., Chueh, W.C., and Storey, B.D. (2021). Perspective—combining physics and machine learning to predict battery lifetime. J. Electrochem. Soc. *168*, 030525. https://doi.org/10.1149/1945-7111/abec55.

Beelen, H., Bergveld, H.J., and Donkers, M.C.F. (2021). Joint estimation of battery parameters and state of charge using an extended Kalman filter: a single-parameter tuning approach. IEEE Trans. Control Syst. Technol. *29*, 1087–1101. https://doi.org/10.1109/TCST.2020.2992523.

Bi, Y., and Choe, S.-Y. (2020). An adaptive sigma-point Kalman filter with state equality constraints for online state-of-charge estimation of a Li (NiMnCo) O_2/Carbon battery using a reduced-order electrochemical model. Appl. Energy *258*, 113925.

Bian, X., Liu, L., Yan, J., Zou, Z., and Zhao, R. (2020). An open circuit voltage-based model for state-of-health estimation of lithium-ion batteries: model development and validation. J. Power Sources *448*, 227401. https://doi.org/10.1016/j.jpowsour.2019.227401.

Bian, X., Wei, Z., He, J., Yan, F., and Liu, L. (2021a). A novel model-based voltage construction method for robust state-of-health estimation of lithium-ion batteries. IEEE Trans. Ind. Electron. *68*, 12173–12184. https://doi.org/10.1109/TIE.2020.3044779.

Bian, X., Wei, Z., He, J., Yan, F., and Liu, L. (2021b). A two-step parameter optimization method for low-order model-based state-of-charge estimation. IEEE Trans. Transp. Electrif. *7*, 399–409. https://doi.org/10.1109/TTE.2020.3032737.

Cai, L., Meng, J., Stroe, D.-I., Peng, J., Luo, G., and Teodorescu, R. (2020). Multiobjective optimization of data-driven model for lithium-ion battery SOH estimation with short-term feature. IEEE Trans. Power Electron. *35*, 11855–11864.

Chang, Y., Fang, H., and Zhang, Y. (2017). A new hybrid method for the prediction of the remaining useful life of a lithium-ion battery. Appl. Energy *206*, 1564–1578. https://doi.org/10.1016/j.apenergy.2017.09.106.

Chaoui, H., and Ibe-Ekeocha, C.C. (2017). State of charge and state of health estimation for lithium batteries using recurrent neural networks. IEEE Trans. Veh. Technol. *66*, 8773–8783. https://doi.org/10.1109/TVT.2017.2715333.

Che, Y., Deng, Z., Lin, X., Hu, L., and Hu, X. (2021). Predictive battery health management with transfer learning and online model correction. IEEE Trans. Veh. Technol. *70*, 1269–1277. https://doi.org/10.1109/TVT.2021.3055811.

Chehade, A.A., and Hussein, A.A. (2020). A collaborative Gaussian process regression model for transfer learning of capacity trends between Li-ion battery cells. IEEE Trans. Veh. Technol. *69*, 9542–9552. https://doi.org/10.1109/TVT.2020.3000970.

Chen, J., Ouyang, Q., Xu, C., and Su, H. (2018a). Neural network-based state of charge observer design for lithium-ion batteries. IEEE Trans. Control Syst. Technol. *26*, 313–320. https://doi.org/10.1109/TCST.2017.2664726.

Chen, Z., Sun, M., Shu, X., Xiao, R., and Shen, J. (2018b). Online state of health estimation for lithium-ion batteries based on support vector machine. Appl. Sci. *8*, 925. https://doi.org/10.3390/app8060925.

Chen, Z., Xue, Q., Wu, Y., Shen, S., Zhang, Y., and Shen, J. (2020). Capacity prediction and validation of lithium-ion batteries based on long short-term memory recurrent neural network. IEEE Access *8*, 172783–172798. https://doi.org/10.1109/ACCESS.2020.3025766.

Chen, C., Xiong, R., Yang, R., Shen, W., and Sun, F. (2019a). State-of-charge estimation of lithium-ion battery using an improved neural network model and extended Kalman filter. J. Clean. Prod. *234*, 1153–1164.

Chen, J., Jing, H., Chang, Y., and Liu, Q. (2019b). Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process. Reliab. Eng. Syst. Saf. *185*, 372–382. https://doi.org/10.1016/j.ress.2019.01.006.

Chen, Z., Shu, X., Xiao, R., Yan, W., Liu, Y., and Shen, J. (2019c). Optimal charging strategy design for lithium-ion batteries considering minimization of temperature rise and energy loss. Int. J. Energy Res. *43*, 4344–4358. https://doi.org/10.1002/er.4560.

Chen, Z., Xue, Q., Xiao, R., Liu, Y., and Shen, J. (2019d). State of health estimation for lithium-ion batteries based on fusion of autoregressive moving average model and elman neural network. IEEE Access *7*, 102662–102678.

Chen, L., Wang, H., Liu, B., Wang, Y., Ding, Y., and Pan, H. (2021a). Battery state-of-health estimation based on a metabolic extreme learning machine combining degradation state model and error compensation. Energy *215*, 119078.

Chen, Z., Zhao, H., Shu, X., Zhang, Y., Shen, J., and Liu, Y. (2021b). Synthetic state of charge estimation for lithium-ion batteries based on long short-term memory network modeling and adaptive H-Infinity filter. Energy *228*, 120630. https://doi.org/10.1016/j.energy.2021.120630.

Cheng, Y., Song, D., Wang, Z., Lu, C., and Zerhouni, N. (2020). An ensemble prognostic method for lithium-ion battery capacity estimation based on time-varying weight allocation. Appl. Energy *266*, 114817. https://doi.org/10.1016/j.apenergy.2020.114817.

Dai, H., Zhao, G., Lin, M., Wu, J., and Zheng, G. (2019). A novel estimation method for the state of health of lithium-ion battery using prior knowledge-based neural network and Markov chain. IEEE Trans. Ind. Electron. *66*, 7706–7716. https://doi.org/10.1109/TIE.2018.2880703.

