Scalable and generalizable deep learning for battery state of health estimation in on-road electric vehicles

Hao Jing[a,b], Jianyao Hu[c], Shiqi (Shawn) Ou[a,b,*], Zhilong Lv[d], Renzhi Lyu[e], Jingyuan Zhao[f,*]

[a]School of Future Technology, South China University of Technology, Guangzhou 511442, Guangdong, China

[b]Guangdong Artificial Intelligence and Digital Economy Laboratory (Guangzhou), Guangzhou 510335, Guangdong, China

[c]China Electronic Product Reliability and Environmental Testing Research Institute, Guangzhou 510610, Guangdong, China

[d]Hubei Longzhong Laboratory, Hubei University of Arts and Science, Xiangyang 441000, Hubei, China

[e]School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100811, China

[f]Institute of Transportation Studies, University of California Davis, Davis 95616, CA, USA

*Corresponding authors.
E-mail addresses: sou@scut.edu.cn (S. Ou), jyzhao@ucdavis.edu (J. Zhao).

Abstract: Accurate battery health diagnostics are essential for timely maintenance, replacement, and the safe operation of electric vehicles (EVs). For on-road EVs, leveraging operational data for accurate state-of-health (SOH) estimation remains challenging due to varied degradation patterns across different driving conditions, vehicle types, and battery chemistries. Thus, developing an on-road-specific efficient feature system and a generalized SOH estimation framework adaptable to diverse EV models and chemistries is essential. To address these limitations, this study proposes a vehicle operational data-driven approach that integrates multi-dimensional feature fusion with a hybrid deep neural network architecture. Specifically, 12.83 million on-road data points spanning a wide range of vehicle types and battery chemistries are processed. Capturing representational, driving behavioral, and electrochemical characteristics, this study proposes a three-dimensional feature system comprising shallow, intermediate, and deep descriptors. To tackle challenges posed by long time spans and the limited effectiveness of Transformer models on multivariate inputs, a hybrid framework combining temporal convolutional networks with an enhanced iTransformer is developed, incorporating a differential attention mechanism to suppress attention noise. Experimental results demonstrate that the proposed method achieves high accuracy across two test sets, with an average R^2, MAPE, MAE, and RMSE of 98.88%, 0.35%, 0.31%, and 0.40%, respectively. This represents an 81.4% reduction in RMSE compared to the best-performing baseline. Data scarcity experiments using reduced training data demonstrate that even when the training set is decreased from 80% to 30%, model performance remains stable, with the RMSE remaining below 0.16%. Feature attribution analysis using Shapley additive explanations (SHAP) confirms the indispensability of all three feature dimensions, with driving behavior features being particularly influential. Following feature optimization, training time is reduced by 17.3%. This study presents a robust SOH estimation framework tailored for intelligent cloud battery management systems, proactive maintenance, and the safe operation of EV batteries in practical environments.

Keywords: Lithium-ion battery; SOH; Deep learning; Electric vehicles; iTransformer; On-road field data

1. Introduction

1.1. Background

Electric vehicles (EVs), known for producing zero tailpipe emissions, have emerged as a key strategy for reducing dependence on fossil fuels and advancing carbon neutrality. Over the past decade, the global EV industry has expanded rapidly, particularly in China, where annual production surpassed 10 million units by November 2024 [1–3]. Lithium-ion batteries, the primary energy storage technology for EVs, are widely adopted due to their high energy density and long cycle life [4,5]. However, the electrochemical processes intrinsic to lithium-ion batteries inevitably lead to irreversible aging over time. As aging progresses, battery capacity diminishes while internal resistance increases, compromising vehicle range, energy efficiency, and overall reliability. In addition to performance deterioration, battery aging can also elevate safety risks [6,7]. To ensure the long-term performance, safety, and lifespan of EV batteries, it is therefore critical to accurately assess battery aging, refine battery management system (BMS) strategies, and implement timely maintenance interventions [8,9]. The state of health (SOH) metric is commonly used to quantify battery aging, typically defined as the ratio of the current remaining capacity to the initial rated capacity, or alternatively, the ratio of the current internal resistance to that measured at the point of manufacture [10,11].

1.2 Literature review

SOH estimation methods are generally divided into direct measurement, model-based, and data-driven approaches [12–14]. Direct measurement methods, such as electrochemical impedance spectroscopy and ampere-hour integration [15–17], rely on high-precision instruments, involve long testing durations, and are susceptible to environmental disturbances, limiting their applicability in real-time scenarios [18,19]. Model-based approaches primarily include equivalent circuit models (ECMs) and electrochemical models [20–22]. The ECM simulates battery behavior using electrical components such as resistors and capacitors [23,24], while the latter captures electrochemical processes at the microscale with high accuracy. However, both methods face significant challenges in acquiring model parameters and adapting to dynamic conditions and environmental variability, which hinder their application in real-time SOH estimation for EVs [12,25]. The availability of extensive battery data, combined with advancements in computational resources, has significantly enhanced the effectiveness of data-driven methods in SOH estimation [26]. These methods

extract health features from operational data and utilize machine learning models to characterize the nonlinear relationship between these features and SOH [27]. Common approaches include conventional machine learning, deep learning, and hybrid models [28]. In traditional machine learning, support vector machines, Gaussian process regression, random forests, and decision trees are widely used for SOH estimation [29–32]. However, their performance tends to degrade as the dataset size increases [33]. In contrast, deep learning methods such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks can automatically extract features from large datasets and capture complex patterns [34–37]. Transformer models, which incorporate self-attention mechanisms, effectively overcome the limitations of recurrent architectures in handling long sequences [38,39]. To improve both estimation accuracy and model interpretability, physics-informed neural networks have been introduced [40]. Moreover, Che et al. introduced a self-supervised learning framework that improves SOH estimation accuracy using fewer labeled samples [41]. In addition, Tao et al. developed a physics-informed model that integrates thermodynamic and kinetic parameter estimation, achieving a 25-fold increase in prediction speed and maintaining an average accuracy of 95.1% under varying conditions, thereby demonstrating the efficacy of combining physics-based feature engineering with learning-based methods [42]. To further address the limitations of individual models, hybrid deep learning approaches have garnered attention [43,44]. For example, Zhao et al. combined CNNs with an enhanced Transformer to create a transfer learning framework, enabling rapid training and accurate predictions across various battery materials [45,46]. Similarly, Lv et al. integrated CNNs with a fast attention mechanism, achieving accurate SOH estimation while reducing GPU power consumption by over 50% compared to traditional Transformer models [47].

Despite these advances, the majority of the techniques mentioned above rely on laboratory-collected data under controlled conditions. These data are typically obtained using high-precision charge-discharge equipment designed for battery aging tests, which often involve lengthy test cycles. On the other hand, field data collected by BMS face challenges such as limited sampling frequency, incomplete charge-discharge cycles, and high data heterogeneity. However, due to their closer alignment with actual application environments, field data offer greater practical value. Accordingly, a study utilized six available field datasets and designed multiple independent feature extractors to capture aging information across different

scales, thereby enabling precise estimation of SOH under varying battery chemistries and operational conditions [48]. To address the challenge of estimating the remaining capacity of second-life batteries with minimal field data, another study proposed a deep generative transfer learning framework. This framework generates voltage response dynamics data across various SOC conditions. It employs a deep correlation alignment (CORAL) method to rapidly and accurately estimate remaining capacity using only limited field data [49]. However, despite the aforementioned data experiments being as close as possible to actual operation, the operation of EVs in real-world road conditions is even more complex [50]. The diverse and highly dynamic influencing factors present on real roads, such as constantly changing ambient temperatures, relative humidity, and load variations caused by road condition uncertainties, as well as importantly, differences in driving behavior, including charging habits, significantly impact battery performance and aging processes [51,52]. Furthermore, models based solely on cell-level data have certain limitations, as a battery pack consists of multiple cells, each of which may exhibit performance discrepancies due to manufacturing variations. At the same time, the overall performance of the battery pack in the vehicle is also influenced by operational factors such as thermal management, energy scheduling, and system coordination, which are difficult to accurately simulate under cell-level field data conditions [53]. To obtain more accurate and reliable battery SOH estimates for EVs, it is crucial to model operational data at the pack level in on-road conditions. Accordingly, several studies have applied models such as the gated recurrent unit (GRU), XGBoost, LSTM, and SVR to estimate battery capacity in actual vehicles [51,54–56]. Similarly, Zhao et al. employed particle swarm optimization to fine-tune the hyperparameters of an extreme learning machine neural network for accurate SOH estimation [57]. However, these studies are often limited to a small number of the same vehicle types, with training and testing sets usually divided within the same vehicle, lacking separation between cross-type or cross-vehicle datasets, which limits the generalization and application of the models. Thus, Deng et al. utilized operational data from 20 EVs to achieve a prediction error of less than 1.6% using a sequence-to-sequence architecture and over 20 features [58]. In summary, despite promising progress, existing methods often fall short in generalization to on-road vehicle operations involving diverse vehicle types and battery chemistries, as well as in efficient feature engineering, which motivates this study.

1.3 Contribution of this work

Despite recent advances in battery health estimation, several challenges persist in on-road applications. First, most existing methods remain focused on SOH estimation at the cell level. Research on full vehicle

battery systems is limited, with most studies relying on modeling using a single battery chemistry and vehicle type. The generalization capability of these models remains insufficiently validated, and only a small fraction of studies have made on-road vehicle data publicly available. Moreover, a critical challenge lies in balancing feature redundancy and computational efficiency when extracting health features from real-vehicle datasets. Several studies incorporate over 20 features [53–56,58], which significantly increases both computational cost and model complexity. Such high-dimensional feature sets also introduce risks associated with the curse of dimensionality, including overfitting, which limits their practical deployment. Furthermore, studies based on road operation battery pack-level data often rely on numerous representational parameters, failing to effectively utilize highly relevant driver behavior information from on-road vehicles, and lacking an in-depth exploration of electrochemical characteristics. Moreover, most existing studies do not systematically evaluate model performance under data-scarce conditions, which is particularly critical given the practical constraints of data availability and computational resources. Finally, the interpretability of most data-driven models remains limited. These models often function as "black boxes," providing outputs without transparent reasoning. Understanding how such models arrive at specific predictions or decisions remains a major barrier to trust and adoption. To address these challenges, this study proposes a three-dimensional health feature system tailored for real-vehicle applications, combined with a temporal convolutional network (TCN)– a Specialized iTransformer hybrid deep learning architecture. As shown in Fig. 1, this framework enables accurate SOH estimation and robustness under data scarcity for EVs under on-road operating conditions. The main contributions are summarized as follows.

(a) A comprehensive dataset is compiled from the historical operational data of ten EVs, spanning diverse vehicle types and battery chemistries. This high-quality dataset will be made publicly available to support broader research efforts, promote transparency, and enhance academic impact.

(b) A compact and interpretable three-dimensional feature system is developed, comprising only seven variables that capture information from vehicle-level representation, driver behavior patterns, and internal electrochemical parameters.

(c) To address the challenges posed by long time-series data and high feature dimensionality, a hybrid deep learning model is proposed that combines a TCN with an iTransformer. This model incorporates a differential attention mechanism to mitigate attention noise and enhance robustness. The framework is

5

validated across multiple vehicle types and datasets, demonstrating strong generalization and robustness in the face of data scarcity.

(d) Model interpretability is enhanced through Shapley additive explanations (SHAP), which quantify the contribution of each feature dimension to the estimation outcome. This analysis informs feature optimization, leading to improved model efficiency and enhanced SOH estimation accuracy.
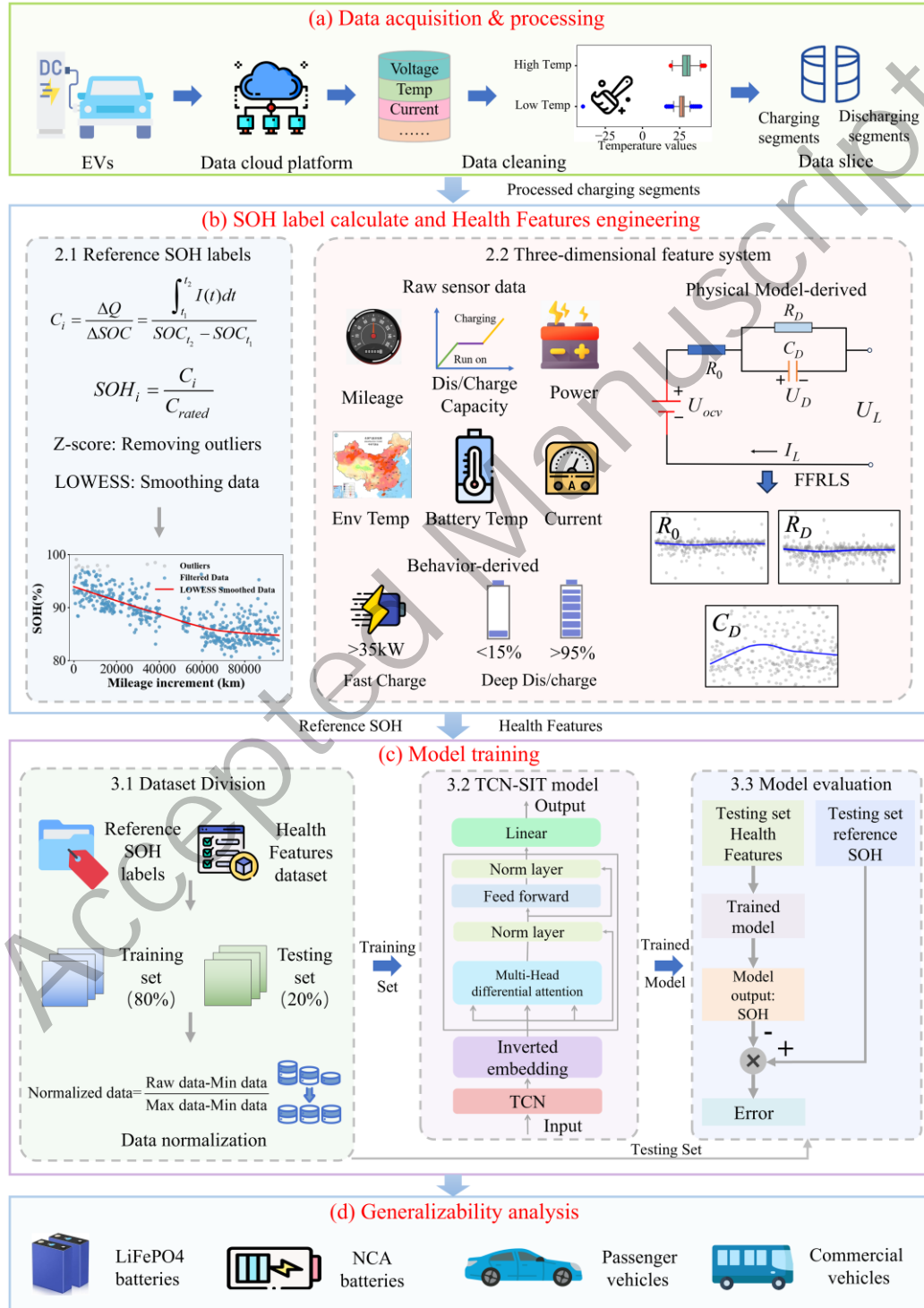

