Journal of Energy Storage 153 (2026) 120956

Contents lists available at ScienceDirect

Journal of Energy Storage

journal homepage: www.elsevier.com/locate/est

Research papers

Battery state-of-health estimation via thermoelectric graph learning and interpretable temporal decomposition for real-world electric vehicles

Zhilong Lv [a], Shiqi (Shawn) Ou [a,b,*], Hao Jing [a,b], Haobo Dong [a], Dapai Shi [c], Jingyuan Zhao [d]

[a] School of Future Technology, South China University of Technology, Guangzhou, Guangdong, 511442, China
[b] Guangdong Artificial Intelligence and Digital Economy Laboratory (Guangzhou), Guangzhou, Guangdong, 510335, China
[c] School of Mechanical Engineering, Hubei University of Arts and Science, Xiangyang, 441053, China
[d] Institute of Transportation Studies, University of California Davis, Davis, CA, 95616, USA

ARTICLE INFO

Keywords:
Electric vehicle
Graph attention networks
Interpretable deep learning
Lithium-ion batteries
Field data
State-of-health

ABSTRACT

Accurate and interpretable estimation of state-of-health (SOH) is critical for the safe and efficient operation of electric vehicles (EVs). Existing data-driven methods, however, struggle to capture spatial heterogeneity across cells and the variability introduced by user-specific charging behaviors. These limitations often lead to poor generalizability and limited interpretability. To address these challenges, this study proposes a framework that combines the dynamic graph attention network (DGAT) and neural basis expansion analysis for interpretable time series forecasting (N-BEATS) to achieve spatial modeling and interpretability. The framework is evaluated using real-world on-road operational data from 300 EVs, comprising approximately 850 million data points. The results demonstrate that the proposed method significantly outperforms state-of-the-art baselines, achieving a root mean square error of 1.22% and a mean absolute percentage error (MAPE) of 0.98%. Furthermore, by incorporating an adapter-based transfer learning strategy, the model maintains a low MAPE of 1.43% using only 10% of the target domain data. Beyond estimation accuracy, the proposed relative charging preference (RCP) metric quantifies user behavior, revealing that users with RCP greater than 0.8 exhibit 38.1% higher degradation rates and SOH estimation errors above 20%. These findings highlight the strengths of the proposed DGAT–N-BEATS framework in spatiotemporal modeling, generalizability, and interpretability, demonstrating its substantial potential for battery health management applications.

1. Introduction

Guided by the strategic goals of “carbon peaking” and “carbon neutrality,” the transportation sector is experiencing a fundamental shift from internal combustion to electrification [1,2]. As a critical enabler of transportation electrification, power battery technology serves as a vital bridge between clean energy generation and efficient energy utilization [3]. Lithium-ion batteries (LIBs) have been widely adopted in electric vehicles (EVs), energy storage systems, and renewable energy integration due to their high energy density, long cycle life, and favorable charge–discharge characteristics [4–7]. However, prolonged usage of LIBs inevitably leads to performance degradation [8]. This degradation process is influenced by intrinsic material properties, electrochemical reaction kinetics, and multi-physical coupling effects, resulting in highly complex and nonlinear degradation behavior [9–12]. Battery degradation, driven by electrochemical aging and operating stressors, leads to a gradual decline in capacity and performance. State-of-health (SOH) is a critical metric for assessing battery degradation, typically defined as the ratio of present capacity to its rated initial capacity [13]. Accurate SOH estimation is essential for advanced battery management systems (BMS) to ensure the operational safety of EVs and optimize energy utilization. In real-world conditions, SOH estimation methods must exhibit high accuracy, robustness, and adaptability—requirements that are often not satisfied by models developed under idealized laboratory assumptions. Therefore, developing SOH estimation methods tailored to real-world EV batteries is essential for bridging the gap between theoretical modeling and practical application.

Recent years have witnessed considerable progress in modeling approaches for lithium-ion battery SOH estimation, ranging from physics-based to data-driven techniques. Physical models provide insights into the physical and chemical changes within the battery [14,15]. Conventional machine learning algorithms, such as support vector machines

---

* Corresponding author at: School of Future Technology, South China University of Technology, Guangzhou, Guangdong, 511442, China.
E-mail address: sou@scut.edu.cn (S.(S. Ou).