Deng, Y., Ying, H., Jiaqiang, E., Zhu, H., Wei, K., Chen, J., Zhang, F., and Liao, G. (2019). Feature parameter extraction and intelligent estimation of the state-of-health of lithium-ion batteries. Energy *176*, 91–102.

Deng, Z., Hu, X., Lin, X., Xu, L., Che, Y., and Hu, L. (2021). General discharge voltage information enabled health evaluation for lithium-ion batteries. IEEE/ASME Trans. Mechatron. *26*, 1295–1306. https://doi.org/10.1109/TMECH.2020.3040010.

Domingos, P., and Hulten, G. (2000). Mining high-speed data streams. Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 71-80.

Eddahech, A., Briat, O., Bertrand, N., Delétage, J.-Y., and Vinassa, J.-M. (2012). Behavior and state-of-health monitoring of Li-ion batteries using impedance spectroscopy and recurrent neural networks. Int. J. Electr. Power Energy Syst. *42*, 487–494. https://doi.org/10.1016/j.ijepes.2012.04.050.

Fan, Y., Xiao, F., Li, C., Yang, G., and Tang, X. (2020). A novel deep learning framework for state of health estimation of lithium-ion battery. J. Energy Storage *32*, 101741. https://doi.org/10.1016/j.est.2020.101741.

Farmann, A., Waag, W., Marongiu, A., and Sauer, D.U. (2015). Critical review of on-board capacity estimation techniques for lithium-ion batteries in electric and hybrid electric vehicles. J. Power Sources *281*, 114–130. https://doi.org/10.1016/j.jpowsour.2015.01.129.

Fei, Z., Yang, F., Tsui, K.-L., Li, L., and Zhang, Z. (2021). Early prediction of battery lifetime via a machine learning based framework. Energy *225*, 120205. https://doi.org/10.1016/j.energy.2021.120205.

Feng, X., Weng, C., He, X., Han, X., Lu, L., Ren, D., and Ouyang, M. (2019). Online state-of-health estimation for Li-ion battery using partial charging segment based on support vector machine. IEEE Trans. Veh. Technol. *68*, 8583–8592.

Feng, F., Teng, S., Liu, K., Xie, J., Xie, Y., Liu, B., and Li, K. (2020a). Co-estimation of lithium-ion battery state of charge and state of temperature based on a hybrid electrochemical-thermal-neural-network model. J. Power Sources *455*, 227935.

Feng, Y., Xue, C., Han, Q., Han, F., and Du, J. (2020b). Robust estimation for state-of-charge and state-of-health of lithium-ion batteries using integral-type terminal sliding-mode observers. IEEE Trans. Ind. Electron. *67*, 4013–4023. https://doi.org/10.1109/TIE.2019.2916389.

Gou, B., Xu, Y., and Feng, X. (2020). State-of-Health estimation and remaining-useful-life prediction for lithium-ion battery using a hybrid data-driven method. IEEE Trans. Veh. Technol. *69*, 10854–10867. https://doi.org/10.1109/TVT.2020.3014932.

Gou, B., Xu, Y., and Feng, X. (2021). An ensemble learning-based data-driven method for online state-of-health estimation of lithium-ion batteries. IEEE Trans. Transp. Electrif. *7*, 422–436. https://doi.org/10.1109/TTE.2020.3029295.

Goud, J.S., K, R., and Singh, B. (2021). An online method of estimating state of health of a Li-ion battery. IEEE Trans. Energy Convers. *36*, 111–119. https://doi.org/10.1109/TEC.2020.3008937.

Guo, P., Cheng, Z., and Yang, L. (2019). A data-driven remaining capacity estimation approach for lithium-ion batteries based on charging health feature extraction. J. Power Sources *412*, 442–450.

Guo, N., Zhang, X., Zou, Y., Guo, L., and Du, G. (2021a). Real-time predictive energy management of plug-in hybrid electric vehicles for coordination of fuel economy and battery degradation. Energy *214*, 119070. https://doi.org/10.1016/j.energy.2020.119070.

Guo, Y., Huang, K., and Hu, X. (2021b). A state-of-health estimation method of lithium-ion batteries based on multi-feature extracted from constant current charging curve. J. Energy Storage *36*, 102372. https://doi.org/10.1016/j.est.2021.102372.

Guo, Y., Yang, Z., Liu, K., Zhang, Y., and Feng, W. (2021c). A compact and optimized neural network approach for battery state-of-charge estimation of energy storage system. Energy *219*, 119529. https://doi.org/10.1016/j.energy.2020.119529.

Han, M., and Liu, B. (2015). Ensemble of extreme learning machine for remote sensing image classification. Neurocomputing *149*, 65–70. https://doi.org/10.1016/j.neucom.2013.09.070.

Hannan, M.A., How, D.N.T., Mansor, M.B., Lipu, M.S.H., Ker, P.J., and Muttaqi, K.M. (2021). State-of-charge estimation of Li-ion battery using gated recurrent unit with one-cycle learning rate policy. IEEE Trans. Ind. Appl. *57*, 2964–2971. https://doi.org/10.1109/TIA.2021.3065194.

He, J., Wei, Z., Bian, X., and Yan, F. (2020a). State-of-health estimation of lithium-ion batteries using incremental capacity analysis based on voltage–capacity model. IEEE Trans. Transp. Electrif. *6*, 417–426. https://doi.org/10.1109/TTE.2020.2994543.

He, Z., Yang, Z., Cui, X., and Li, E. (2020b). A method of state-of-charge estimation for EV power lithium-ion battery using a novel adaptive extended Kalman filter. IEEE Trans. Veh. Technol. *69*, 14618–14630. https://doi.org/10.1109/TVT.2020.3032201.

Hong, J., Wang, Z., Chen, W., Wang, L., Lin, P., and Qu, C. (2021). Online accurate state of health estimation for battery systems on real-world electric vehicles with variable driving conditions considered. J. Clean. Prod. *294*, 125814. https://doi.org/10.1016/j.jclepro.2021.125814.