Fig. 1. Research framework.

1.4 Organization of the paper

This paper is structured into five sections. Section 1 introduces the research background and motivation. Section 2 describes the data sources, provides an overview of the dataset, and outlines the data processing workflow. Section 3 explains the generation of SOH labels, the construction of the three-dimensional health feature framework, and the architecture of the proposed hybrid neural network model. Section 4 presents the SOH estimation results and includes a SHAP-based interpretability analysis. Finally, Section 5 summarizes the key findings and implications of the study.

2. Data acquisition and pre-processing

2.1 Source of data

To better represent on-road operating conditions of EVs, this research utilizes data obtained from the *Guangzhou New Energy Intelligent Vehicle Big Data Monitoring Platform*. The study involves 10 fully EVs, with a total of approximately 12.83 million data points collected over one year. To comprehensively improve and validate the robustness and generalization ability of the algorithm, these 10 vehicles represent different types of EVs, including both passenger cars and commercial vehicles, and the battery materials include lithium iron phosphate (LFP) and nickel cobalt manganese (NCM) batteries. The sensor signals of these vehicles are transmitted to the BMS via the in-vehicle controller area network (CAN) bus, and then uploaded by the onboard telematics box (T-BOX) at a fixed sampling frequency to the signal base station, which subsequently transfers the data to the original equipment manufacturer (OEM) and the big data monitoring platform for storage. Table 1 presents the specific information for each vehicle. The collected data includes over 70 signals such as data sampling time, drive motor data, power battery data, and vehicle speed. Considering that an excessive amount of data may lead to memory overflow and prolonged processing times, this study retains only 11 data points closely related to battery health, which include: data sampling time, vehicle speed, charging status, cumulative mileage, total voltage, current, state of charge (SOC), and cell-level parameters such as the minimum and maximum voltage and temperature, and battery pack structure information is provided in Supplementary Note 1. Fig. 2 illustrates the distribution of raw sampling data for vehicle #1 for one year. It shows that the charging conditions and environmental temperature during on-road operation significantly differ from those in laboratory tests, presenting more complex characteristics.

Table 1. Overview of sampled vehicle characteristics.

Vehicle number	Vehicle type	Battery material	Initial rated capacity/Ah	Number of data points	Cumulative Mileage/km	Sampling frequency/Hz
Vehicle#1	Passenger vehicle	NCM	150	954754	69043	0.1
Vehicle#2	Passenger vehicle	NCM	150	998243	73950	0.1
Vehicle#3	Passenger vehicle	NCM	160	997098	79440	0.1
Vehicle#4	Passenger vehicle	NCM	160	1150999	96279	0.1
Vehicle#5	Passenger vehicle	NCM	160	1096073	114431	0.1
Vehicle#6	Passenger vehicle	NCM	160	501031	27318	0.1
Vehicle#7	Passenger vehicle	LFP	120	5304111	32496	0.5
Vehicle#8	Electric bus	LFP	645	675236	82668	0.1
Vehicle#9	Electric bus	LFP	645	443806	43988	0.1
Vehicle#10	Electric bus	LFP	505	715956	27677	0.1

2.2 Data cleaning and processing

Laboratory environments offer stable, interference-free conditions for battery testing, whereas on-road vehicle sensor data are subject to environmental disturbances such as weather, surrounding infrastructure, and data transmission inconsistencies. Consequently, the quality of data uploaded to cloud platforms is often variable. As illustrated in Fig. 2(a, b), unlike the full SOC cycling between 0% and 100% typically performed in laboratories, on-road charging events usually begin around 30%–40% SOC and terminate at 80%–90%. Fig. 2(c) shows that ambient operating temperatures range from approximately 0 to 30°C, in contrast to the controlled, constant-temperature conditions of laboratory tests. Fig. 2(d, e) highlight the dynamic variations in voltage and current during vehicle operation, underscoring the complexity of on-road load profiles compared to the constant loads used in lab environments. Fig. 2(f–i) depict intra-pack variations in cell voltages and temperatures. Notably, Fig. 2(f) shows some cell voltage readings exceeding 6V—values that violate the physical limits of lithium-ion chemistry—suggesting likely sampling or transmission errors. Fig. 2(h, i) indicate that cell temperatures fluctuate between 10 and 40°C, driven by external conditions and usage patterns. In summary, vehicle-collected field data differ substantially from laboratory benchmarks and often contain noise and anomalies. Thus, rigorous data cleaning is imperative to ensure the reliability and utility of field-collected battery datasets.
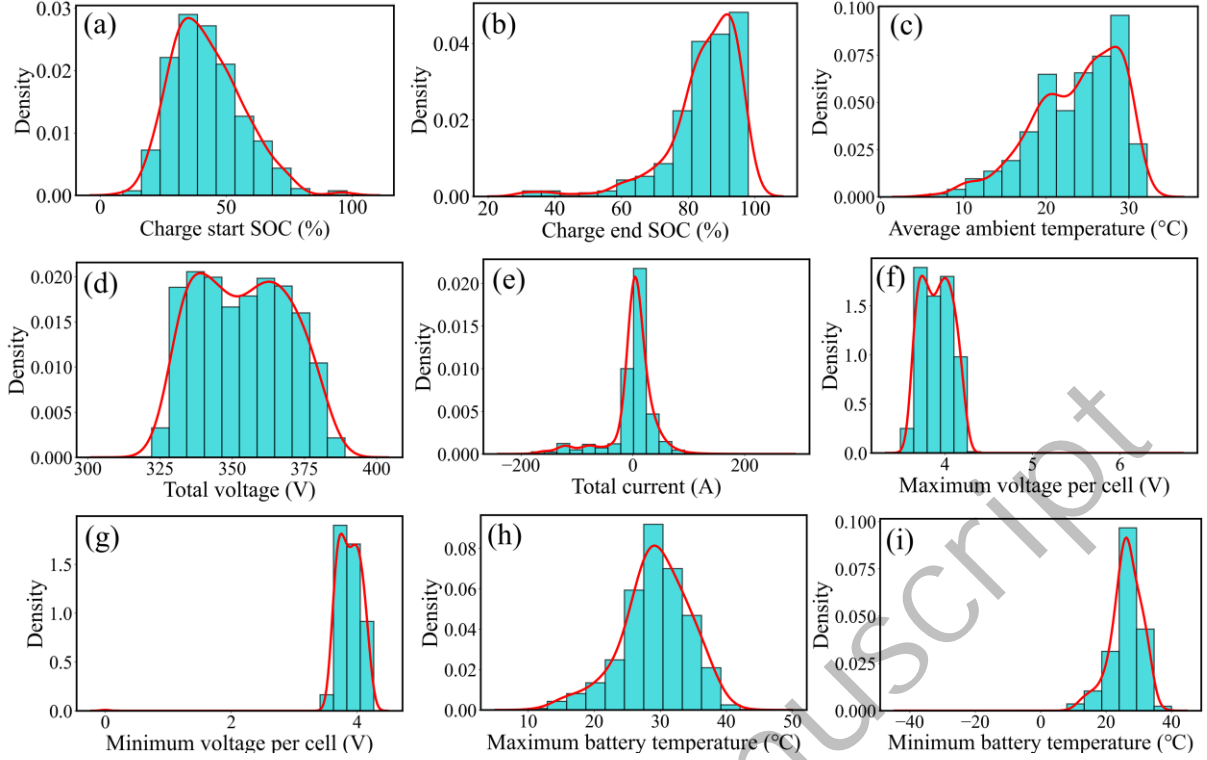
Fig. 2. Raw data distributions for Vehicle #1. (a) Starting SOC during charging. (b) Ending SOC during charging. (c) Ambient temperature. (d) Total battery pack voltage. (e) Total battery pack current. (f) Maximum single-cell voltage. (g) Minimum single-cell voltage. (h) Maximum battery temperature. (i) Minimum battery temperature.

The data cleaning process in this study consists of six steps, shown in Fig. 3. First, duplicate data is removed based on the sampling time. Since environmental temperature significantly impacts battery performance and can vary greatly under actual vehicle operating conditions, the second step involves querying historical environmental temperature data using Visual Crossing Weather based on vehicle location information, and matching it with the corresponding vehicle data [59]. In the third step, outliers are detected using the Interquartile Range (IQR) approach and subsequently substituted with adjacent values. After these steps, the next step is to categorize the vehicle data into driving, charging, and resting segments based on the vehicle charging status and speed information. Since current fluctuations are significant and intense during vehicle operation, and due to the large sampling intervals, using driving segments to calculate health features and labels may lead to substantial errors. In contrast, as shown in Fig. 3(d), the current variation in charging segments is stepwise and remains stable for a period. The features and labels calculated from these segments are more accurate [53,57,58]. Therefore, this study extracts features and calculates labels based on the processed charging segments. However, during charging, certain charging segments are short in duration

and show minimal changes in SOC, which may lead to insufficient internal reactions, unstable chemical properties, and large fluctuations in the calculation results. Thus, this study sets a threshold for filtering charging segments, retaining only those where the SOC changes exceed 20% [53,57]. The filtered charging segments undergo missing value detection and are then filled using linear interpolation. After processing through these six steps, the segments are used as the dataset for feature extraction and label calculation, then for model training.
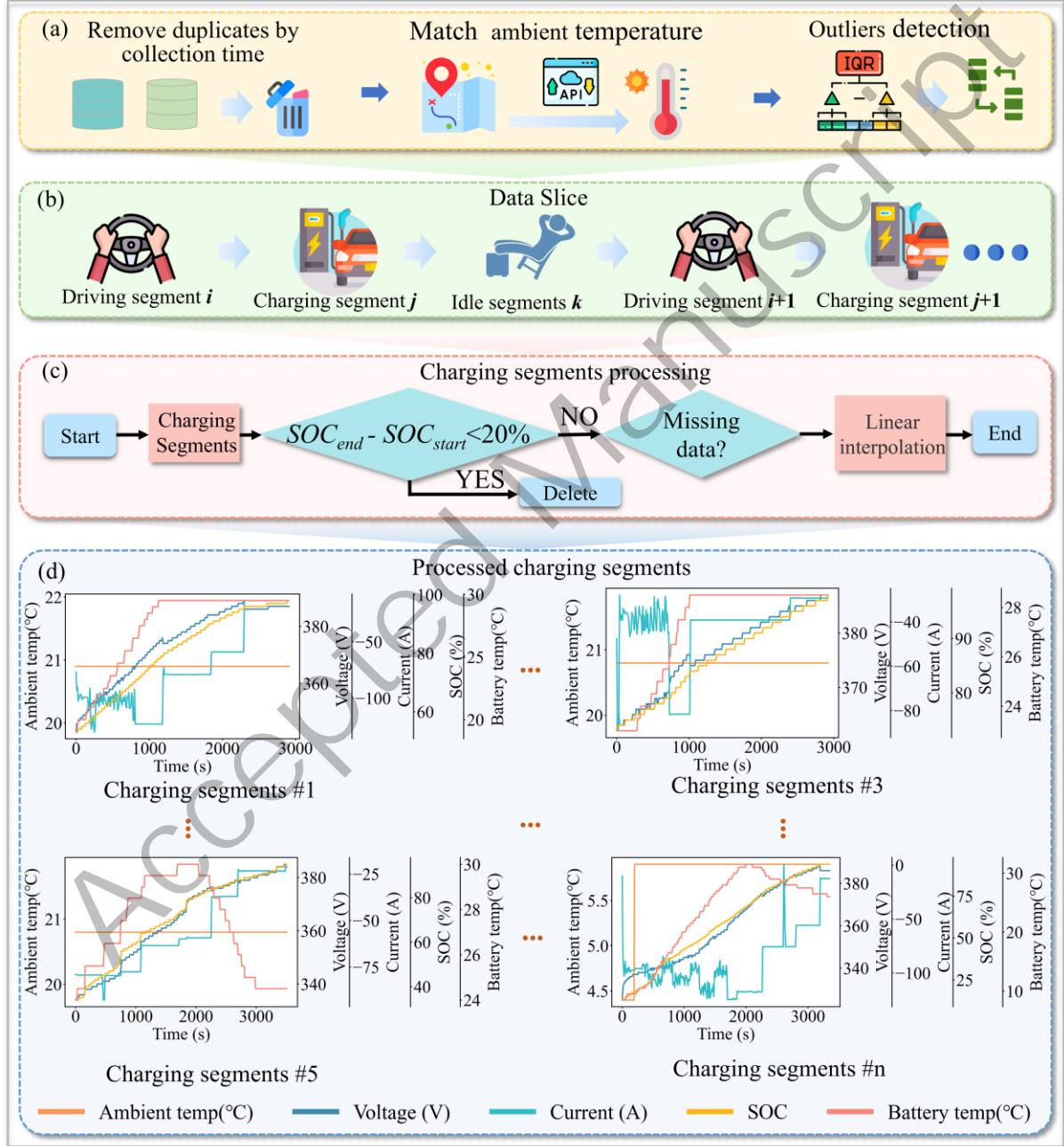
Fig. 3. Data cleaning steps. (a) Three steps: removal of duplicates, matching with environmental temperature, and outlier handling. (b) Data slicing. (c) Two steps: removal of invalid segments and interpolation of missing values. (d) Visualization of processed segments.