https://doi.org/10.1016/j.est.2026.120956
Received 10 August 2025; Received in revised form 30 December 2025; Accepted 1 February 2026

and extreme learning machines, have gained attention for their robustness and ability to capture nonlinear patterns in battery data [16,17]. However, physics-based models typically rely on complex mathematical formulations and extensive parameter calibration, making them difficult to adapt to diverse battery chemistries and operational conditions [18,19]. In addition, conventional machine learning methods rely on handcrafted feature engineering and struggle to efficiently handle high-dimensional, complex datasets [20]. These limitations have prompted a shift toward deep learning approaches. Several studies have demonstrated the potential of deep learning to overcome the limitations of conventional battery diagnostics, achieving high accuracy, scalability, and generalization in modeling complex degradation patterns from large-scale, high-dimensional operational data [21–25]. However, most of these approaches treat battery measurements as independent sequences without explicitly modeling the spatial or relational dependencies among cells, which motivates the use of graph-based architectures. For example, a study developed a graph neural network–based framework for data aggregation and feature fusion, significantly improving the accuracy of capacity estimation under incomplete charging conditions [26]. Another work proposed a graph neural network architecture that integrates temporal and spatial health indicators, achieving high-precision SOH prediction across multiple public datasets [27]. However, the non-stationarity, operational variability, and behavioral diversity inherent in real-world conditions pose significant challenges to model performance and generalization beyond laboratory settings.

To bridge the gap between laboratory research and engineering deployment, researchers developed a deep multimodal framework for SOH estimation using three years of operational data from 300 EVs, achieving a mean absolute percentage error (MAPE) of 2.83% [28]. Despite the powerful modeling capabilities of deep learning methods, they often suffer from limited interpretability and a strong reliance on large amounts of labeled data, which are difficult to obtain in practical applications. Transfer learning provides a scalable alternative by facilitating cross-domain knowledge transfer and improving adaptability in data-scarce scenarios [29,30]. A study developed a transfer learning-based framework combining the Transformer and the convolutional neural network (CNN) for SOH estimation, achieving less than 1% error across lithium iron phosphate and nickel cobalt aluminum oxide batteries [31]. Another study achieved an MAPE of 0.23% on data from 707 EVs by integrating condition alignment and deep fusion networks within a transfer learning framework, demonstrating strong cross-domain generalization [32]. Recent studies have proposed a driving-behavior-aware transfer learning framework based on an exponential smoothing Transformer to enhance real-world SOH estimation from discharging profiles, thereby improving EV-level prediction without requiring complete retraining [33]. In addition, by leveraging laboratory data, a study developed a cell-to-pack transfer learning framework that estimates battery pack SOH using single-cell aging knowledge and limited pack data [34]. Hybrid modeling approaches have emerged as a key direction for improving model performance and interpretability [35–39]. For example, a study employing Bayesian optimization and extreme gradient boosting with an attention mechanism achieved SOH estimation errors below 3.2% using real-world vehicle data [40]. Additionally, a model combining natural-gradient boosting and categorical boosting (N-CatBoost), trained on 28 million EV samples, achieved an MAPE of 0.817% for SOH estimation and supported uncertainty quantification [41].

Despite the notable progress in SOH estimation enabled by deep learning and hybrid modeling techniques, several critical challenges remain unresolved in real-world deployment. First, existing models often fail to explicitly capture the spatiotemporal heterogeneity of battery degradation. Second, many frameworks rely on static graph structures that are predefined rather than dynamically constructed in real-time, which limits their adaptability across battery systems with varying pack configurations or sensor layouts. Finally, while high prediction accuracy has been achieved in controlled environments, model robustness and interpretability in large-scale, noisy fleet scenarios remain limited. Furthermore, current transfer learning strategies typically require substantial data from the target domain, hindering rapid adaptation to new battery systems. To address these limitations, this study proposes an interpretable and transferable SOH estimation framework that integrates the dynamic graph attention network (DGAT) with neural basis expansion analysis for interpretable time series forecasting (N-BEATS). Extensive validation on large-scale EV fleet data demonstrates the framework's ability to achieve high accuracy, provide transparent diagnostics, and adapt rapidly, even in highly variable user behavior and sparse data scenarios. The principal contributions of this study are outlined as follows:

(1) An interpretable framework is developed for EV battery health diagnostics: By integrating DGAT with N-BEATS, the proposed model jointly learns spatial thermoelectric dependencies and decomposes temporal degradation trends.
(2) A real-time dynamic graph construction method based on multi-head attention mechanisms is proposed: In contrast to conventional methods relying on offline precomputed adjacency matrices, this approach does not require a predefined battery topology.
(3) The superior performance and interpretability of the proposed model are validated on large-scale real-world fleet data: Operational data from 300 EVs were used for training and evaluation, achieving an MAPE of 0.98%.
(4) An efficient transfer learning scheme based on adapter-based fine-tuning is carefully designed: Adapter modules facilitate rapid adaptation to new battery systems with differing specifications using only 10% of the target data, resulting in root mean square error (RMSE) and MAPE of 2.1% and 1.43%, respectively.
(5) The impact of diverse user behaviors on battery capacity degradation and model performance is analyzed: Users predominantly employing fast charging experience up to 38.1% greater battery capacity degradation, with maximum model prediction errors exceeding 20%.