Hong, S., and Zeng, Y. (2021). A health assessment framework of lithium-ion batteries for cyber defense. Appl. Soft Comput. *101*, 107067. https://doi.org/10.1016/j.asoc.2020.107067.

Hosen, M.S., Jaguemont, J., Van Mierlo, J., and Berecibar, M. (2021). Battery lifetime prediction and performance assessment of different modeling approaches. Iscience *24*, 102060.

Hossain, M., Saha, S., Arif, M.T., Oo, A.M.T., Mendis, N., and Haque, M.E. (2020). A parameter extraction method for the Li-ion batteries with wide-range temperature compensation. IEEE Trans. Industry Appl. *56*, 5625–5636. https://doi.org/10.1109/TIA.2020.3011385.

Houlian, W., and Gongbo, Z. (2018). State of charge prediction of supercapacitors via combination of Kalman filtering and backpropagation neural network. IET Electr. Power Appl. *12*, 588–594.

Hu, C., Jain, G., Schmidt, C., Strief, C., and Sullivan, M. (2015). Online estimation of lithium-ion battery capacity using sparse Bayesian learning. J. Power Sources *289*, 105–113.

Hu, X., Che, Y., Lin, X., and Deng, Z. (2020). Health prognosis for electric vehicle battery packs: a data-driven approach. IEEE/ASME Trans. Mechatron. *25*, 2622–2632.

Hu, X., Che, Y., Lin, X., and Onori, S. (2021). Battery health prediction using fusion-based feature selection and machine learning. IEEE Trans. Transp. Electrif. *7*, 382–398. https://doi.org/10.1109/TTE.2020.3017090.

Hu, X., Yuan, H., Zou, C., Li, Z., and Zhang, L. (2018). Co-estimation of state of charge and state of health for lithium-ion batteries based on fractional-order calculus. IEEE Trans. Veh. Technol. *67*, 10319–10329.

Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. (2006). Extreme learning machine: theory and applications. Neurocomputing *70*, 489–501.

Islam, S.M.R., and Park, S. (2020). Precise online electrochemical impedance spectroscopy strategies for Li-ion batteries. IEEE Trans. Ind. Appl. *56*, 1661–1669. https://doi.org/10.1109/TIA.2019.2958555.

Jia, S., Ma, B., Guo, W., and Li, Z.S. (2021). A sample entropy based prognostics method for lithium-ion batteries using relevance vector machine. J. Manuf. Syst. https://doi.org/10.1016/j.jmsy.2021.03.019.

Jiang, L., Huang, Y., Li, Y., Yu, J., Qiao, X., Huang, C., and Cao, Y. (2021). Optimization of variable-current charging strategy based on SOC segmentation for Li-ion battery. IEEE Trans. Intell. Transp. Syst. *22*, 622–629. https://doi.org/10.1109/TITS.2020.3006092.

Jiang, Y., Jiang, J., Zhang, C., Zhang, W., Gao, Y., and Li, N. (2018). State of health estimation of second-life $LiFePO_4$ batteries for energy storage applications. J. Clean. Prod. *205*, 754–762.

Johnson, V.H. (2002). Battery performance models in ADVISOR. J. Power Sources *110*, 321–329. https://doi.org/10.1016/S0378-7753(02)00194-5.

Kavzoglu, T., and Colkesen, I. (2013). An assessment of the effectiveness of a rotation forest ensemble for land-use and land-cover mapping. Int. J. Remote Sens. *34*, 4224–4241. https://doi.org/10.1080/01431161.2013.774099.

Khodadadi Sadabadi, K., Jin, X., and Rizzoni, G. (2021). Prediction of remaining useful life for a composite electrode lithium ion battery cell using an electrochemical model to estimate the state of health. J. Power Sources *481*, 228861. https://doi.org/10.1016/j.jpowsour.2020.228861.

Kim, Y., and Sohn, S.Y. (2012). Stock fraud detection using peer group analysis. Expert Syst. Appl. *39*, 8986–8992. https://doi.org/10.1016/j.eswa.2012.02.025.

Koseoglou, M., Tsioumas, E., Papagiannis, D., Jabbour, N., and Mademlis, C. (2021). A novel on-board electrochemical impedance spectroscopy system for real-time battery impedance estimation. IEEE Trans. Power Electron. *36*, 10776–10787. https://doi.org/10.1109/TPEL.2021.3063506.

Lee, C., Jo, S., Kwon, D., and Pecht, M.G. (2021). Capacity-fading behavior analysis for early detection of unhealthy Li-ion batteries. IEEE Trans. Ind. Electron. *68*, 2659–2666. https://doi.org/10.1109/TIE.2020.2972468.

Lee, J., Wang, W., Harrou, F., and Sun, Y. (2020). Reliable solar irradiance prediction using ensemble learning-based models: a comparative study. Energy Convers. Manag. *208*, 112582. https://doi.org/10.1016/j.enconman.2020.112582.

Li, P., Zhang, Z., Xiong, Q., Ding, B., Hou, J., Luo, D., Rong, Y., and Li, S. (2020a). State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network. J. Power Sources *459*, 228069. https://doi.org/10.1016/j.jpowsour.2020.228069.

Li, W., Fan, Y., Ringbeck, F., Jöst, D., Han, X., Ouyang, M., and Sauer, D.U. (2020b). Electrochemical model-based state estimation for lithium-ion batteries with adaptive unscented Kalman filter. J. Power Sources *476*, 228534. https://doi.org/10.1016/j.jpowsour.2020.228534.

Li, X., Yuan, C., Li, X., and Wang, Z. (2020c). State of health estimation for Li-ion battery using incremental capacity analysis and Gaussian process regression. Energy *190*, 116467.

Li, D., Zhang, Z., Liu, P., Wang, Z., and Zhang, L. (2021a). Battery fault diagnosis for electric vehicles based on voltage abnormality by combining the long short-term memory neural network and the equivalent circuit model. IEEE Trans. Power Electron. *36*, 1303–1315. https://doi.org/10.1109/TPEL.2020.3008194.