3. Method for on-road SOH estimation

After preprocessing, the raw vehicle data is converted into stable charging segments, from which SOH labels and health features can be extracted. These extracted labels and features are then used for model construction and training.

3.1 Reference SOH calculation

The SOH of a battery is generally quantified as the ratio between its present full charge capacity and the original capacity specified at the time of manufacture [53,57]. Since this value cannot be directly obtained from the data collected from the vehicle, it needs to be indirectly calculated using the collected current and SOC data. The detailed calculation methods are provided in Equations (1) and (2) [58]

$$SOH = \frac{C_{now}}{C_{rated}} \tag{1}$$

$$C_{now} = \frac{\int_{t_1}^{t_2} I(t)dt}{SOC_{t_2} - SOC_{t_1}} = \frac{\int_{t_1}^{t_2} I(t)dt}{DOC_{t_{12}}} \tag{2}$$

where C_{now} denotes the present full charge capacity of the battery, while C_{rated} indicates the factory-specified rated capacity. $I(t)$ represents the instantaneous current during vehicle operation, and SOC is used to indicate the state of charge. t_1 and t_2 correspond to the start and end times of the charging process, respectively. DOC refers to the depth of charge. As shown in Fig. 3(c), this study has filtered out segments with a DOC less than 20%, as smaller DOC values lead to fluctuations in the computed values and larger errors. In comparison, larger DOC values result in too few remaining segments [53,57].

Through the above steps, SOH labels for the ten vehicles are calculated. Fig. 4 displays the SOH labels for four randomly selected vehicles from this set. However, the computed SOH labels exhibit considerable noise, which can be attributed to factors such as sensor inaccuracies, low data collection frequency, and signal transmission interference. Therefore, preprocessing of the SOH labels is necessary [53,57,58]. To address this, the study first applies a Z-score filtering algorithm to detect and remove outliers in the capacity data, effectively minimizing the impact of anomalous points on the overall capacity curve. The Z-score algorithm computes the standardized score of each data point, identifying and eliminating values that deviate significantly from the mean. This approach removes extreme outliers while preserving the core data structure,

thereby improving the robustness of subsequent analysis. The Z-score formula is provided in Supplementary Note 2.
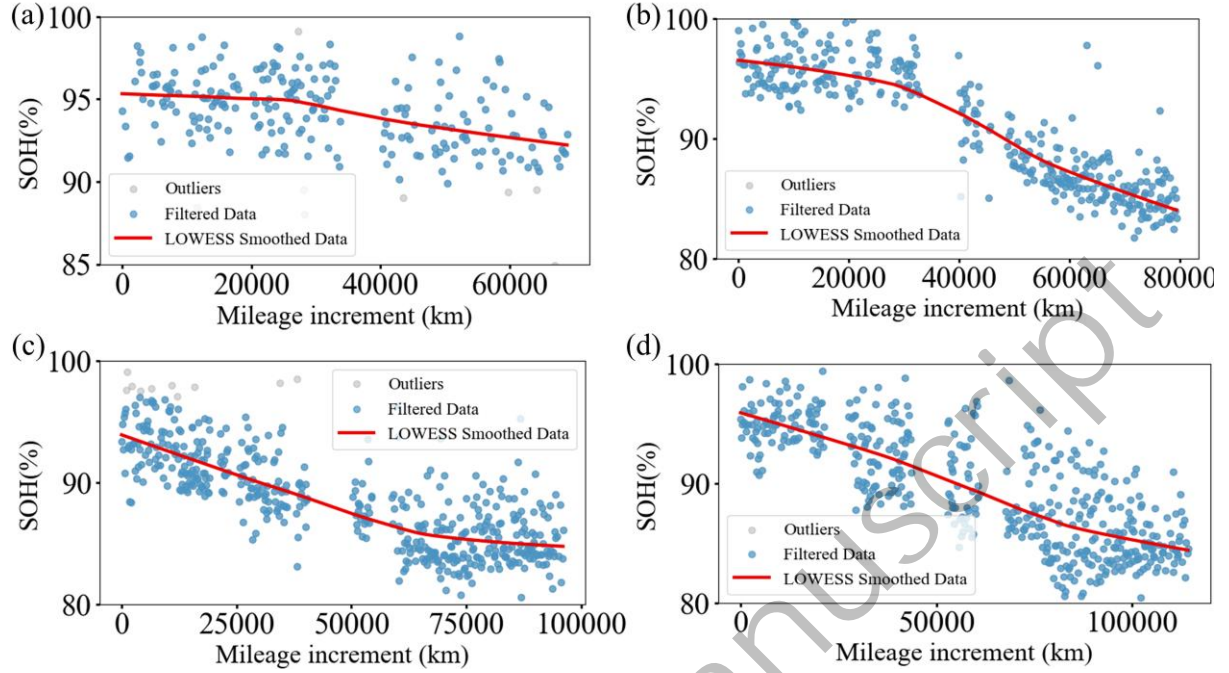
Fig. 4. SOH reference labels calculated for partial vehicles. (a) SOH reference labels of Vehicle#1. (b) SOH reference labels of Vehicle#3. (c) SOH reference labels of Vehicle#4. (d) SOH reference labels of Vehicle#5.

As shown in Fig. 4, the SOH labels still exhibit significant fluctuations after removing the gray outliers. For on-road vehicle applications, focusing on the SOH value of each charging segment is unnecessary, as it can be significantly influenced by factors such as the environment, driving habits, and charging station locations. Instead, the focus for EVs should be on the degradation trend of the SOH [52,57,58]. As a result, it is essential to apply smoothing to the SOH values in order to qualitatively characterize the degradation trend. This study uses the locally weighted scatterplot smoothing (LOWESS) algorithm for data smoothing [60]. The red solid line illustrates the smoothing effect in Fig. 4. This indicates that the degradation patterns among different vehicles vary significantly, which may result from multiple influencing factors, including ambient temperature, individual driving behavior, and other environmental conditions. The validation of the proposed label calculation method is presented in Supplementary Note 3.

3.2 Health feature extraction

The selection of appropriate health features can significantly enhance the effectiveness of SOH estimation [1,54–58]. This study utilizes the processed EV charging segments to propose a three-dimensional feature system, comprising shallow, intermediate, and deep-level features, for the first time. This system

encompasses a range of multidimensional features, including battery voltage, current, statistical user charging and discharging habits, and internal resistance and capacitance, aiming to capture the effects of various factors on battery degradation characteristics in a comprehensive and multi-layered manner.

3.2.1 Shallow dimensional features—raw sensor health features

It first extracts relevant health features based on direct sampling data of the vehicle. As is well-known, with increasing battery cycle counts, irreversible changes in the crystalline structure of internal electrode materials can occur, which is the main cause of gradual battery performance degradation [1,2]. However, unlike laboratory data, the charge-discharge behavior of on-road vehicles does not always follow standard cycle patterns. As shown in Fig. 2(a, b), vehicles often do not fully discharge and recharge the battery during each use. Therefore, to better reflect actual EV usage scenarios, this study selects cumulative mileage and cumulative charge-discharge capacity, which are highly correlated with battery cycle counts, as health features [57,61]. Cumulative mileage reflects the operational load on the battery during actual use. In contrast, cumulative charge-discharge capacity directly quantifies the total energy transferred to the battery, which can intuitively capture key factors of battery degradation.

Furthermore, the health status of the battery is closely related to current intensity and charging methods, particularly high currents and fast charging behaviors, which can significantly damage the battery. Excessive current can generate high heat within the battery, accelerating the aging of the electrode material and electrolyte decomposition. Additionally, excessive current may trigger the formation of lithium dendrites, which, in severe cases, can cause internal short circuits, thereby shortening the lifespan of the battery [62]. Based on these influencing factors, this study incorporates the average current and average power calculated for each charging segment as health features.

Temperature is considered a crucial external factor that impacts both the efficiency and lifespan of lithium-ion batteries. Previous research indicates that these batteries achieve optimal performance when operating within the range of 20 to 25°C. Elevated temperatures can accelerate electrolyte degradation and electrode aging, while excessively low temperatures reduce lithium-ion mobility and raise internal resistance [63]. To reflect the influence of temperature on battery condition, this study incorporates the deviation of the average battery temperature, collected by the BMS, from 25°C as a model input. Additionally, ambient

temperature data are incorporated to assess the potential impact of external conditions on battery performance.

3.2.2 Intermediate dimensional features—behavior-derived health features

The key difference between real-vehicle and laboratory battery data lies in user variability. In laboratory settings, batteries are typically tested under standardized operating procedures, whereas in real-life usage, habits vary across EV users. Even with the same vehicle type, the battery degradation trends can vary significantly between different users. Studies have shown that, with L1 slow charging and DC fast charging, battery SOH after 10 years can differ by more than 20% under the same driving mode [62]. Therefore, it is essential to identify how various user behaviors collectively influence battery condition. Given that the dataset is organized by charging sessions, this study begins by analyzing the frequency of fast charging cycles. Fast charging is defined as the number of charging sessions where the average charging power exceeds 35 kW. Additionally, deep discharge and charge cycles cause irreversible damage to lithium batteries, as widely studied and confirmed [64,65]. This study includes the cumulative number of deep discharge and charge cycles as important feature variables. Deep discharge is defined as discharging with SOC below 15%, while deep charging is defined as charging with SOC above 95%.

3.2.3 Deep dimensional features—physical model-derived health features

Deep dimensional features are constructed based on shallow and intermediate features, allowing further investigation into the internal electrochemical mechanisms of the battery. These processes involve extracting key parameters, including internal resistance and polarization capacitance. As EVs operate, the resistance of the battery pack tends to rise with the decline of SOH, a phenomenon closely related to changes in electrochemical characteristics [38,51]. As the battery ages, the internal chemical reactions become increasingly incomplete, and electrode materials may undergo crystallization or morphological changes. At the same time, the electrolyte concentration may decrease, leading to an increase in internal resistance. Additionally, as the number of charge and discharge cycles increases, the electrolyte inside the battery may degrade or deplete, reducing ionic conductivity and further increasing the internal resistance of the battery. Due to the increased internal resistance, the energy conversion efficiency during charge and discharge cycles decreases, manifesting as a gradual decline in battery voltage and capacity. Therefore, internal resistance, an