The rest of the paper is organized as follows: Section 2 introduces the dataset sources, preprocessing methods, and feature construction. Section 3 outlines the proposed framework. Section 4 presents experimental results and discusses the influence of user behavior on degradation and estimation accuracy. Section 5 summarizes the main contributions and findings of this paper and provides an outlook on future research directions.

2. Field datasets

2.1. Data source

This study uses two datasets collected during in-field LIB operation in EVs to develop and validate a data-driven framework for SOH estimation under realistic driving conditions. The two datasets differ in scale, structure, and battery configuration, thereby supporting both large-scale model training and cross-domain generalization analysis. Dataset #1 was obtained from a publicly available large-scale database, comprising operational data from 300 decommissioned EVs equipped with nickel cobalt manganese oxide (NCM) battery packs [28]. These vehicles experienced a wide range of usage scenarios, with total mileage varying from 0 to 500,000 km and service lifespans ranging from 0.5 to 4 years. The dataset was recorded at a 10-s sampling interval, resulting in approximately 850 million data points. Each data point contains key vehicle and battery parameters. These include the timestamp, state of charge (SOC), vehicle speed, cumulative odometer reading, charge/discharge status, pack voltage and current, minimum and maximum cell voltages, sensor temperatures, and full lists of individual cell voltages and probe temperatures. The battery packs in both datasets adopt a series-connected configuration consisting of 96 and 91 cells, respectively. Temperature measurements are obtained from 32 and 34 BMS-embedded sensor probes. These sensors are installed at the module or sub-pack level and record the local surface temperatures representative of the overall thermal behavior during vehicle operation. Although the exact probe locations and detailed cell specifications are not publicly disclosed, such probe layouts are standard practice in automotive BMS design for capturing the thermal state of series-connected packs. Consequently, the temperature inputs used in this study reflect realistic BMS-measured values and are representative of SOH-related thermal analysis.

Fig. 1 summarizes the key statistical distributions extracted from Dataset #1. Specifically, Fig. 1(a) shows the long-tail distribution of cumulative mileage, reflecting varying usage intensity across the fleet. Fig. 1(b) and (c) illustrates the distributions of SOC and charging power, respectively, revealing user-dependent charging behaviors and system-level variability. Fig. 1(d) and (e) presents histograms of the minimum and maximum cell voltages and temperatures, respectively, which indicate nonuniformity within the battery modules. The joint distribution of voltage and temperature in Fig. 1(f) further captures their coupling effects under real-world operating conditions. Fig. 1(g) and (h) displays the degradation trajectories of capacity and total energy with respect to mileage. Linear fitting highlights the overall downward trend associated with long-term cycling and usage. Table 1 outlines the nominal specifications of the battery packs used across the fleet. Dataset #2 was collected through the Guangzhou New Energy Intelligent Vehicle Big Data Monitoring Platform. It comprises approximately 4.68 million data frames collected over a continuous one-year period from five EVs, each sampled every 10 s. Notably, these five vehicles are equipped with two distinct battery configurations, differing in both nominal capacity and internal cell layout. This diversity in structural specifications provides a valuable basis for evaluating the proposed model's generalization ability through transfer learning experiments.

Table 1
Nominal specifications for EV field data.

Parameters	Dataset #1	Dataset #2
Number of vehicles	300	5
Nominal capacity	155 Ah	150 Ah/160 Ah
Number of cells	96	91
Number of sensor probes	32	34
Battery materials	NCM	NCM
Collecting period	0.5–4 years	1 year
Sampling interval	10s	10s
Number of data frames	850 million	4.68 million
Total energy	62.49 kWh	57.33 kWh /61.15 kWh
Battery pack topology	Series connection	Series connection

2.2. *Data preprocessing*

Due to the inherent complexity and variability of real-world data collected from EV batteries, comprehensive data preprocessing was essential to ensure data quality and reliability for accurate modeling of battery SOH. Therefore, to ensure the reliability of the modeling, the raw data collected from the EV battery have undergone a systematic preprocessing workflow to address common issues such as quality