Li, W., Sengupta, N., Dechent, P., Howey, D., Annaswamy, A., and Sauer, D.U. (2021b). Online capacity estimation of lithium-ion batteries with deep long short-term memory networks. J. Power Sources *482*, 228863. https://doi.org/10.1016/j.jpowsour.2020.228863.

Li, X., Huang, Z., Tian, J., and Tian, Y. (2021c). State-of-charge estimation tolerant of battery aging based on a physics-based model and an adaptive cubature Kalman filter. Energy *220*, 119767. https://doi.org/10.1016/j.energy.2021.119767.

Li, Y., Abdel-Monem, M., Gopalakrishnan, R., Berecibar, M., Nanini-Maury, E., Omar, N., van den Bossche, P., and Van Mierlo, J. (2018a). A quick on-line state of health estimation method for Li-ion battery with incremental capacity curves processed by Gaussian filter. J. Power Sources *373*, 40–53.

Li, Y., Zou, C., Berecibar, M., Nanini-Maury, E., Chan, J.C.-W., Van den Bossche, P., Van Mierlo, J., and Omar, N. (2018b). Random forest regression for online capacity estimation of lithium-ion batteries. Appl. Energy *232*, 197–210.

Li, W., Jiao, Z., Du, L., Fan, W., and Zhu, Y. (2019a). An indirect RUL prognosis for lithium-ion battery under vibration stress using Elman neural network. Int. J. Hydrogen Energy *44*,

12270–12276. https://doi.org/10.1016/j.ijhydene.2019.03.101.

Li, X., Wang, Z., and Zhang, L. (2019b). Co-estimation of capacity and state-of-charge for lithium-ion batteries in electric vehicles. Energy *174*, 33–44.

Li, X., Zhang, L., Wang, Z., and Dong, P. (2019c). Remaining useful life prediction for lithium-ion batteries based on a hybrid model combining the long short-term memory and Elman neural networks. J. Energy Storage *21*, 510–518. https://doi.org/10.1016/j.est.2018.12.011.

Li, Y., Liu, K., Foley, A.M., Zülke, A., Berecibar, M., Nanini-Maury, E., Van Mierlo, J., and Hoster, H.E. (2019d). Data-driven health estimation and lifetime prediction of lithium-ion batteries: a review. Renew. Sustain. Energy Rev. *113*, 109254.

Li, Y., Zhong, S., Zhong, Q., and Shi, K. (2019e). Lithium-ion battery state of health monitoring based on ensemble learning. IEEE Access *7*, 8754–8762. https://doi.org/10.1109/ACCESS.2019.2891063.

Liang, N., Huang, G., Saratchandran, P., and Sundararajan, N. (2006). A fast and accurate online sequential learning algorithm for feedforward networks. IEEE Trans. Neural Netw. *17*, 1411–1423. https://doi.org/10.1109/TNN.2006.880583.

Lin, M., Zeng, X., and Wu, J. (2021). State of health estimation of lithium-ion battery based on an adaptive tunable hybrid radial basis function network. J. Power Sources *504*, 230063. https://doi.org/10.1016/j.jpowsour.2021.230063.

Lipu, M.S.H., Hannan, M.A., Hussain, A., Hoque, M.M., Ker, P.J., Saad, M.H.M., and Ayob, A. (2018). A review of state of health and remaining useful life estimation methods for lithium-ion battery in electric vehicles: challenges and recommendations. J. Clean. Prod. *205*, 115–133. https://doi.org/10.1016/j.jclepro.2018.09.065.

Lipu, M.S.H., Hannan, M.A., Hussain, A., Saad, M.H., Ayob, A., and Uddin, M.N. (2019). Extreme learning machine model for state-of-charge estimation of lithium-ion battery using gravitational search algorithm. IEEE Trans. Ind. Appl. *55*, 4225–4234. https://doi.org/10.1109/TIA.2019.2902532.

Liu, K., Li, Y., Hu, X., Lucu, M., and Widanage, W.D. (2019). Gaussian process regression with automatic relevance determination kernel for calendar aging prediction of lithium-ion batteries. IEEE Trans. Ind. Inform. *16*, 3767–3777.

Liu, B., Tang, X., and Gao, F. (2020a). Joint estimation of battery state-of-charge and state-of-health based on a simplified pseudo-two-dimensional model. Electrochim. Acta *344*, 136098. https://doi.org/10.1016/j.electacta.2020.136098.

Liu, W., Xu, Y., and Feng, X. (2020b). A hierarchical and flexible data-driven method for online state-of-health estimation of Li-ion battery. IEEE Trans. Veh. Technol. *69*, 14739–14748. https://doi.org/10.1109/TVT.2020.3037088.

Liu, K., Hu, X., Zhou, H., Tong, L., Widanalage, D., and Macro, J. (2021a). Feature analyses and modelling of lithium-ion batteries manufacturing based on random forest classification. IEEE/ASME Trans. Mechatron. 1. https://doi.org/10.1109/TMECH.2020.3049046.

Liu, K., Shang, Y., Ouyang, Q., and Widanage, W.D. (2021b). A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery. IEEE Trans. Ind. Electron. *68*, 3170–3180. https://doi.org/10.1109/TIE.2020.2973876.

Liu, Y., Ma, R., Pang, S., Xu, L., Zhao, D., Wei, J., Huangfu, Y., and Gao, F. (2021c). A nonlinear observer SOC estimation method based on electrochemical model for lithium-ion battery. IEEE Trans. Ind. Appl. *57*, 1094–1104. https://doi.org/10.1109/TIA.2020.3040140.

Liu, Y., Shu, X., Yu, H., Shen, J., Zhang, Y., Liu, Y., and Chen, Z. (2021d). State of charge prediction framework for lithium-ion batteries incorporating long short-term memory network and transfer learning. J. Energy Storage *37*, 102494. https://doi.org/10.1016/j.est.2021.102494.