important feature of battery degradation, reflects the degree of battery aging and typically increases as SOH decreases [51].
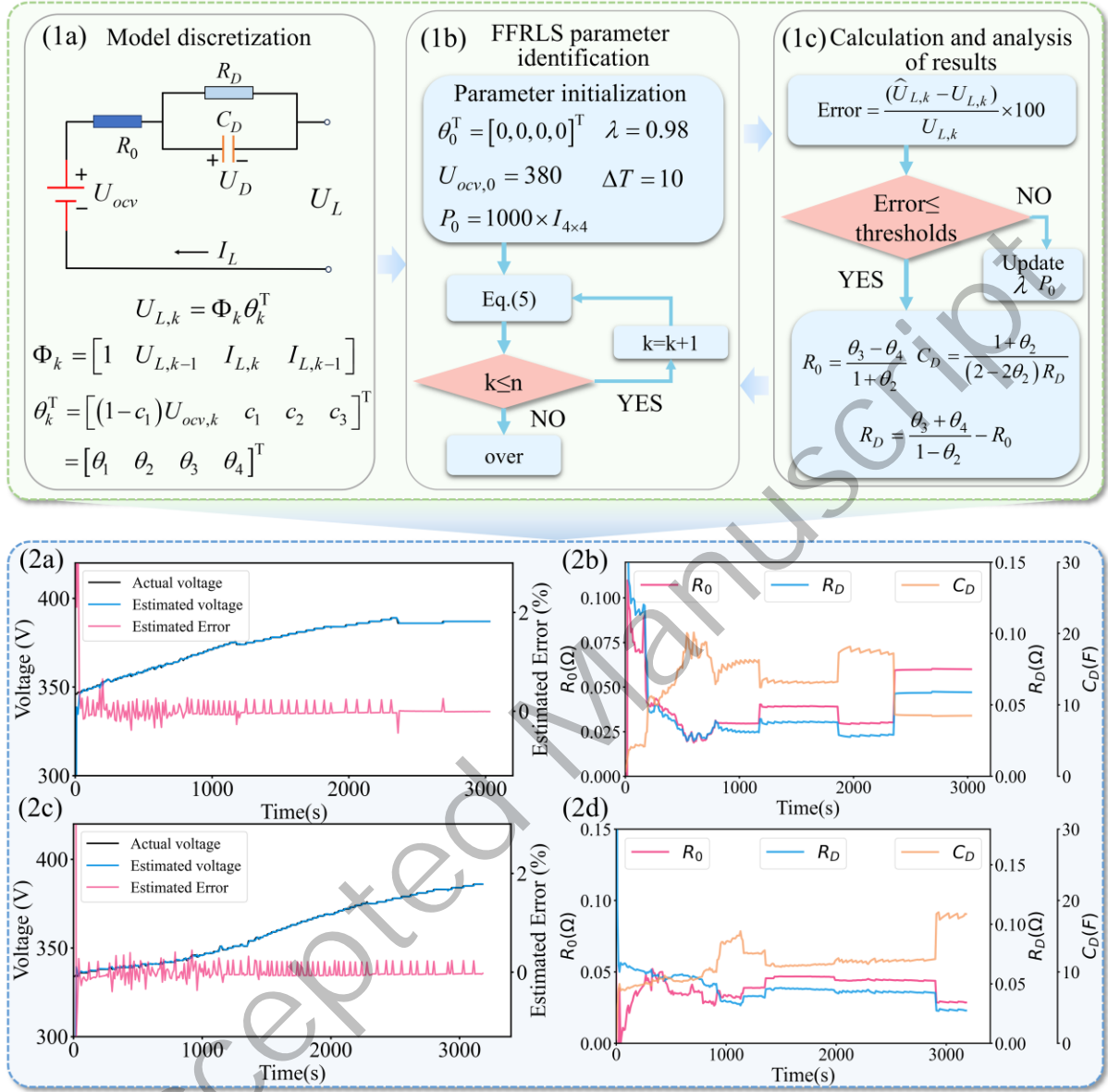
Fig. 5. Parameter identification based on equivalent circuits to obtain deep internal health features. (1a) Equivalent circuit model and its discretized equations. (1b-1c) Flowchart of the FFRLS-based parameter identification process. (2a, 2c) Voltage estimation results and corresponding errors for two arbitrarily selected segments. (2b, 2d) Identified values of three parameters corresponding to the two segments.

The ECM is commonly used to simulate the dynamic behavior of internal resistance and other characteristics in lithium batteries. Typical ECMs include the Rint, Thevenin, and RC models [54]. Among them, the Thevenin model is frequently adopted in battery modeling due to its high accuracy in simulation

and relatively low complexity and parameter count. Thus, this study uses the Thevenin model for battery pack modeling and parameter identification. The specific flowchart is shown in Fig. 5. The Thevenin model, as illustrated in Fig. 5(1a), consists of polarization capacitance C_D, polarization resistance R_D, and polarization voltage U_D, with load current and voltage denoted as I_L and U_L, respectively, and open-circuit voltage and internal resistance denoted as U_{ocv} and R_0. Based on Kirchhoff current-voltage law, the mathematical expressions for the Thevenin model are given in Equations (3) and (4).

$$C_D \frac{\mathrm{d}U_D}{\mathrm{d}t} + \frac{U_D}{R_D} = I_L \tag{3}$$

$$U_L = U_{ocv} + U_d + I_L R_0 \tag{4}$$

To apply the above mathematical model for parameter identification, it needs to be discretized, as shown in Fig. 5(1a). The recursive least squares (RLS) algorithm is widely used for identifying model parameters. Its core principle involves continuously refining prior predictions by incorporating newly available data at each time step, which allows the algorithm to determine parameter values corresponding to the current moment. The new parameter estimate is the sum of the old parameter estimate and the correction term. However, traditional least squares methods suffer from a "filter saturation" phenomenon, where, as the iteration number increases, the values of the gain K and the covariance matrix P gradually decrease, weakening the ability of the algorithm to correct the data and increasing the parameter identification error. To address this issue, the forgetting factor recursive least squares (FFRLS) algorithm is introduced. Due to its lightweight structure and high computational efficiency, the FFRLS algorithm has been validated and is widely integrated into real-world BMS implementations for EVs, supporting reliable and efficient online parameter identification [66,67].The forgetting factor is a number between 0 and 1, usually close to 1, which serves to diminish the influence of old data and enhance the impact of new data, thereby preventing the occurrence of the "filter saturation" phenomenon [54]. As the forgetting factor increases, the parameter identification results generally become more stable and robust, but the response speed may decrease. Therefore, a balance between identification accuracy and response speed needs to be achieved. After multiple rounds of tuning, the forgetting factor is set to 0.98. In this study, the FFRLS method is employed for parameter identification of real vehicle battery packs, with the core of the algorithm presented in Equation (5).

$$
\begin{cases}
\hat{\theta}_k = \hat{\theta}_{k\text{-}1} + K_k \left(y_k - \Phi_k \hat{\theta}_{k\text{-}1} \right) \\
P_k = \dfrac{\left(I - K_k \Phi_k \right) P_{k\text{-}1}}{\lambda} \\
K_k = \dfrac{P_{k\text{-}1} \Phi_k}{\left(\lambda + \Phi_k P_{k\text{-}1} \Phi_k^T \right)}
\end{cases}
\tag{5}
$$

where, λ is the forgetting factor, $\hat{\theta}$ represents the estimated parameter matrix θ, K_k represents the gain matrix, P_k denotes the covariance matrix, and I stands for the identity matrix. During algorithm execution, as shown in Fig. 5(1b, 1c), the parameters are first initialized, and then Equation (5) is used iteratively. The parameters are adjusted based on the error until the identification error reaches an acceptable range. To verify the effectiveness of the proposed method, the identification results for two randomly selected data segments are illustrated in Fig. 5(2a–2d). Fig. 5(2a, 2c) show that the identification errors are both less than 1%. Specifically, in Fig. 5(2a), the estimated load voltage closely matches the actual measured voltage curve, indicating that the algorithm provides an accurate estimate of the load voltage during the identification process. Furthermore, the identification results in Fig. 5(2c) also exhibit minimal error, further demonstrating the high precision and reliability of this method in real vehicle battery pack parameter identification. Experimental findings indicate that the FFRLS algorithm effectively mitigates the filter saturation issue and demonstrates strong performance in terms of robustness and precision for identifying a complex battery pack model. Parameter identification is discussed in more detail in Supplementary Note 4.

3.2.4 Health features overall analysis

The three-dimensional feature extraction process extracts 11 relevant health features, as shown in Table 2, with specific classifications and information. Pearson correlation is applied to examine the feature distribution, the linear relationships among features, as well as those between features and labels, as illustrated in Fig. 6 [53]. The formula for the Pearson correlation coefficient is provided in Supplementary Note 5. The specific values in Fig. 6 represent the Pearson correlation between features and between features and labels. As expected, a strong negative correlation exists between battery capacity, cumulative mileage, fast charge counts, and internal resistance with SOH. However, the average current, power, and temperature during the charging phase show a weak correlation with SOH, respectively. This does not suggest that these factors are irrelevant to SOH, but rather reflects the limitation of the Pearson correlation coefficient, which

is only capable of identifying linear associations. To explore this aspect further, Section 4 employs SHAP analysis for additional validation.

Additionally, the data distribution of features and labels is shown along the diagonal in Fig. 6 for a more intuitive observation of the distribution characteristics of each variable. For example, the red box in the Fig. 6 shows that the average charging current of users follows an approximately normal distribution, with most values concentrated in a moderate range of charging currents. This suggests that the charging behavior of most users tends to be stable, with current values that are neither extremely high nor low, which aligns with typical charging habits.

In the lower left part of the Fig. 6, the linear relationships between features and between features and labels are further displayed. This section not only helps reveal the basic associations between variables but also provides valuable insights for subsequent modeling and analysis. For example, the yellow box in Fig. 6 clearly shows a significant positive linear correlation between cumulative mileage and internal resistance. Specifically, with an increase in cumulative mileage, the internal resistance of the battery pack progressively rises. This pattern indicates performance degradation of the battery under prolonged usage. As mileage accumulates, the battery experiences more frequent charge and discharge cycles. This leads to less ideal internal chemical reactions and a subsequent rise in internal resistance, which negatively influences overall battery performance. This finding is consistent with expectations and further proves that battery aging is closely related to its service life, particularly evident in the increase in internal resistance.

Table 1. Overview of the health features.

| Feature dimension	Concrete features description	Abbreviation
Shallow dimensional features	Accumulated mileages	M_{total}
	Cumulative charge/discharge capacity	C_{total}
	Average charging current	C_{mean}
	Average charging power	P_{mean}
	Ambient temperature	T_a
	BMS collected battery temperature absolute deviation from 25°C	T_{offset}
Intermediate dimensional features	Fast charge count	FCC
	Deep charge-discharge count	DCC
Deep dimensional features	Ohmic resistance	R_0
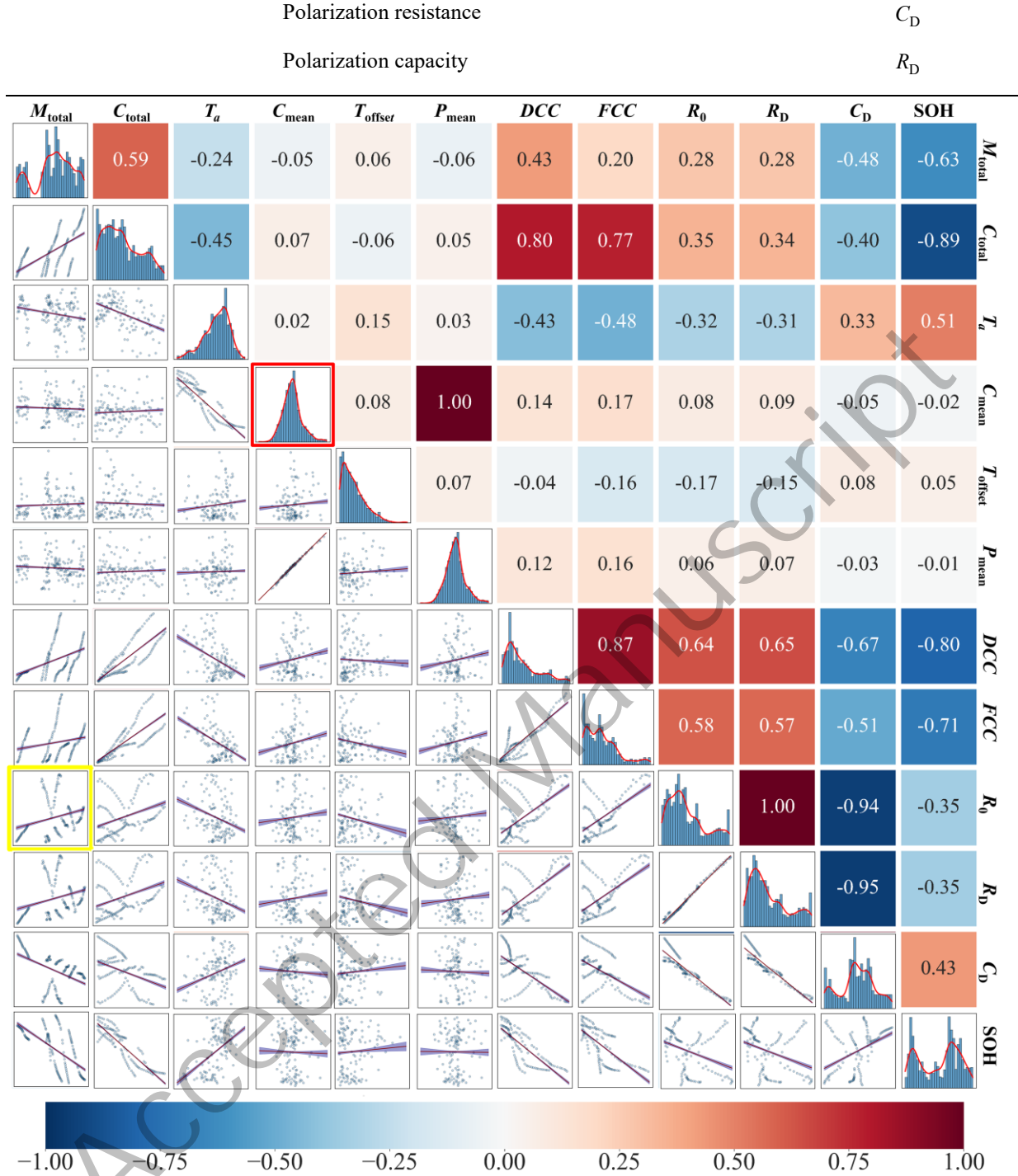
Fig. 6. Pearson correlation analysis results between SOH and the input features. The diagonal displays the distribution plots of each feature and the target variable. For example, the red box highlights the distribution of average charging current. The lower-left panel contains scatter plots illustrating the linear regression relationships between pairs of features and between features and the target. In the yellow box, for instance, the X-axis represents accumulated driving mileage and the Y-axis represents ohmic resistance, which together exhibit a positive linear correlation.