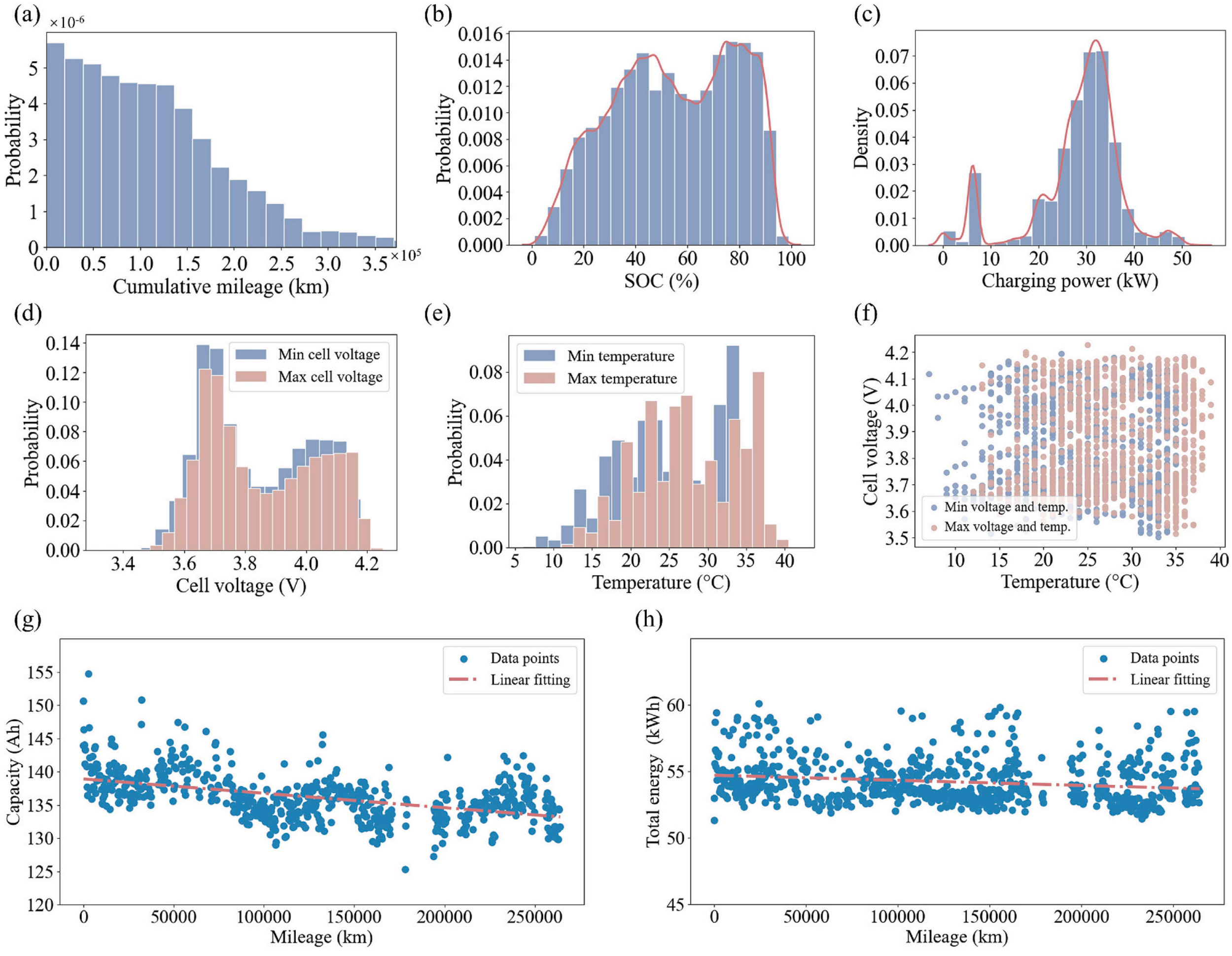

Fig. 1. Distributions of real-world EV operation data.

inconsistencies, sensor anomalies, and structural misalignment in the on-site dataset. Duplicate entries were removed based on timestamp consistency. Outlier detection was then performed using the interquartile range method, targeting extreme values in cell voltages, pack current, and temperature measurements. Observed anomalies—such as physically implausible cell voltages exceeding 6 V, which are inconsistent with the electrochemical characteristics of LIBs—were corrected through localized temporal smoothing. Following outlier correction, the vehicle data were segmented into charging, discharging, and idle periods using a combination of SOC gradient, current polarity, and vehicle speed. Only charging segments were retained for subsequent modeling, as they provide more stable operating conditions and clearer electrochemical signatures than dynamic driving periods, where frequent current reversals and rapid state transitions may obscure degradation indicators.