Liu, W., and Xu, Y. (2020). Data-driven online health estimation of Li-ion batteries using a novel energy-based health indicator. IEEE Trans. Energy Convers. *35*, 1715–1718. https://doi.org/10.1109/TEC.2020.2995112.

Losing, V., Hammer, B., and Wersing, H. (2018). Incremental on-line learning: a review and comparison of state of the art algorithms. Neurocomputing *275*, 1261–1274. https://doi.org/10.1016/j.neucom.2017.06.084.

Lyu, Z., Gao, R., and Chen, L. (2021). Li-ion battery state of health estimation and remaining useful life prediction through a model-data-fusion method. IEEE Trans. Power Electron. *36*, 6228–6240. https://doi.org/10.1109/TPEL.2020.3033297.

Marelli, S., and Corno, M. (2021). Model-based estimation of lithium concentrations and temperature in batteries using soft-constrained dual unscented Kalman filtering. IEEE Trans. Control Syst. Technol. *29*, 926–933. https://doi.org/10.1109/TCST.2020.2974176.

Mawonou, K.S.R., Eddahech, A., Dumur, D., Beauvois, D., and Godoy, E. (2021). State-of-health estimators coupled to a random forest approach for lithium-ion battery aging factor ranking. J. Power Sources *484*, 229154. https://doi.org/10.1016/j.jpowsour.2020.229154.

Meng, H., and Li, Y.-F. (2019). A review on prognostics and health management (PHM) methods of lithium-ion batteries. Renew. Sustain. Energy Rev. *116*, 109405.

Meng, J., Cai, L., Stroe, D.-I., Luo, G., Sui, X., and Teodorescu, R. (2019a). Lithium-ion battery state-of-health estimation in electric vehicle using optimized partial charging voltage profiles. Energy *185*, 1054–1062.

Meng, J., Stroe, D., Ricco, M., Luo, G., and Teodorescu, R. (2019b). A simplified model-based state-of-charge estimation approach for lithium-ion battery with dynamic linear model. IEEE Trans. Ind. Electron. *66*, 7717–7727. https://doi.org/10.1109/TIE.2018.2880668.

Meng, J., Cai, L., Stroe, D.I., Huang, X., Peng, J., Liu, T., and Teodorescu, R. (2021). An automatic weak learner formulation for lithium-ion battery state of health estimation. IEEE Trans. Ind. Electron. 1. https://doi.org/10.1109/TIE.2021.3065594.

Merla, Y., Wu, B., Yufit, V., Brandon, N.P., Martinez-Botas, R.F., and Offer, G.J. (2016). Novel application of differential thermal voltammetry as an in-depth state-of-health diagnosis method for lithium-ion batteries. J. Power Sources *307*, 308–319.

Naseri, F., Farjah, E., Ghanbari, T., Kazemi, Z., Schaltz, E., and Schanen, J. (2020). Online parameter estimation for supercapacitor state-of-energy and state-of-health determination in vehicular applications. IEEE Trans. Ind. Electron. *67*, 7963–7972. https://doi.org/10.1109/TIE.2019.2941151.

Ng, M.-F., Zhao, J., Yan, Q., Conduit, G.J., and Seh, Z.W. (2020). Predicting the state of charge and health of batteries using data-driven machine learning. Nat. Mach. Intell. *2*, 161–170. https://doi.org/10.1038/s42256-020-0156-7.

Ouyang, Q., Chen, J., and Zheng, J. (2020). State-of-charge observer design for batteries with online model parameter identification: a robust approach. IEEE Trans. Power Electron. *35*, 5820–5831. https://doi.org/10.1109/TPEL.2019.2948253.

Ouyang, T., Xu, P., Lu, J., Hu, X., Liu, B., and Chen, N. (2021). Co-estimation of state-of-charge and state-of-health for power batteries based on multi-thread dynamic optimization method. IEEE Trans. Ind. Electron. 1. https://doi.org/10.1109/TIE.2021.3062266.

Pan, H., Lü, Z., Wang, H., Wei, H., and Chen, L. (2018). Novel battery state-of-health online estimation method using multiple health indicators and an extreme learning machine. Energy *160*, 466–477.

Pan, W., Chen, Q., Zhu, M., Tang, J., and Wang, J. (2020). A data-driven fuzzy information granulation approach for battery state of health forecasting. J. Power Sources *475*, 228716. https://doi.org/10.1016/j.jpowsour.2020.228716.

Peng, J., Luo, J., He, H., and Lu, B. (2019). An improved state of charge estimation method based on cubature Kalman filter for lithium-ion batteries. Appl. Energy *253*, 113520. https://doi.org/10.1016/j.apenergy.2019.113520.

Pernici, F., and Bimbo, A.D. (2014). Object tracking by Oversampling local features. IEEE Trans. Pattern Anal. Machine Intelligence *36*, 2538–2551. https://doi.org/10.1109/TPAMI.2013.250.

Ren, L., Dong, J., Wang, X., Meng, Z., Zhao, L., and Deen, M.J. (2021). A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life. IEEE Trans. Ind. Inform. *17*, 3478–3487. https://doi.org/10.1109/TII.2020.3008223.

Richardson, R.R., Birkl, C.R., Osborne, M.A., and Howey, D.A. (2018). Gaussian process regression for in situ capacity estimation of lithium-ion batteries. IEEE Trans. Ind. Inform. *15*, 127–138.

Richardson, R.R., Osborne, M.A., and Howey, D.A. (2017). Gaussian process regression for forecasting battery state of health. J. Power Sources *357*, 209–219.

Roman, D., Saxena, S., Robu, V., Pecht, M., and Flynn, D. (2021). Machine learning pipeline for battery state-of-health estimation. Nat. Mach. Intell. *3*, 447–456. https://doi.org/10.1038/s42256-021-00312-3.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1986). Learning representations by back-propagating errors. Nature *323*, 533–536.