3.3 TCN- iTransformer for SOH estimation

3.3.1 Model development

As shown in Fig. 7, to address the challenges faced by real-vehicle SOH estimation, such as long periods, multiple features, and data noise, a TCN-iTransformer hybrid deep learning model is proposed. This model primarily consists of a TCN and an iTransformer module. First, the TCN module can effectively capture long-term dependencies, offering high computational efficiency and parallel processing capability, which makes it well-suited for processing battery datasets spanning extended periods [68]. Next, the iTransformer module addresses the issue of multi-feature input by mapping time-series data of the same variable into high-dimensional feature representations via an inverted embedding module, overcoming the information loss and computational bottleneck issues that traditional Transformers face when processing multivariate time-series data [69]. To address the noise issue in the traditional attention mechanism, a multi-head differential attention mechanism is introduced, which reduces attention noise interference by computing the differences between attention spectra, thereby enhancing the robustness of the model [70]. Finally, the projection layer integrates the extracted features and outputs the final SOH estimation results.
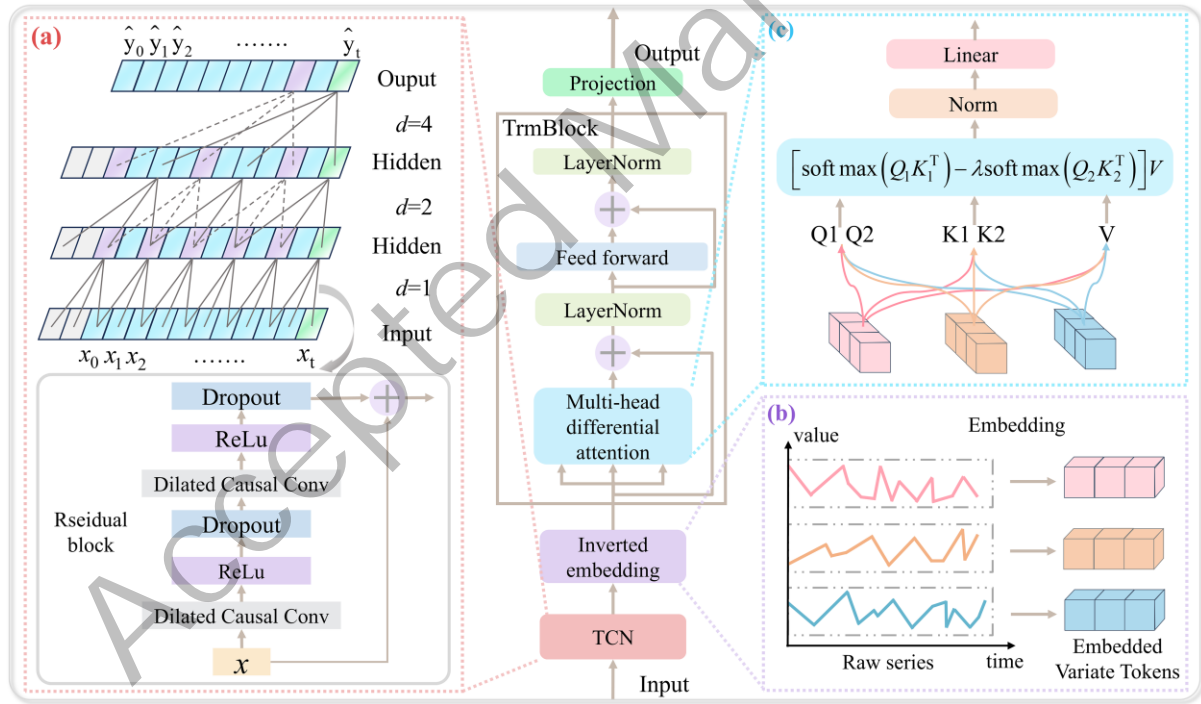
Fig. 7. TCN-specialized iTransformer model architecture diagram. (a) TCN structure diagram, including causal convolutions, dilated convolutions, and residual blocks. (b) Inverted embedding schematic diagram. (c) Differential attention mechanism schematic diagram.

3.3.2 Temporal convolution network

TCN has been shown to be effective in addressing certain time-series tasks, outperforming traditional models such as RNN and LSTM in specific scenarios [68,71]. Especially in real-vehicle SOH estimation, when facing challenges like long time spans and large data volumes, TCN can significantly enhance network running efficiency through large-scale parallel processing, making it better suited to handle complex multivariate time-series data. Moreover, TCN can adaptively adjust the receptive range of the network by modifying the dilation rate. As illustrated in Fig. 7(a), the key structural elements of TCN consist of causal convolutions, dilated convolutions, and residual connections. The corresponding formula is given in Equation (6).

$$f(t) = \sum_{i=0}^{k-1} w_i \times x_{t-d \bullet i} \tag{6}$$

where x_t is the input time series, $f(t)$ is the convolution output, w_i is the convolution kernel weight, d is the dilation rate, and k represents the kernel size. The details of the TCN are provided in Supplementary Note 6.

3.3.3 iTransformer and multi-head differential attention

Real-vehicle SOH estimation is a multi-feature time-series problem. However, conventional Transformer models do not fully exploit their potential when dealing with multivariate time-series data [69]. The iTransformer, as a novel sequence prediction model, is based on the original Transformer encoder architecture, as shown in Fig. 7, the architecture comprises an inverted embedding layer, followed by an attention module, the normalization component, a feedforward network, and a projection unit, with no modifications to the individual Transformer modules. Instead, it performs an inversion-based modeling approach on top of the original Transformer. Additionally, iTransformer innovatively adopts an inverted perspective during the variable input phase, mapping the entire time series of the same variable into a high-dimensional feature representation. This approach changes the traditional sequence embedding method, where the resulting feature vector is centered around the variable itself, independently showcasing the historical process it reflects. This allows the attention mechanism to function more effectively, capturing the correlations between variables more efficiently. Subsequently, the feed-forward network and layer normalization collaborate to eliminate range discrepancies between different variable measurement units, enabling the model to learn sequence features suitable for time-series predictions. Finally, the projection

layer integrates the extracted features and outputs the final prediction results of the model. This method significantly overcomes the limitations of traditional Transformers in multivariate time-series forecasting and has proven to achieve significant breakthroughs in training on six benchmark datasets [69].

3.3.3.1 iTransformer

The mainstream approach of traditional Transformer models is to represent all variables at the same time step as "Tokens," resulting in a time-point-based word sequence. The model mainly focuses on dependencies between different time steps, often overlooking interactions between variables. However, in multivariate time-series modeling, such as SOH estimation in this study, there are typically complex interrelations between features. For example, a certain linear relationship exists between cumulative mileage and internal resistance, and a representation based solely on time steps is insufficient for fully modeling these dependencies.

In iTransformer, as shown in Fig. **7**(b), the inverted embedding treats the entire time series of each variable as an independent input unit, meaning each variable forms its independent embedding token, thereby enabling "variable-centered" modeling. For a given time length T and number of variables N in the multivariate time-series $X \in \mathbb{R}^{T \times N}$, $X_{t,:}$ represents all features at the same time step, and $X_{:,n}$ represents the full sequence of a single variable. Since $X_{:,n}$ is the entire sequence of the same variable, it carries more consistent semantics and measurement units than $X_{t,:}$. As shown in Equation (7), the inverted embedding layer maps $X_{:,n}$ to a high-dimensional space using a multi-layer perceptron (MLP), resulting in an embedding feature matrix $H = \{h_1, \cdots, h_N\} \in \mathbb{R}^{N \times D}$ that contains N Variate Tokens, where each $h_i \in \mathbb{R}^D$ contains all the time-series variations of the corresponding variable in the past time steps, referred to as a Variate Token. In the subsequent layers, the self-attention mechanism facilitates interaction between each Variate Token, while layer normalization standardizes the measurement units and feature distributions of different variables. The feed-forward network performs fully connected feature encoding. Finally, the projection layer outputs the estimation results for each Variate Token. The entire computational process is represented by Equations (7) to (9).

$$h_n^0 = \text{Embedding}\left(X_{:,n}\right) \tag{7}$$

$$H^{l+1} = \text{TrmBlock}\left(H^l\right), l = 0,\ldots,L-1 \tag{8}$$

$$Y_{:,n} = \text{Projection}\left(h_n^L\right) \tag{9}$$

where, $Y_{:,n}$ represents the final prediction result of $X_{:,n}$. Both the embedding layer and the projection layer are implemented using an MLP. The TrmBlock includes the self-attention mechanism, layer normalization, and feed-forward neural network.

3.3.3.2 Multi-head differential attention

The attention mechanism is designed to mimic how the human brain selectively focuses on different stimuli based on their relevance. This concept has been extensively applied in natural language processing. In its conventional form, attention computes the dot product between the query (Q), key (K), and value (V), followed by *softmax* weighting to perform global feature integration [68]. However, traditional attention mechanisms often face challenges in accurately retrieving key information from the context, especially when handling complex or noisy data, as irrelevant information can interfere with the retrieval process. To address this issue, the differential attention mechanism is introduced. The core idea is to effectively eliminate noise interference by computing the difference between two independent attention spectra. This design is inspired by the differential amplifier principle in electrical engineering, which removes common-mode noise by comparing the differences between two signals [70]. In this way, the differential attention mechanism can suppress irrelevant noise from different sources while preserving important information, thereby improving the ability of the model to extract key information. Similar to traditional attention mechanisms, the Q, K, and V vectors in differential attention are obtained by performing linear transformations on the input sequence X. Attention scores are then computed by comparing the query and key vectors, followed by a weighted sum of the value vectors. The key innovation lies in utilizing a pair of softmax functions to filter out noise from the attention scores. The specific formulas are shown in Equations (10) and (11).

$$\left[Q_1; Q_2\right] = XW^Q, \quad \left[K_1; K_2\right] = XW^K, \quad V = XW^V \tag{10}$$

$$\text{DiffAtnn}(X) = \left(\text{softmax}\left(\frac{Q_1 K_1^T}{\sqrt{d}}\right) - \lambda\,\text{softmax}\left(\frac{Q_2 K_2^T}{\sqrt{d}}\right)\right)V \tag{11}$$

where $X \in \mathbb{R}^{N \times d_{\text{model}}}$ is the input sequence, N is the length of the input sequence, d_{model} is the input feature dimension, and $W^Q, W^K, W^V \in \mathbb{R}^{d_{\text{model}} \times 2d}$ are the projection matrix. The query, key, and value vectors are

represented by $Q_1, Q_2, K_1, K_2 \in \mathbb{R}^{N \times d}, V \in \mathbb{R}^{N \times 2d}$, respectively. A learnable scaling factor λ is used to control the weighted difference between the two attention spectra, as shown in Equation (12)

$$\lambda = \exp\left(\lambda_{q_1} \cdot \lambda_{k_1}\right) - \exp\left(\lambda_{q_2} \cdot \lambda_{k_2}\right) + \lambda_{\text{init}} \tag{12}$$

where, $\lambda_{q_1}, \lambda_{k_1}, \lambda_{q_2}, \lambda_{k_2} \in \mathbb{R}^d$ are learnable vectors used to adjust the attention weights, and $\lambda_{\text{init}} \in (0,1)$ is a constant initialization parameter, typically set to $\lambda_{\text{init}} = 0.8 - 0.6 \times \exp(-0.3 \cdot (l-1))$, where $l \in [1, L]$ represents the layer index, and L is the total number of layers [70].

3.3.4 Model loss function and normalization

The performance of the hybrid model is enhanced by utilizing the mean squared error (MSE) loss function. This function calculates the average squared deviation between the estimated and actual SOH values. In this study, min-max normalization is applied to preprocess all input features before training. The formulas for MSE and min-max normalization are provided in Supplementary Notes 7 and 8, respectively.

4. Results and discussion

In this study, to validate the proposed TCN-Specialized iTransformer model based on multi-dimensional health features, practical generalization tests and model robustness under data scarcity conditions are conducted. Finally, SHAP analysis is performed on the multi-dimensional feature system of the model using the trained model. Notably, most previous real-vehicle SOH models were typically validated within the same battery material or vehicle type. However, this study encompasses a wide range of vehicle types and battery materials, thereby broadening the range of model application and enhancing generalization capability.