To ensure sufficient depth of electrochemical reaction during each charging process, segments were further filtered by enforcing a minimum threshold for SOC variation. Only charging events with a cumulative SOC increase of at least 30% were retained. This criterion ensures that each selected segment reflects a meaningful amount of electrochemical activity, which is critical for extracting degradation-relevant temporal features. Incomplete charging segments with missing sensor measurements were identified and reconstructed using linear interpolation, enabling consistent temporal alignment. The cleaned charging records were then processed using a sliding-window approach, with each input sample comprising 20 consecutive frames. The sliding step was set to 1 frame to preserve temporal continuity and capture fine-grained feature transitions. This preprocessing procedure generated a structured dataset suitable for both supervised learning and cross-domain generalization, thereby enabling the model to learn temporal degradation patterns with enhanced robustness, interpretability, and scalability under real-world operating conditions.

2.3. *Calculation of reference capacity labels*

Battery aging is a cumulative process driven by internal electrochemical degradation mechanisms, such as lithium inventory loss and active material loss. SOH is a quantitative indicator that characterizes the extent of degradation and reflects a battery's remaining ability to store and deliver energy. In practical applications, capacity and internal resistance are the two most commonly used parameters for evaluating battery SOH [42,43]. Although resistance growth is an important aging manifestation, its practical use in EV applications is limited because internal resistance is highly sensitive to instantaneous operating temperature, current profile, and measurement protocol, all of which vary significantly during real-world driving. Unlike resistance-based indicators, capacity loss reflects the actual loss of cyclable lithium and active material, providing a physically grounded and chemistry-agnostic measure of performance degradation. Moreover, the maximum available capacity directly determines the vehicle's driving range and overall energy storage capability, both of which are critical for user experience and alleviating range anxiety. Therefore, considering the operational priorities of electric vehicles, this study uses capacity degradation to quantify battery SOH, as defined in Eq. (1),

$$\mathrm{SOH} = \frac{C_t}{C_0} \tag{1}$$

where C_0 represents the nominal rated capacity specified by the battery manufacturer, and C_t denotes the current maximum available capacity of the battery pack.

Determining the precise value of C_t is a prerequisite for SOH quantification. Since the real-world dataset employed in this study does not contain logged SOH estimates or capacity values from the onboard BMS, the ground truth labels must be derived directly from the raw sensor measurements. To achieve this, we use the high-quality charging segments isolated during preprocessing. Compared to dynamic driving cycles, charging processes offer stable current profiles and consistent electrochemical states, making them ideal for SOH estimation. For each retained charging segment satisfying the SOC interval threshold, the maximum available capacity C_t at that aging state is computed using the Ampere-hour integration method and is defined by Eq. (2),

$$C_t = \frac{-\int_{t_0}^{t_i} I(t)dt}{\Delta SOC} \tag{2}$$

where I represents the charging current, and t_0 and t_i denote the start and end times of the charging process, respectively. The denominator, ΔSOC, denotes the corresponding proportional change in the SOC. Physically, this ratio provides a meaningful link between partial charging data and the battery's total capacity. By normalizing the accumulated charge over a segment by the corresponding SOC fraction, the ratio effectively extrapolates the measured partial energy throughput to reconstruct the maximum available capacity of the battery pack at that specific aging state. As the equation represents a ratio, when the ΔSOC is small, minor sensor noise or quantization errors in the BMS SOC readings are disproportionately amplified, leading to severe fluctuations and singularities in the estimated capacity C_t. Conversely, imposing an excessively strict threshold (e.g., $\Delta SOC > 80\%$) would discard a vast majority of real-world charging segments, leading to data scarcity that hinders the model's generalization. To achieve an optimal balance between estimation accuracy and data availability, this study adopts a threshold of $\Delta SOC \geq 30\%$ for selecting charging segments [44]. This selection ensures that the integration interval is sufficiently long to suppress transient sensor noise and minimize the relative error in the ratio, while retaining a statistically significant number of samples for robust model training.

Notably, the SOC values estimated by the BMS and current measurements in real-world settings are subject to noise and cumulative errors, which can introduce outliers in the reference capacity labels calculated using the above method. Furthermore, variability in operational conditions and user behavior often leads to non-negligible errors in SOC estimation. To mitigate these disturbances and extract the robust nonlinear trend of capacity degradation, this study adopts the locally weighted scatterplot smoothing (LOWESS) method [45,46], as shown in Fig. 2. Furthermore, to rigorously quantify the confidence in these labels, a cluster-bootstrapping method is integrated into this calibration process. As depicted in Fig. 2, each charging segment is treated as a sampling cluster, and bootstrap resampling with replacement is performed iteratively. For each iteration, a new LOWESS fit is computed, generating an ensemble of degradation trajectories. From this ensemble, pointwise 95% confidence intervals are constructed using the 2.5th and 97.5th percentiles. This calibration step effectively isolates the true irreversible electrochemical aging signal from stochastic measurement noise, while also providing a measure of segment-aware uncertainty. In the subsequent prediction process, these calibrated capacity values are used as ground-truth labels for the supervised learning framework. Specifically, during the training phase, the model learns to map the input multi-modal charging features to these corresponding calibrated capacity labels. Once trained, the model can directly estimate the battery's current maximum available capacity during inference based on short-term charging snippets, eliminating the need for long-duration charging events or explicit external calibration procedures.