Sarrafan, K., Muttaqi, K.M., and Sutanto, D. (2020). Real-time estimation of model parameters and state-of-charge of Li-ion batteries in electric vehicles using a new mixed estimation model. IEEE Trans. Ind. Appl. *56*, 5417–5428. https://doi.org/10.1109/TIA.2020.3002977.

Sbarufatti, C., Corbetta, M., Giglio, M., and Cadini, F. (2017). Adaptive prognosis of lithium-ion batteries based on the combination of particle filters and radial basis function neural networks. J. Power Sources *344*, 128–140. https://doi.org/10.1016/j.jpowsour.2017.01.105.

Schaltz, E., Stroe, D.I., Nrregaard, K., Ingvardsen, L.S., and Christensen, A. (2021). Incremental capacity analysis applied on electric vehicles for battery state-of-health estimation. IEEE Trans. Ind. Appl. *57*, 1810–1817. https://doi.org/10.1109/TIA.2021.3052454.

She, C., Wang, Z., Sun, F., Liu, P., and Zhang, L. (2019). Battery aging assessment for real-world electric buses based on incremental capacity analysis and radial basis function neural network. IEEE Trans. Ind. Inform. *16*, 3345–3354.

Shen, F., and Hasegawa, O. (2008). A fast nearest neighbor classifier based on self-organizing incremental neural network. Neural Netw. *21*, 1537–1547. https://doi.org/10.1016/j.neunet.2008.07.001.

Shen, S., Sadoughi, M., Li, M., Wang, Z., and Hu, C. (2020). Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries. Appl. Energy *260*, 114296. https://doi.org/10.1016/j.apenergy.2019.114296.

Shibagaki, T., Merla, Y., and Offer, G.J. (2018). Tracking degradation in lithium iron phosphate batteries using differential thermal voltammetry. J. Power Sources *374*, 188–195. https://doi.org/10.1016/j.jpowsour.2017.11.011.

Shrivastava, P., Soon, T.K., Idris, M.Y.I.B., Mekhilef, S., and Adnan, S.B.R.S. (2021). Combined state of charge and state of energy estimation of lithium-ion battery using dual forgetting factor-based adaptive extended Kalman filter for electric vehicle applications. IEEE Trans. Veh. Technol. *70*, 1200–1215. https://doi.org/10.1109/TVT.2021.3051655.

Shu, X., Li, G., Shen, J., Lei, Z., Chen, Z., and Liu, Y. (2020a). An adaptive multi-state estimation algorithm for lithium-ion batteries incorporating temperature compensation. Energy *207*, 118262. https://doi.org/10.1016/j.energy.2020.118262.

Shu, X., Li, G., Shen, J., Lei, Z., Chen, Z., and Liu, Y. (2020b). A uniform estimation framework for state of health of lithium-ion batteries considering feature extraction and parameters optimization. Energy *204*, 117957. https://doi.org/10.1016/j.energy.2020.117957.

Shu, X., Li, G., Shen, J., Yan, W., Chen, Z., and Liu, Y. (2020c). An adaptive fusion estimation algorithm for state of charge of lithium-ion batteries considering wide operating temperature and degradation. J. Power Sources *462*, 228132. https://doi.org/10.1016/j.jpowsour.2020.228132.

Shu, X., Li, G., Zhang, Y., Shen, J., Chen, Z., and Liu, Y. (2020d). Online diagnosis of state of health for lithium-ion batteries based on short-term charging profiles. J. Power Sources *471*, 228478. https://doi.org/10.1016/j.jpowsour.2020.228478.

Shu, X., Li, G., Zhang, Y., Shen, S., Chen, Z., and Liu, Y. (2021a). Stage of charge estimation of lithium-ion battery packs based on improved cubature Kalman filter with long short-term memory model. IEEE Trans. Transp. Electrif. *7*, 1271–1284. https://doi.org/10.1109/TTE.2020.3041757.

Shu, X., Shen, J., Li, G., Zhang, Y., Chen, Z., and Liu, Y. (2021b). A flexible state of health prediction scheme for lithium-ion battery packs with long short-term memory network and transfer learning. IEEE Trans. Transp. Electrif. 1. https://doi.org/10.1109/TTE.2021.3074638.

Singh, K.V., Bansal, H.O., and Singh, D. (2021). Fuzzy logic and Elman neural network tuned energy management strategies for a power-split HEVs. Energy *225*, 120152.

Song, Y., Peng, Y., and Liu, D. (2021). Model-based health diagnosis for lithium-ion battery pack in space applications. IEEE Trans. Ind. Electron. *68*, 12375–12384. https://doi.org/10.1109/TIE.2020.3045745.

Song, Z., Wu, X., Li, X., Sun, J., Hofmann, H.F., and Hou, J. (2019). Current profile optimization for combined state of charge and state of health estimation of lithium ion battery based on Cramer–Rao bound analysis. IEEE Trans. Power Electron. *34*, 7067–7078. https://doi.org/10.1109/TPEL.2018.2877294.

Sun, C., Lin, H., Cai, H., Gao, M., Zhu, C., and He, Z. (2021). Improved parameter identification and state-of-charge estimation for lithium-ion battery with fixed memory recursive least squares and sigma-point Kalman filter. Electrochim. Acta *387*, 138501. https://doi.org/10.1016/j.electacta.2021.138501.

Tagade, P., Hariharan, K.S., Ramachandran, S., Khandelwal, A., Naha, A., Kolake, S.M., and Han, S.H. (2020). Deep Gaussian process regression for lithium-ion battery health prognosis and degradation mode diagnosis. J. Power Sources *445*, 227281.

Tan, X., Zhan, D., Lyu, P., Rao, J., and Fan, Y. (2021). Online state-of-health estimation of lithium-ion battery based on dynamic parameter identification at multi timescale and support vector regression. J. Power Sources *484*, 229233. https://doi.org/10.1016/j.jpowsour.2020.229233.

Tan, Y., and Zhao, G. (2020). Transfer learning with long short-term memory network for state-of-health prediction of lithium-ion batteries. IEEE Trans. Ind. Electron. *67*, 8723–8731. https://doi.org/10.1109/TIE.2019.2946551.