4.1 Model hyperparameters

The effectiveness of deep learning models is closely related to the selection of hyperparameters, which influence both the training dynamics and architectural configuration [46]. In the case of the TCN module, the dilation rate determines the receptive field size of the convolution kernel, which in turn impacts the capacity of the model to handle long time-span data. Appropriately setting the dilation rate enables the model to learn long-range temporal dependencies without increasing the depth of the network. This is especially important for tasks like SOH estimation, particularly when handling long time-series data. A larger dilation rate enables the model to obtain more extensive historical information, thereby providing a more accurate

reflection of changes in battery health status over time. Additionally, the length of the sliding window determines the time span that the model can capture. A longer window helps to identify long-term dependencies but may increase computational load, while a shorter window focuses more on local features and is computationally more efficient. In this work, several critical hyperparameters were optimized through a grid search strategy. For example, as shown in Fig. 8, the *x*-axis represents batch size, increasing by a step size of 16, and the *y*-axis represents the sliding window size, increasing by a step size of 5. For the purpose of hyperparameter tuning, vehicle # 4 was randomly selected, and the feature and label data were split into training and test sets in an 8:2 ratio, with the first 80% used for training and the remaining 20% for testing. After training the model, the relative error of the test set is statistically analyzed. The figure shows that the relative error does not follow a simple linear relationship with batch size and window size. For example, when the batch size is 16, the average relative error of the test set first decreases, then increases, and decreases again as the window size increases. After analysis, this study selected a batch size of 32 and a window size of 30. The main hyperparameters of the model are shown in Table 3.
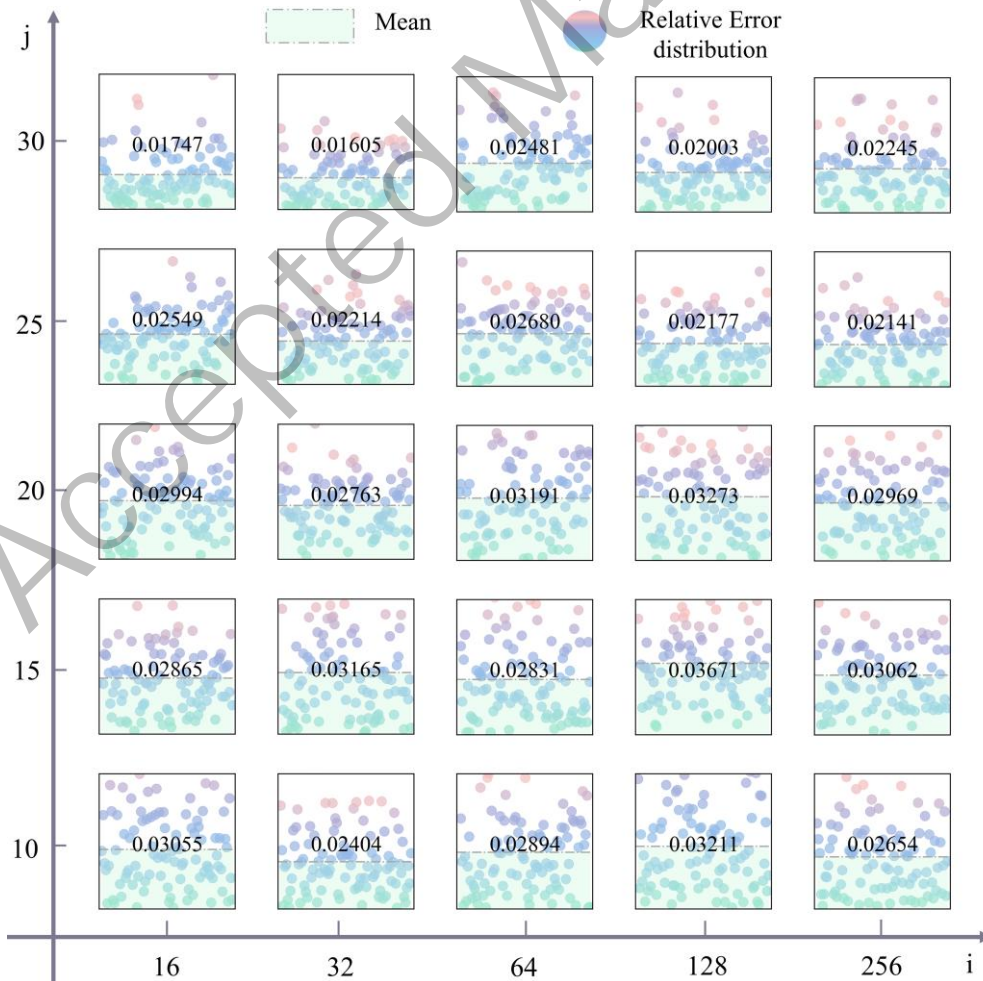
Fig. 8. Relative error distribution and mean value of the test set under different batch sizes and window sizes. i represents the batch size, and j represents the window size. Taking vehicle #4 as an example, its data is split in an 8:2 ratio, and training is performed with different batch sizes and window sizes.

Table 3. Model hyperparameters.

Hyperparameters	Value	Hyperparameters	Value
Seed	42	TCN kernel size	3
Window size	30	TCN stride	1
Model dimension	64	TCN dilation rate (Layer 1, 2, 3)	1, 2, 4
Window size	1	TCN dropout rate	0.2
Batch size	32	Causal padding for TCN (Layer 1, 2, 3)	1, 2, 4
Optimizer	Adam	Number of TCN layers	2
Number of epochs	400	Number of encoder layers	2
The number of attention heads	4	Initial value of the differential attention mechanism scaling factor	0.8
Encoder dropout rate	0.1	Embedding mapping dimension	64
TCN dropout rate	0.2	TCN activation function	ReLU
Number of encoder layers	2	Encoder activation function	GELU
Feedforward network dimension	32	Learning rate	0.001

4.2 Model generalizability validation

In this study, during the model generalization test, 20%–30% randomly selected vehicles are used as the test set to demonstrate the accuracy of the model [44]. The results are compared with three other mainstream models: Transformer, CNN, and LSTM. To ensure reproducibility and fairness, the training processes for all four models are set with the same learning rate strategy. Fig. 9(a, b) illustrate the differences between estimated SOH and actual values across various test datasets. In Fig. 9(a), the test set includes two passenger vehicles, #3 and #5. While in Fig. 9(b), the test set is expanded to include the commercial vehicle #10 in addition to #3 and #5. These two test sets cover different vehicle types and battery materials, but the proposed SOH estimation method achieved the best estimation accuracy on both test sets.

To further quantify and evaluate the performance, this study compared the four methods across five key dimensions: R^2, MAE, MAPE, RMSE, and maximum error (ME). The formulas for these metrics are provided in Supplementary Note 7. Fig. 9(c, d) show that the method proposed in this study achieves a maximum error of less than 2% in both the first and second test sets. Additionally, the proposed method outperforms the other three methods in the remaining four evaluation metrics. In contrast, the other three methods exhibit varying strengths and weaknesses across different dimensions; LSTM demonstrates relatively better performance in terms of average R^2, ME, and RMSE across the two test sets. Compared to LSTM, the proposed method achieves substantial improvements, including reductions of 83.8% in MAE, 83.7% in MAPE, 81.4% in RMSE, and 47.6% in ME, along with a 30.6% increase in R^2. Overall, the

proposed method demonstrates the ability to achieve accurate estimates across different test sets. The additional cross-vehicle validation is discussed in Supplementary Note 9.
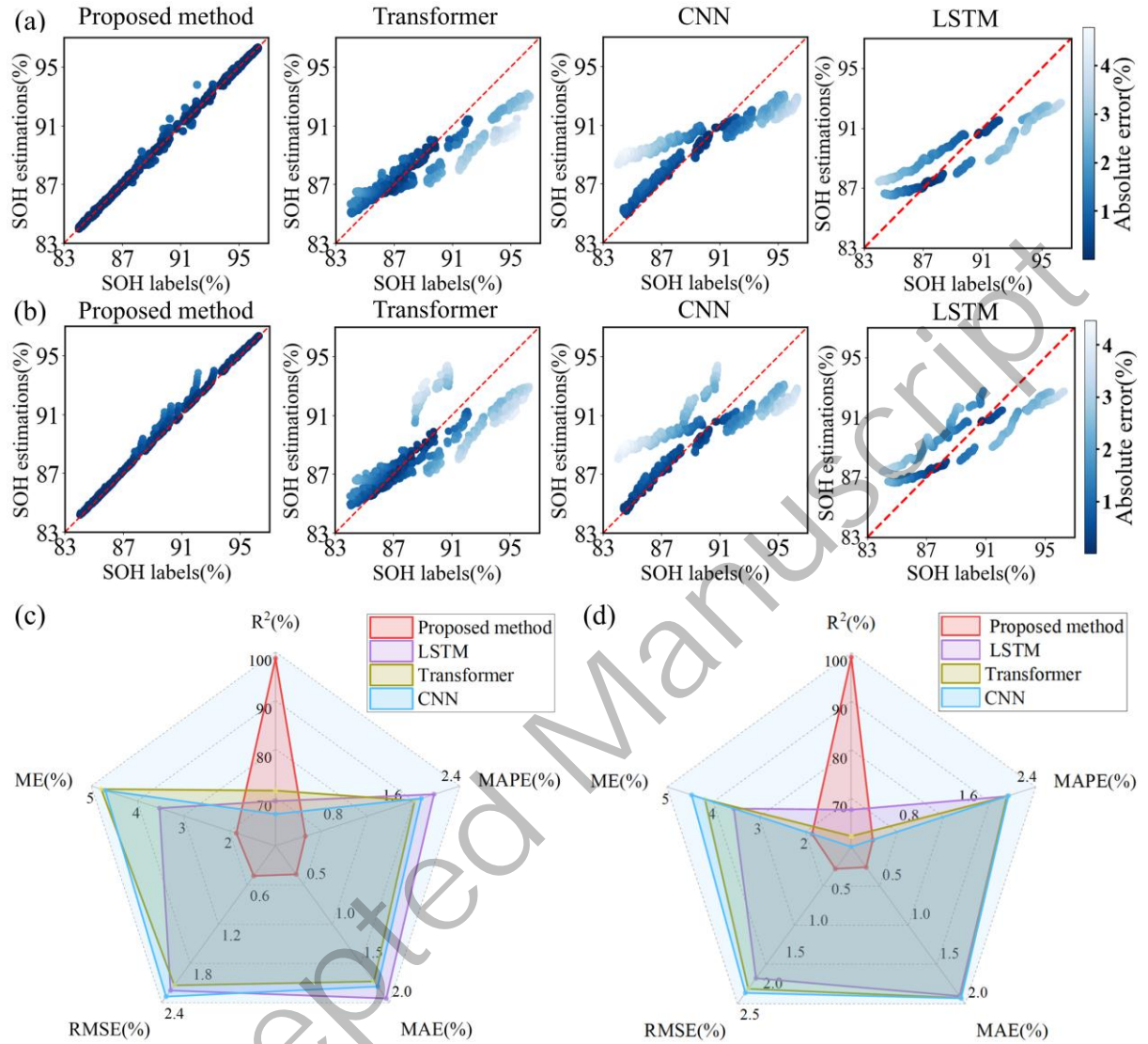
Fig. 9. SOH estimation results compared with different test sets and algorithms. (a) SOH estimation results on the test set consisting of vehicles #3 and #5, with the remaining vehicles used for training, under different algorithms. (b) SOH estimation results on the test set consisting of vehicles #3, #5, and #10, with the remaining vehicles used for training, under different algorithms. (c) Comparison of estimation error metrics for various algorithms from panel (a). (d) Comparison of estimation error metrics for various algorithms from panel (b).

4.3 Model robustness under data scarcity conditions

To evaluate the data efficiency and practical applicability of the proposed framework in large-scale realistic EV fleet deployment scenarios, this section examines the performance of the model under data scarcity conditions. In practical applications, particularly on cloud-based platforms that manage large EV fleets, it is often impractical to collect extensive historical data for each newly connected vehicle. Therefore,

an SOH estimation model must be capable of achieving reliable fitting performance with limited training data, thereby reducing data acquisition efforts, training time, and computational costs.

In this study, vehicles #1, #5, and #8 are randomly selected as case studies. For each vehicle, two training configurations are evaluated, using 80% and 30% of the available data, respectively. The remaining data serve as the test set to assess model performance. As illustrated in Fig. 10, the proposed method consistently achieves superior results across all test cases under both training conditions, outperforming baseline models such as Transformer, CNN, and LSTM. Even when trained with only 30% of the data, the proposed model maintains stable fitting accuracy across different vehicle types and battery chemistries. As shown in Table 4, the proposed method achieves the lowest RMSE across all configurations, with a maximum RMSE below 0.16%. In contrast, some baseline models may occasionally achieve lower errors on certain individual test cases, their performance generally fluctuates significantly across different vehicles and training conditions, indicating weaker robustness and generalization. In contrast, the proposed method delivers consistent and reliable performance across diverse practical operating scenarios. In summary, the proposed method establishes stable SOH estimation mappings under data scarcity conditions. It provides a practical and scalable solution for intelligent battery health management in on-road EV deployment.

Table 4. The RMSE of the test sets under different training data proportions and across different vehicles.