2.4. *Feature engineering*

As shown in Table 2, this study integrates battery aging mechanisms, engineering practice experience, and operational characteristics derived from real-world vehicle data. Based on these factors, a total of 11 health-state features are constructed across thermal, electrical, and temporal dimensions as model inputs. Accumulated mileage, serving as a direct quantitative indicator of vehicle usage intensity, exhibits a long-term

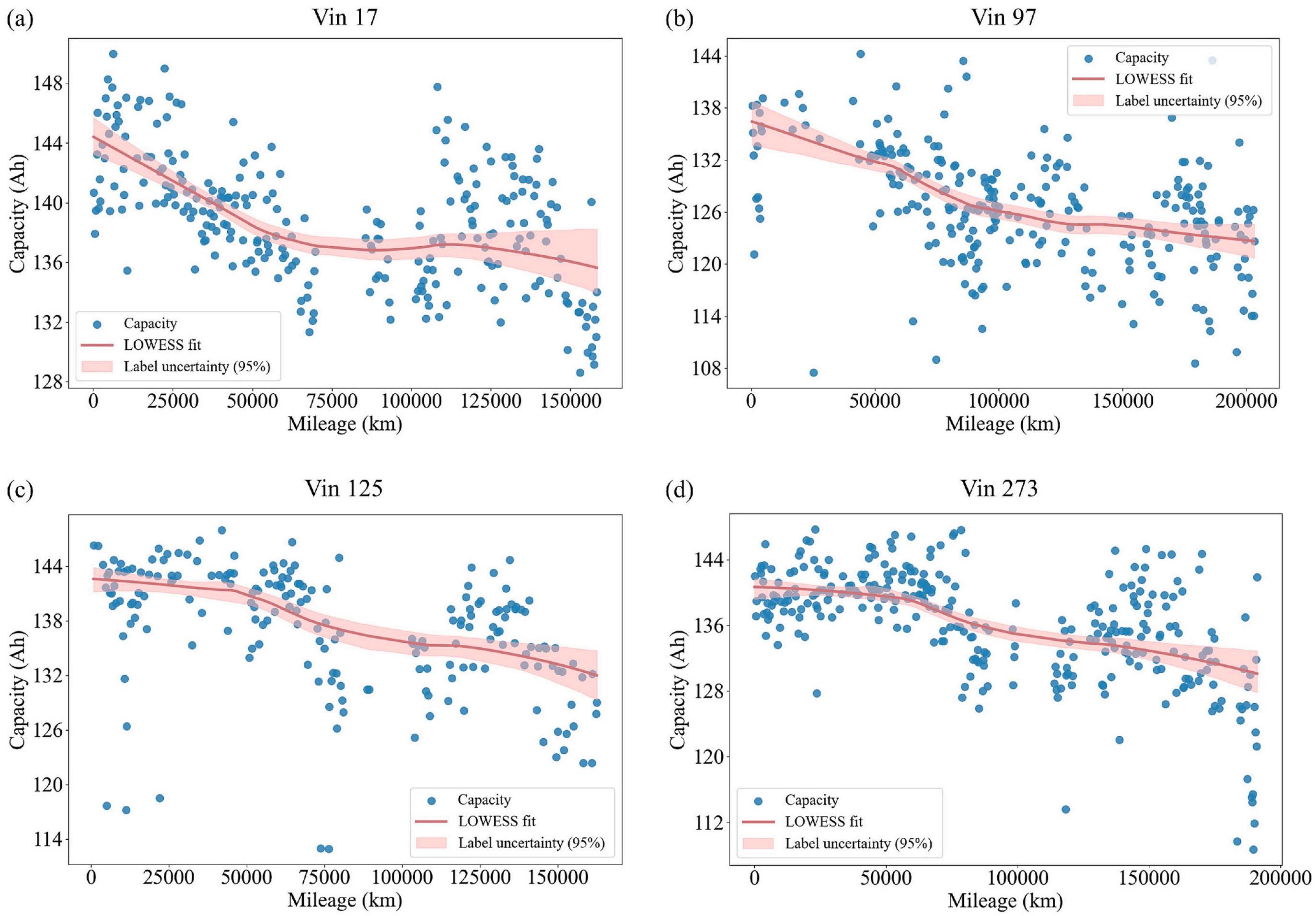

Fig. 2. Capacity degradation trajectories with quantified label uncertainty for four representative EVs.