Tang, X., Liu, K., Lu, J., Liu, B., Wang, X., and Gao, F. (2020). Battery incremental capacity curve extraction by a two-dimensional Luenberger–Gaussian-moving-average filter. Appl. Energy *280*, 115895. https://doi.org/10.1016/j.apenergy.2020.115895.

Tang, X., Zou, C., Yao, K., Chen, G., Liu, B., He, Z., and Gao, F. (2018). A fast estimation algorithm for lithium-ion battery state of health. J. Power Sources *396*, 453–458.

Tian, H., and Qin, P. (2021). State of health prediction for lithium-ion batteries with a novel online sequential extreme learning machine method. Int. J. Energy Res. *45*, 2383–2397. https://doi.org/10.1002/er.5934.

Tian, H., Qin, P., Li, K., and Zhao, Z. (2020a). A review of the state of health for lithium-ion batteries: research status and suggestions. J. Clean. Prod. *261*, 120813.

Tian, J., Xiong, R., and Shen, W. (2020b). State-of-health estimation based on differential temperature for lithium ion batteries. IEEE Trans. Power Electron. *35*, 10363–10373.

Tian, J., Xiong, R., and Shen, W. (2019). A review on state of health estimation for lithium ion batteries in photovoltaic systems. eTransportation *2*, 100028. https://doi.org/10.1016/j.etran.2019.100028.

Tian, J., Xiong, R., Shen, W., and Sun, F. (2021). Electrode ageing estimation and open circuit voltage reconstruction for lithium ion batteries. Energy Storage Mater. *37*, 283–295. https://doi.org/10.1016/j.ensm.2021.02.018.

Tian, J., Xiong, R., and Yu, Q. (2018). Fractional-order model-based incremental capacity analysis for degradation state recognition of lithium-ion batteries. IEEE Trans. Ind. Electron. *66*, 1576–1584.

Tipping, M.E. (2001). Sparse Bayesian learning and the relevance vector machine. J. Mach. Learn. Res. *1*, 211–244.

Ungurean, L., Cârstoiu, G., Micea, M.V., and Groza, V. (2017). Battery state of health estimation: a structured review of models, methods and commercial devices. Int. J. Energy Res. *41*, 151–181. https://doi.org/10.1002/er.3598.

Ungurean, L., Micea, M.V., and Cârstoiu, G. (2020). Online state of health prediction method for lithium-ion batteries, based on gated recurrent unit neural networks. Int. J. Energy Res. *44*, 6767–6777.

Vidal, C., Malysz, P., Kollmeyer, P., and Emadi, A. (2020). Machine learning applied to electrified vehicle battery state of charge and state of health estimation: state-of-the-art. IEEE Access *8*, 52796–52814. https://doi.org/10.1109/ACCESS.2020.2980961.

Wang, Z., Ma, J., and Zhang, L. (2017). State-of-health estimation for lithium-ion batteries based on the multi-island genetic algorithm and the Gaussian process regression. Ieee Access *5*, 21286–21295.

Wang, Z., Yuan, C., and Li, X. (2021). Lithium battery state-of-health estimation via differential thermal voltammetry with Gaussian process regression. IEEE Trans. Transp. Electrif. *7*, 16–25. https://doi.org/10.1109/TTE.2020.3028784.

Wei, M., Gu, H., Ye, M., Wang, Q., Xu, X., and Wu, C. (2021). Remaining useful life prediction of lithium-ion batteries based on Monte Carlo dropout and gated recurrent unit. Energy Rep. *7*, 2862–2871. https://doi.org/10.1016/j.egyr.2021.05.019.

Wu, L., Liu, K., and Pang, H. (2021). Evaluation and observability analysis of an improved reduced-order electrochemical model for lithium-ion battery. Electrochim. Acta *368*, 137604. https://doi.org/10.1016/j.electacta.2020.137604.

Xiao, D., Fang, G., Liu, S., Yuan, S., Ahmed, R., Habibi, S., and Emadi, A. (2020). Reduced-coupling coestimation of SOC and SOH for lithium-ion batteries based on convex optimization. IEEE Trans. Power Electron. *35*, 12332–12346.

Xiong, R., Li, L., and Tian, J. (2018a). Towards a smarter battery management system: a critical review on battery state of health monitoring methods. J. Power Sources *405*, 18–29.

Xiong, R., Zhang, Y., Wang, J., He, H., Peng, S., and Pecht, M. (2018b). Lithium-ion battery health prognosis based on a real battery management system used in electric vehicles. IEEE Trans. Veh. Technol. *68*, 4110–4121.

Xiong, R., Tian, J., Shen, W., and Sun, F. (2019). A novel fractional order model for state of charge estimation in lithium ion batteries. IEEE Trans. Veh. Technol. *68*, 4130–4139. https://doi.org/10.1109/TVT.2018.2880085.

Xu, T., Peng, Z., and Wu, L. (2021a). A novel data-driven method for predicting the circulating capacity of lithium-ion battery under random variable current. Energy *218*, 119530. https://doi.org/10.1016/j.energy.2020.119530.

Xu, Z., Wang, J., Lund, P.D., and Zhang, Y. (2021b). Estimation and prediction of state of health of electric vehicle batteries using discrete incremental capacity analysis based on real driving data. Energy *225*, 120160. https://doi.org/10.1016/j.energy.2021.120160.

Xu, X., Yu, C., Tang, S., Sun, X., Si, X., and Wu, L. (2019). State-of-health estimation for lithium-ion batteries based on Wiener process with modeling the relaxation effect. IEEE Access *7*, 105186–105201. https://doi.org/10.1109/ACCESS.2019.2923095.

Yan, W., Zhang, B., Zhao, G., Tang, S., Niu, G., and Wang, X. (2019). A battery management system with a Lebesgue-sampling-based extended Kalman filter. IEEE Trans. Ind. Electron. *66*, 3227–3236. https://doi.org/10.1109/TIE.2018.2842782.