Training set ratio	Vehicle number	Proposed method	CNN	Transformer	LSTM
80%	#1	0.08‰	0.65‰	0.59‰	0.84‰
	#5	0.42‰	4.10‰	2.99‰	2.57‰
	#8	1.06‰	2.86‰	3.79‰	4.15‰
30%	#1	0.51‰	3.83‰	2.66‰	3.88‰
	#5	1.43‰	9.00‰	12.61‰	12.19‰
	#8	1.55‰	5.61‰	7.87‰	4.58‰

4.4 Model interpretability based on SHAP

Due to the intricate internal mechanisms and high nonlinearity of deep learning models, they are often regarded as "black boxes." Although such models are capable of autonomously extracting meaningful features from large-scale data, their internal structures and the knowledge they acquire are often difficult for humans to comprehend. This complexity reduces model transparency and poses challenges in interpreting the reasoning behind estimation results, thereby affecting the reliability and interpretability of the model outputs. To address this issue, this study introduces SHAP-based analysis to explore how various features influence the results of the model, with a particular focus on the proposed shallow, intermediate, and deep three-dimensional feature structures. SHAP, rooted in Shapley values from cooperative game theory,

quantifies the contribution of each input feature by computing its marginal effect on the prediction through

an aggregation over all possible feature subsets [53]. The mathematical formulation is
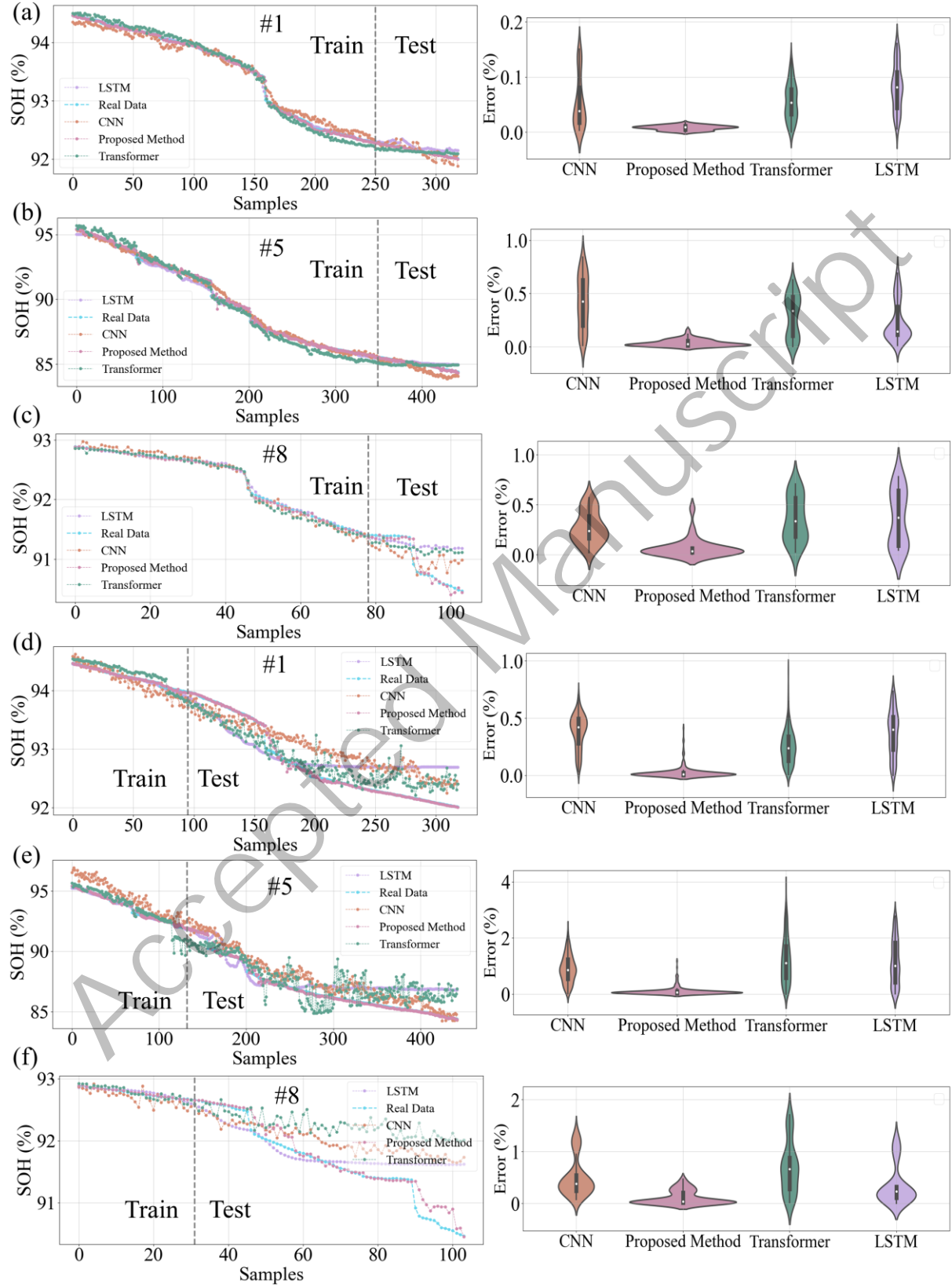
Fig. 10. Comparison of different algorithms under data scarcity conditions. (a–c) Comparison of models on vehicles #1, #5, and #8 with 80% training and 20% testing data. The right side shows violin plots of relative estimation errors for different algorithms. (d–f) Comparison of models on vehicles #1, #5, and #8 with 30% training and 70% testing data, with corresponding violin plots of relative estimation errors.

$$\phi_r = \sum_{S \subseteq Z \setminus \{r\}} \frac{|S|!(|Z|-|S|-1)!}{|Z|!}(f(S \cup \{r\}) - f(S)) \tag{13}$$

where, Z is the complete feature set in the study, S represents the subset of features excluding feature r, f is the model function. The expression $(f(S \cup \{r\}) - f(S))$ reflects the impact of adding feature r to subset S.

This study conducts a global explanation of the importance of the three-dimensional feature system using SHAP analysis. Global explanation refers to an analysis of the entire model, aiming to help us understand the importance of each feature and how they influence the estimation results of the model. As shown in Fig. 11, the global significance of each feature in the proposed model is displayed, and the average contribution of each feature is described. Among the 11 features, the contributions of the first seven features reached 99%. Specifically, the first-dimensional feature contributed 20.6%, the second-dimensional feature contributed 43.5%, and the third-dimensional feature contributed 35.9%. The dimensional contribution analysis shows that although the second-dimensional feature contains only two features, its contribution exceeds that of the other two dimensions.
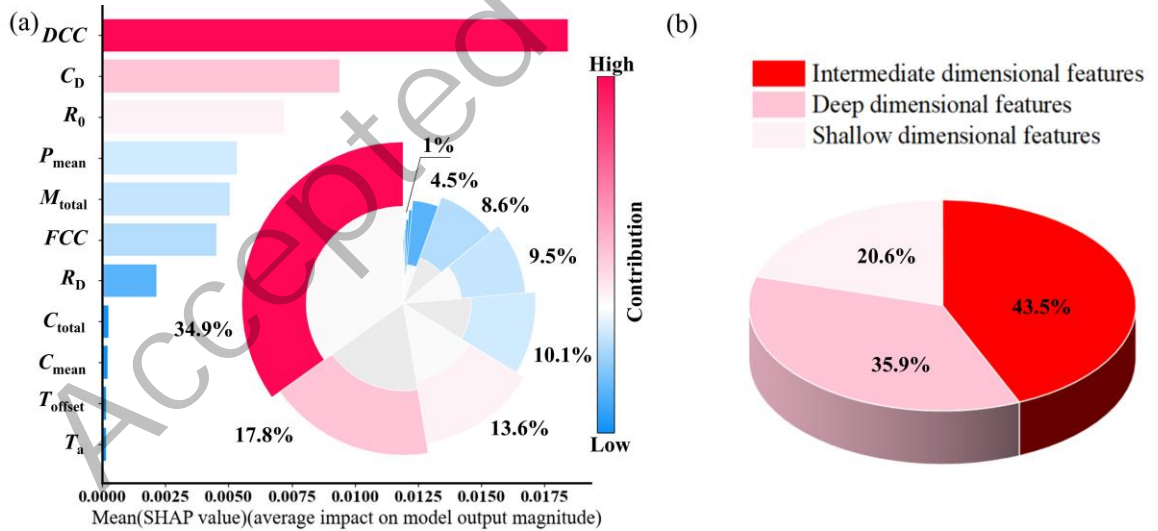
Fig. 11. Global feature contribution analysis of the proposed model based on SHAP. (a) Average SHAP value analysis and heatmap display of feature contributions. (b) Total contribution proportion of different dimension features.

Based on the previous analysis, the second-dimensional features primarily stem from the driving behaviors, which indicates that, in real-vehicle SOH estimation, the heterogeneity of drivers significantly impacts SOH degradation. Therefore, feature extraction from the perspective of the user is particularly

important in on-road SOH estimation modeling. Additionally, although the third-dimensional feature includes only three features, its contribution remains as high as 35.9%. Notably, the average power, which shows a low Pearson correlation in Fig. 6, contributes as much as 10.1%. This suggests that a low linear correlation between a feature and the label does not imply that the feature has no contribution to the model. Furthermore, for features such as cumulative charge-discharge capacity, which exhibit high linear correlation with SOH, the corresponding global contribution appears relatively lower. This may be attributed to strong collinearity between cumulative charge-discharge capacity and cumulative mileage. As these two features contain partially redundant information, the model integrates such redundancy during training, which dilutes the marginal contribution of any single feature in the global attribution analysis [72]. In summary, features from each dimension make significant contributions to the estimation results of the model. The absence of any dimension would compromise the estimation result, further validating the importance of the three-dimensional feature system proposed in this study.

4.5 Comparison of model performance after feature optimization

As indicated by the analysis above, there are four features with a contribution of only 1%. Thus, this study excluded them from the dataset and retrained the proposed model using the first dataset from Section 4.2, followed by a comparison, shown in Table 5. The performance of the model before and after removing the features was evaluated using R^2, RMSE, MAE and MAPE. After removing these four features, the most significant improvement was in training efficiency, with a 17.3% reduction in training time. Additionally, R^2 increased, while RMSE, MAE and MAPE decreased. This improvement in performance is primarily due to the elimination of irrelevant features, which reduces the susceptibility of the model to noise and enables it to focus more on informative variables. Additionally, Supplementary Note 10 provides a comparison of different algorithms, including CNN, LSTM, and Transformer, after feature optimization.

Table 5. Performance metrics before and after feature optimization based on SHAP importance.

Evaluation metric	Before optimization	After optimization	Improvement
Training time (s)	120.75	99.79	17.3%
R^2 (%)	98.67	98.71	0.04%
RMSE (%)	0.4598	0.4521	1.67%
MAE (%)	0.3633	0.3483	4.13%
MAPE (%)	0.4056	0.3887	4.17%

By utilizing SHAP-based interpretation, the model can identify and remove features with limited contribution, thereby improving accuracy. It is essential to note that, in certain specific cases, even these weak features may have contributed to the estimation result of the model, potentially reducing the estimation

31

error. However, these features could negatively impact the overall performance of the model. Based on the above analysis and optimization, the final three-dimensional feature system established in this study is presented in Table 6.

Table 2. Health features after optimization.

Feature dimension	Concrete features description	Abbreviation
Shallow dimensional features	Accumulated mileages	M_{total}
	Average charging power	P_{mean}
Intermediate dimensional features	Fast charge count	FCC
	Deep charge-discharge count	DCC
Deep dimensional features	Ohmic resistance	R_0
	Polarization resistance	R_D
	Polarization capacity	C_D

4.6 Discussion

In practical EV deployments, SOH estimation faces ongoing challenges due to inconsistencies in driver operation, variations in environmental conditions, and the diversity of battery chemistries within vehicle fleets. Unlike laboratory datasets, operational data collected during on-road use are often affected by noise, making it difficult to extract degradation signals reliably. The framework proposed in this study is designed to address these challenges. Its hybrid architecture enables the model to process incomplete temporal sequences and extract meaningful degradation features from data that deviate from idealized charge-discharge patterns. Moreover, the three-dimensional health feature system, developed for vehicle-based applications, integrates physical, behavioral, and electrochemical information, thereby enhancing the model capability to capture degradation heterogeneity. Importantly, the framework is designed for cloud-based deployment, allowing raw vehicle data to be periodically transmitted, preprocessed into standardized feature representations, and used to produce SOH estimates across all connected vehicles through centralized model training. This architecture supports unified training and analysis, reduces the computational burden on onboard hardware, and enables real-time monitoring across vehicle fleets.

Furthermore, reliable SOH estimation is of critical importance for the life cycle management of EV batteries. In the retirement phase, accurate SOH values can be used to identify performance thresholds and inform rational end-of-life strategies. In second-life applications, the safety and durability of the battery system depend on a balanced configuration of cells. SOH-based reuse strategies can optimize cell allocation and reduce the risk of failure due to mismatch. Similarly, in the recycling phase, state classification based on historical health data can improve material recovery efficiency and reduce processing costs. The

compatibility of the proposed framework with cloud deployment further enables integration into residual value assessment workflows for used EVs, where battery condition plays a key role in determining market value. Despite these advantages, future work must address challenges arising from domain shift and varying operational conditions, potentially through online adaptation or continual learning. Such advancements are essential to ensuring that SOH estimation remains reliable not only during active service but also throughout the reuse, resale, and retirement stages, thereby enabling intelligent life-cycle management of EV battery systems.