Table 2
Battery health feature description.

Serial number	Feature type	Health features	Abbreviation
1	Temporal feature	Charging capacity	C_a
2		Cumulative mileage	M_e
3		Charging voltage	V_t
4		Charging current	C_t
5		Charging power	C_p
6		Maximum temperature	T_max
7		Minimum temperature	T_min
8		Maximum cell voltage	V_max
9		Minimum cell voltage	V_min
10	Spatial feature	Single-cell voltage	N_cv
11		Sensor temperature	N_t

correlation with capacity degradation. Its progressive increase reflects cumulative cycling behavior and energy consumption. Charging current and power characterize user charging strategies, indirectly representing the variability in electrical stress conditions and thereby influencing the evolution trajectories of battery life. Extreme values of cell voltage and temperature highlight the inherent nonuniformity of the internal electrothermal field within battery modules, which intensifies over time, ultimately resulting in differential single-cell performance degradation. Furthermore, the joint distribution of voltage and temperature highlights the coupling between extreme thermal conditions and voltage anomalies. These features collectively capture the underlying causes and external manifestations of battery degradation across multiple dimensions.

The Pearson correlation strength of temporal features is illustrated in Fig. 3, where color intensity denotes the correlation of magnitude and direction (red for positive, blue for negative). Scatter plots and histograms are embedded to reveal linear trends and feature distributions. For spatial features, node feature vectors were constructed using single-cell voltage and temperature probe data and subsequently encoded into graph structures through the DGAT module. The graph feature design aims to reflect both the heterogeneity and correlations among individual battery cells, represented as the localized thermoelectric states of each cell unit. Each node corresponds to a single battery cell and features cell voltage and sensor probe temperature. The incorporation of graph-structured spatial features allows the model to explicitly capture spatial interactions and thermoelectric coupling among individual battery cells. Conventional methods often treat battery cells as independent or homogeneous units. In contrast, our graph-based representation captures local anomalies and evolving spatial nonuniformities. This approach enables the model to identify subtle degradation signals that manifest as inter-cell variability. By combining such detailed spatial insights with temporal degradation patterns captured by the aforementioned temporal features, the proposed DGAT-N-BEATS framework achieves improved robustness and generalization capability under diverse operational conditions. This integrated spatial–temporal approach not only enhances estimation accuracy but also provides greater interpretability by elucidating the underlying physical relationships driving battery performance decay.