Yang, D., Wang, Y., Pan, R., Chen, R., and Chen, Z. (2018a). State-of-health estimation for the lithium-ion battery based on support vector regression. Appl. Energy *227*, 273–283.

Yang, D., Zhang, X., Pan, R., Wang, Y., and Chen, Z. (2018b). A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve. J. Power Sources *384*, 387–395. https://doi.org/10.1016/j.jpowsour.2018.03.015.

Yang, F., Li, W., Li, C., and Miao, Q. (2019). State-of-charge estimation of lithium-ion batteries based on gated recurrent neural network. Energy *175*, 66–75.

Yang, R., Xiong, R., Shen, W., and Lin, X. (2021). Extreme learning machine-based thermal model for lithium-ion batteries of electric vehicles under external short circuit. Engineering *7*, 395–405. https://doi.org/10.1016/j.eng.2020.08.015.

Yann Liaw, B., Nagasubramanian, G., Jungst, R.G., and Doughty, D.H. (2004). Modeling of lithium ion cells—a simple equivalent-circuit model approach. Solid State Ionics *175*, 835–839. https://doi.org/10.1016/j.ssi.2004.09.049.

Yao, Z., Lu, S., Li, Y., and Yi, X. (2019). Cycle life prediction of lithium ion battery based on DE-BP neural network. In 2019 International conference on sensing, diagnostics, prognostics, and control (SDPC) held in Beijing, China, 15–17 Aug. 2019 (IEEE), pp. 137–141. https://doi.org/10.1109/SDPC.2019.00033.

Yu, J. (2018). State of health prediction of lithium-ion batteries: multiscale logic regression and Gaussian process regression ensemble. Reliab. Eng. Syst. Saf. *174*, 82–95. https://doi.org/10.1016/j.ress.2018.02.022.

Yu, Q., Xiong, R., Yang, R., and Pecht, M.G. (2019). Online capacity estimation for lithium-ion batteries through joint estimation method. Appl. Energy *255*, 113817. https://doi.org/10.1016/j.apenergy.2019.113817.

Zhang, D., Dey, S., Couto, L.D., and Moura, S.J. (2020a). Battery adaptive observer for a single-particle model with intercalation-induced stress. IEEE Trans. Control Syst. Technol. *28*, 1363–1377. https://doi.org/10.1109/TCST.2019.2910797.

Zhang, Y., Tang, Q., Zhang, Y., Wang, J., Stimming, U., and Lee, A.A. (2020b). Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. Nat. Commun. *11*, 1–6.

Zhang, S., Zhai, B., Guo, X., Wang, K., Peng, N., and Zhang, X. (2019a). Synchronous estimation of state of health and remaining useful lifetime for lithium-ion battery using the incremental capacity and artificial neural networks. J. Energy Storage *26*, 100951. https://doi.org/10.1016/j.est.2019.100951.

Zhang, Y., Xiong, R., He, H., Qu, X., and Pecht, M. (2019b). Aging characteristics-based health diagnosis and remaining useful life prognostics for lithium-ion batteries. eTransportation *1*, 100004. https://doi.org/10.1016/j.etran.2019.100004.

Zhang, X., Guo, L., Guo, N., Zou, Y., and Du, G. (2021a). Bi-level energy management of plug-in hybrid electric vehicles for fuel economy and battery lifetime with intelligent state-of-charge reference. J. Power Sources *481*, 228798. https://doi.org/10.1016/j.jpowsour.2020.228798.

Zhang, Y., Peng, Z., Guan, Y., and Wu, L. (2021b). Prognostics of battery cycle life in the early-cycle stage based on hybrid model. Energy *221*, 119901. https://doi.org/10.1016/j.energy.2021.119901.

Zhang, Y., Xiong, R., He, H., and Pecht, M.G. (2018). Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. IEEE Trans. Veh. Technol. *67*, 5695–5705.

Zhao, R., Kollmeyer, P.J., Lorenz, R.D., and Jahns, T.M. (2019). A compact methodology via a recurrent neural network for accurate equivalent circuit type modeling of lithium-ion batteries. IEEE Trans. Industry Appl. *55*, 1922–1931. https://doi.org/10.1109/TIA.2018.2874588.

Zhao, X., Xuan, D., Zhao, K., and Li, Z. (2020). Elman neural network using ant colony optimization algorithm for estimating of state of charge of lithium-ion battery. J. Energy Storage *32*, 101789. https://doi.org/10.1016/j.est.2020.101789.

Zhou, D., Zheng, W., Chen, S., Fu, P., Zhu, H., Song, B., Qu, X., and Wang, T. (2021). Research on state of health prediction model for lithium batteries based on actual diverse data. Energy *230*, 120851. https://doi.org/10.1016/j.energy.2021.120851.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
Battery	Lishen	LR2170SA 4.0Ah
Software and algorithms		
Matlab	MathWorks	https://www.mathworks.com
CT4008	NEWARE	https://en.neware.com.cn/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Zheng Chen (chen@kust.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All data reported in this paper will be shared by the lead contact upon request.

This work is an experimental study of state of health for lithium-ion battery and there is no code generated.

Any additional information required to reanalyze the data reported in this work paper is available from the Lead Contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The battery adopted in this investigation is a commercial prismatic lithium-ion battery with nickel-cobalt-manganese (NCM) as the cathode and graphite as the anode. The separator consists of polypropylene (PP) with ceramic coating.

METHODS DETAILS

Experimental battery parameters

The nominal capacity of the NCM/graphite battery used in this research is 4 Ah with the nominal voltage of 3.6 V.

Equipment used in the experiment

The charge/discharge machine used in this paper is CT4008, which is developed by Neware (Shenzhen) Co. LTD. During the experiments, the temperature was set at 25°C.

QUANTIFICATION AND STATISTICAL ANALYSIS

The original data are collected on an NEWARE Battery Testing System (CT4008). Figures are produced by Matlab from raw data.

ADDITIONAL RESOURCES

Any additional information about the cell assembly, tests and data reported in this paper is available from the lead contact on request.