5. Conclusions

Reliable estimation of battery capacity degradation is essential for enabling timely maintenance, accurate residual value assessment, and effective second-life applications in EVs. To address this challenge, this study proposes a hybrid deep learning framework—TCN-Specialized iTransformer—built upon a three-layer dimensional feature system using real-vehicle data. To enhance the generalizability and practical applicability of the model, the dataset includes operational data from typical commercial and passenger EVs and covers two mainstream battery chemistries: NCM and LFP. This ensures broad representativeness in terms of battery types and usage scenarios. Based on this dataset, a three-dimensional feature system—comprising shallow, intermediate, and deep features—is constructed from raw vehicle sensor data, driver behavior patterns, and internal electrochemical parameters. To handle the long time spans and high volume of multivariate time-series data inherent in on-road applications, the TCN enables efficient parallel processing and captures temporal dependencies effectively. Additionally, to address the limitations of traditional Transformers in multivariate time-series modeling, the study introduces a Specialized iTransformer module, featuring a differential attention mechanism that mitigates noise in standard attention layers and improves robustness.

Both generalization and data scarcity robustness tests were conducted to evaluate the performance of the proposed SOH estimation method. In the generalization test, the model was validated across different vehicle types and battery chemistries using two distinct test sets. The results demonstrated strong estimation performance, with average values of R^2 = 98.88%, MAE = 0.31%, MAPE = 0.35%, RMSE = 0.40%, and ME = 1.85%. To enable a comprehensive performance assessment, the proposed method was benchmarked

against several widely adopted deep learning models, including Transformer, CNN, and LSTMs. Compared to the best-performing baseline, the proposed method achieves relative reductions of 83.8% in MAE, 83.7% in MAPE, 81.4% in RMSE, and 47.6% in ME, along with a 30.6% increase in R^2. To assess model robustness under data scarcity, the training data was truncated at 80% and 30% of the time series. Even with significantly reduced training data, the model maintained high accuracy, with RMSE values not exceeding 0.16%, again surpassing baseline methods. Model interpretability and feature relevance were further evaluated through SHAP analysis. Results revealed that all three feature dimensions contributed meaningfully to the estimation results of the model. Notably, two features related to driver behavior—part of the intermediate feature dimension—accounted for as much as 43.5% of the total contribution, highlighting the significant influence of driving behavior on battery degradation. Based on this insight, the feature set was optimized by removing four less impactful features, resulting in a 17.3% reduction in training time and improvements of over 4% in both MAE and MAPE. In conclusion, the proposed approach enables accurate, interpretable SOH estimation in practical vehicle settings. It supports proactive maintenance and life cycle management of EV battery systems and provides a scalable foundation for future deployment in cloud-based battery health monitoring platforms.

In future work, preliminary discussions have been conducted with Guangzhou Automobile Group Co., Ltd. regarding data sharing and model application. The proposed method is expected to be extended to a large-scale dataset encompassing millions of vehicles from the company, in order to further validate its generalization capability through cloud-based deployment. By integrating multimodal data from diverse onboard sensors, the model is expected to achieve greater adaptability and robustness across a wider range of on-road scenarios. Additionally, the model will be deployed on mainstream EV cloud platforms, with periodic online updates and retraining mechanisms to enable real-time battery monitoring and proactive maintenance. This integration will facilitate comprehensive lifecycle management and enhance the operational performance of EVs. Ultimately, the proposed framework is anticipated to accelerate the intelligent evolution of battery management systems and support the sustainable development of the EV industry.

Acknowledgment

Conflicts of interest

The authors declare no competing interests.

Reference

[1] H. Liu, C. Li, X. Hu, J. Li, K. Zhang, Y. Xie, Z. Song, Nat. Commun. 16 (2025) 1137.

[2] J. Zhao, Z. Lv, D. Li, X. Feng, Z. Wang, Y. Wu, A.F. Burke, eTransportation 22 (2024) 100364.

[3] State Council Information Office of the People's Republic of China, China's New Energy Vehicle Production Hits 10 Million Units Annually.

[4] Z. Deng, L. Xu, H. Liu, X. Hu, B. Wang, J. Zhou, J. Energy Chem. 89 (2024) 345–354.

[5] Y. Wang, X. Feng, Y. Peng, F. Zhang, D. Ren, X. Liu, M. Ouyang, Joule 6 (2022) 2810–2820.

[6] X. Hu, L. Xu, X. Lin, M. Pecht, Joule 4 (2020) 310–346.

[7] F. Wang, Z. Zhai, Z. Zhao, Y. Di, X. Chen, Nat. Commun. 15 (2024) 4332.

[8] Z. Wang, D. Shi, J. Zhao, Z. Chu, D. Guo, C. Eze, A.F. Burke, eTransportation 19 (2024) 100309.

[9] D.P. Finegan, J. Zhu, X. Feng, M. Keyser, M. Ulmefors, W. Li, S.J. Cooper, Joule 5 (2021) 316–329.

[10] J. Lu, R. Xiong, J. Tian, C. Wang, F. Sun, Nat. Commun. 14 (2023) 2760.

[11] Y. Yang, J. Yang, X. Wu, L. Fu, X. Gao, X. Xie, Q. Ouyang, J. Energy Chem. 106 (2025) 553–564.

[12] C. Liang, S. Tao, X. Huang, Y. Wang, B. Xia, X. Zhang, J. Energy Chem. 106 (2025) 205–219.

[13] S. Khaleghi, Y. Firouz, J. Van Mierlo, P. Van Den Bossche, Appl. Energy 255 (2019) 113813.

[14] J. Zhao, H. Ling, J. Liu, J. Wang, A.F. Burke, Y. Lian, eTransportation 15 (2023) 100214.

[15] B. Jiang, J. Zhu, X. Wang, X. Wei, W. Shang, H. Dai, Appl. Energy 322 (2022) 119502.

[16] K.S. Ng, C.S. Moo, Y.P. Chen, Y.C. Hsieh, Appl. Energy 86 (2009) 1506–1511.

[17] Y. He, Z. Deng, J. Chen, W. Li, J. Zhou, F. Xiang, X. Hu, J. Energy Chem. 98 (2024) 1–11.

[18] L. Vichard, A. Ravey, P. Venet, F. Harel, S. Pelissier, D. Hissel, Energy 225 (2021) 120235.

[19] L. Chen, Z. Lü, W. Lin, J. Li, H. Pan, Measurement 116 (2018) 586–595.

[20] X. Qu, D. Shi, J. Zhao, M.K. Tran, Z. Wang, M. Fowler, A.F. Burke, J. Energy Chem(2024).

[21] J. Li, K. Adewuyi, N. Lotfi, R.G. Landers, J. Park, Appl. Energy 212 (2018) 1178–1190.

[22] C.P. Lin, J. Cabrera, F. Yang, M.H. Ling, K.L. Tsui, S.J. Bae, Appl. Energy 275 (2020) 115338.

[23] X. Shu, G. Li, J. Shen, Z. Lei, Z. Chen, Y. Liu, Energy 207 (2020) 118262.

[24] H. Xu, L. Wu, S. Xiong, W. Li, A. Garg, L. Gao, Energy 276 (2023) 127585.

[25] N. Yang, Z. Song, H. Hofmann, J. Sun, J. Energy Storage 48 (2022) 103857.

[26] J. Zhao, X. Feng, Q. Pang, J. Wang, Y. Lian, M. Ouyang, A.F. Burke, J. Power Sources 581 (2023) 233474.

[27] C. Lin, L. Wu, X. Tuo, C. Liu, W. Zhang, Z. Huang, G. Zhang, J. Energy Chem. 105 (2025) 261–279.

[28] B. Chen, Y. Liu, B. Xiao, J. Energy Storage 98 (2024) 113074.

[29] S. Tao, R. Ma, Y. Chen, Z. Liang, H. Ji, Z. Han, G. Zhou, J. Power Sources 597 (2024) 234156.

[30] S. Liu, Y. Nie, A. Tang, J. Li, Q. Yu, C. Wang, eTransportation 18 (2023) 100296.

[31] K.S. Mawonou, A. Eddahech, D. Dumur, D. Beauvois, E. Godoy, J. Power Sources 484 (2021) 229154.

[32] R. Pan, T. Liu, W. Huang, Y. Wang, D. Yang, J. Chen, Energy 285 (2023) 129460.

[33] N. Cai, X. Que, X. Zhang, W. Feng, Y. Zhou, Energy 302 (2024) 131700.

[34] Y. Lu, D. Guo, G. Xiong, Y. Wei, J. Zhang, Y. Wang, M. Ouyang, eTransportation 22 (2024) 100361.

[35] G. Lee, D. Kwon, C. Lee, Mech. Syst. Signal Process. 188 (2023) 110004.

[36] M. Lin, D. Wu, S. Chen, J. Meng, W. Wang, J. Wu, IEEE Trans. Instrum. Meas. 73 (2024) 1–9.

[37] L. Cai, J. Yan, H. Jin, J. Meng, J. Peng, B. Wang, R. Teodorescu, J. Energy Chem. 103 (2025) 759–772.

[38] L. Chen, S. Xie, A.M. Lopes, H. Li, X. Bao, C. Zhang, P. Li, Energy 286 (2024) 129597.

[39] J. Zhao, X. Han, Y. Wu, Z. Wang, A.F. Burke, J. Energy Chem 102 (2025) 463-496.

[40] R. Xiong, Y. He, Y. Sun, Y. Jia, W. Shen, J. Energy Chem 104 (2025) 618-627.

[41] Y. Che, Y. Zheng, X. Sui, R. Teodorescu, J. Energy Chem. 84 (2023) 335–346.

[42] S. Tao, M. Zhang, Z. Zhao, H. Li, R. Ma, Y. Che, G. Zhou, Energy Environ. Sci. 18 (2025) 1544–1559.

[43] H. Li, X. Xie, X. Zhang, A.F. Burke, J. Zhao, J. Energy Storage 115 (2025) 116000.

[44] J. Hong, F. Liang, H. Yang, C. Zhang, X. Zhang, H. Zhang, J. Yang, eTransportation 20 (2024) 100322.

[45] J. Zhao, Z. Wang, Y. Wu, A.F. Burke, Appl. Energy 377 (2025) 124746.

[46] J. Zhao, Z. Wang, Energy Storage Mater. 71 (2024) 103668.

[47] Z. Lv, J. Zhao, eTransportation 23 (2025) 100383.

[48] X. Huang, S. Tao, C. Liang, R. Ma, X. Wang, B. Xia, X. Zhang, J. Power Sources 642 (2025) 236939.

[49] S. Tao, R. Guo, J. Lee, S. Moura, L.C. Casals, S. Jiang, X. Zhang, Energy Environ. Sci. (2025). https://doi.org/10.1039/D5EE02217G

[50] G. Pozzato, A. Allam, L. Pulvirenti, G.A. Negoita, W.A. Paxton, S. Onori, Joule 7 (2023) 2035–2053.

[51] J. Hong, K. Li, F. Liang, H. Yang, C. Zhang, Q. Yang, J. Wang, Energy 289 (2024) 129918.

[52] Q. Wang, M. Ye, S. Celik, Z. Deng, B. Li, D.U. Sauer, W. Li, J. Energy Chem. 99 (2024) 681–691.

[53] S. Wen, N. Lin, S. Huang, X. Li, Z. Wang, Z. Zhang, J. Energy Storage 96 (2024) 112623.

[54] R. Sun, J. Chen, C. Piao, J. Power Sources 645 (2025) 236784.

[55] X. Li, P. Wang, J. Wang, F. Xiu, Y. Xia, J. Energy Storage 72 (2023) 108247.

[56] C. Zhang, J. Wang, L. Zhang, W. Zhang, T. Zhu, X.G. Yang, A. Cruden, Energy Storage Materials 78 (2025) 104236.

[57] X. Zhao, J. Hu, G. Hu, H. Qiu, J. Energy Storage 63 (2023) 107031.

[58] Z. Deng, L. Xu, H. Liu, X. Hu, Z. Duan, Y. Xu, Appl. Energy 339 (2023) 120954.

[59] Visual Crossing Weather, Historical weather data API.https://www.visualcrossing.com/weather-data(accessed 22 December 2024).

[60] F. Xia, Y. Yu, J. Chen, J. Energy Storage 102 (2024) 114134.

[61] Z. He, X. Shen, Y. Sun, J. Energy Storage 41 (2021) 102867.

[62] S. Ou, J. Power Sources 573 (2023) 233133.

[63] J. Zhang, Y. Wang, B. Jiang, H. He, S. Huang, C. Wang, M. Ouyang, Nat. Commun. 14 (2023) 5940.

[64] S. Bashash, S.J. Moura, J.C. Forman, H.K. Fathy, J. Power Sources 196 (2011) 541–549.

[65] S. Sagaria, M. van der Kam, T. Boström, Appl. Energy 377 (2025) 124546.

[66] D. Zhen, J. Liu, S. Ma, J. Zhu, J. Kong, Y. Gao, F. Gu, Green Energy Intelligent Transp. 3 (2024) 100207.

[67] W. Liu, Y. Che, J. Han, Z. Deng, X. Hu, Z. Song, J. Power Sources 615 (2024) 235114.

[68] X. Yang, J. Hu, G. Hu, X. Guo, J. Energy Storage 55 (2022) 105820.

[69] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, M. Long, arXiv (2023) arXiv:2310.06625.

[70] T. Ye, L. Dong, Y. Xia, Y. Sun, Y. Zhu, G. Huang, F. Wei, arXiv (2024) arXiv:2410.05258 .

[71] P. Westermann, M. Welzel, R. Evins, Appl. Energy 278 (2020) 115563.

[72] J. Ittner, L. Bolikowski, K. Hemker, R. Kennedy, arXiv (2021) arXiv:2107.12436.