3. Methodology

Persistent limitations in spatiotemporal modeling, interpretability,

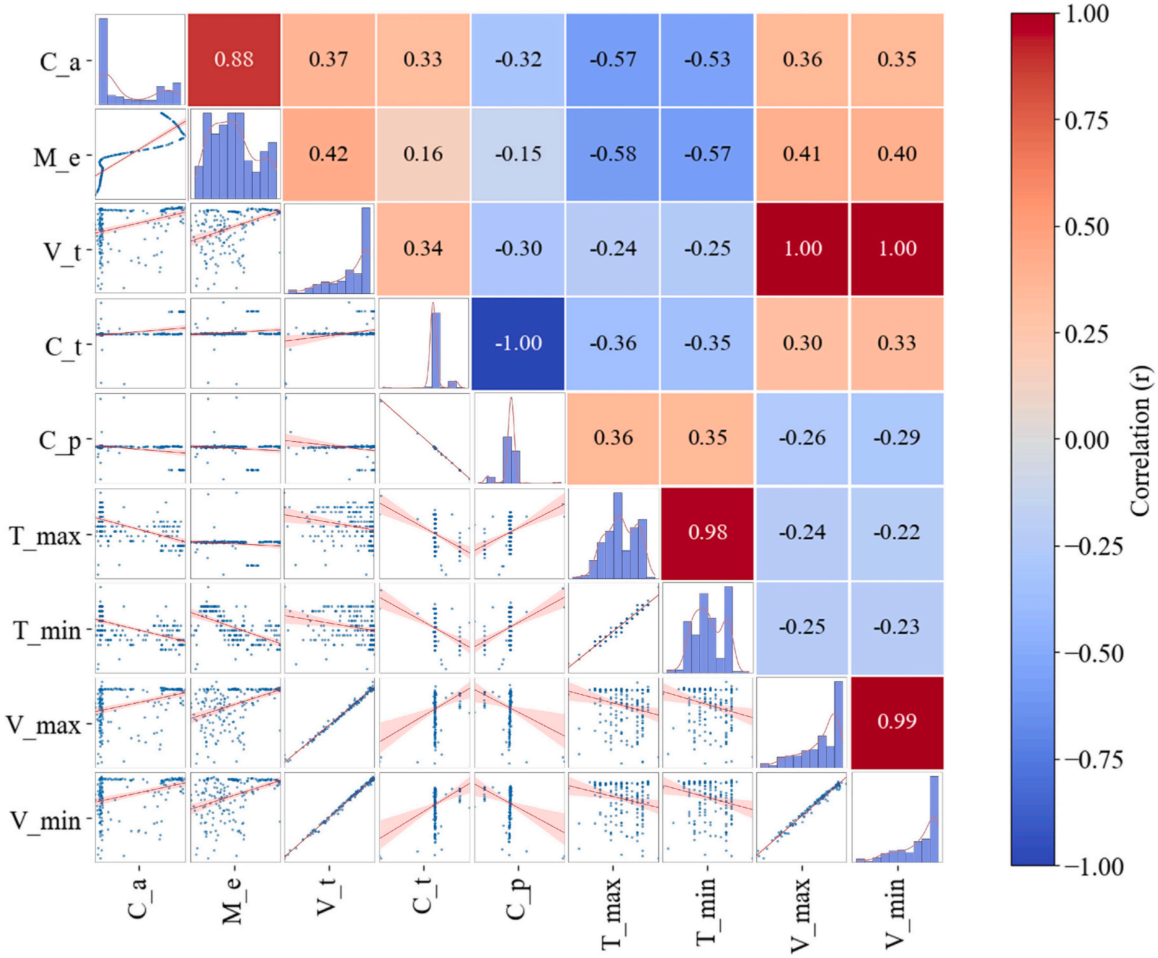

Fig. 3. Pearson correlation heatmap of battery health features.

and cross-domain generalization continue to constrain the advancement of intelligent battery management systems for electric vehicles. To address these challenges, this work introduces a diagnostic framework that integrates spatiotemporal learning capabilities, interpretability, and strong generalization. Fig. 4 presents an end-to-end modeling pipeline encompassing data preprocessing, label construction, spatiotemporal modeling, and final prediction output. Fig. 4 (a) and (b) depicts the data processing pipeline and the spatial dynamic graph construction module, respectively. Fig. 4(c) illustrates the interpretable prediction architecture and the lightweight fine-tuning module. Fig. 4 (d) provides behavior-aware interpretation and prediction analysis. The framework enhances graph construction by learning sample-specific, time-varying inter-cell adjacency directly from raw voltage and temperature signals. A multi-head attention mechanism replaces fixed or heuristic graphs and enables position-aware message passing, revealing dominant cells and critical interactions.

On the temporal side, the N-BEATS forecaster employs a blockwise residual decomposition based on polynomial bases. The resulting coefficients capture global degradation trends and local deviations, aligning predictions with established degradation behaviors and enabling direct attribution. A gated fusion mechanism couples the spatial encoder with the temporal predictor by modulating window-level features using global graph context. This unified representation allows the model to maintain trajectory consistency under heterogeneous usage scenarios, even when local anomalies occur. For cross-domain deployment, the framework utilizes an adapter-based fine-tuning strategy that updates only lightweight modules while keeping the backbone architecture fixed. This design provides behavior-agnostic generalization and maintains stable training costs across different domains.

3.1. Overview of the EV battery health diagnosis framework

The proposed DGAT-N-BEATS framework integrates dynamic spatial encoding, interpretable temporal forecasting, and behavior-agnostic modulation into a unified structure for real-world EV battery capacity estimation. As illustrated in Fig. 5, the architecture departs from conventional black-box predictors by explicitly embedding interpretability and structural flexibility in both its spatial and temporal components. As shown in **Fig. 5**(a), the graph constructor generates a node feature matrix X, derived from voltage and temperature measurements, along with an attention-based adjacency matrix A. As shown in **Fig. 5**(b), these outputs are then fed into the DGAT module, where A is used as edge weights for multi-head message passing to produce the global graph embedding g. Through this design, the model explicitly characterizes localized degradation effects stemming from thermoelectric heterogeneity, while ensuring robust generalization across diverse battery packs and usage patterns. Key innovations include: (1) a learnable attention-based graph constructor that dynamically infers cell-wise interaction graphs from raw telemetry; (2) a residual polynomial forecaster that decouples long-term trends and localized irregularities; (3) an efficient parameter fine-tuning strategy based on adapters; and (4) a modulation layer that couples spatial and temporal features through global graph-aware embeddings, enabling localized behavior to propagate across time.

To capture the latent thermoelectric dependencies among cells,

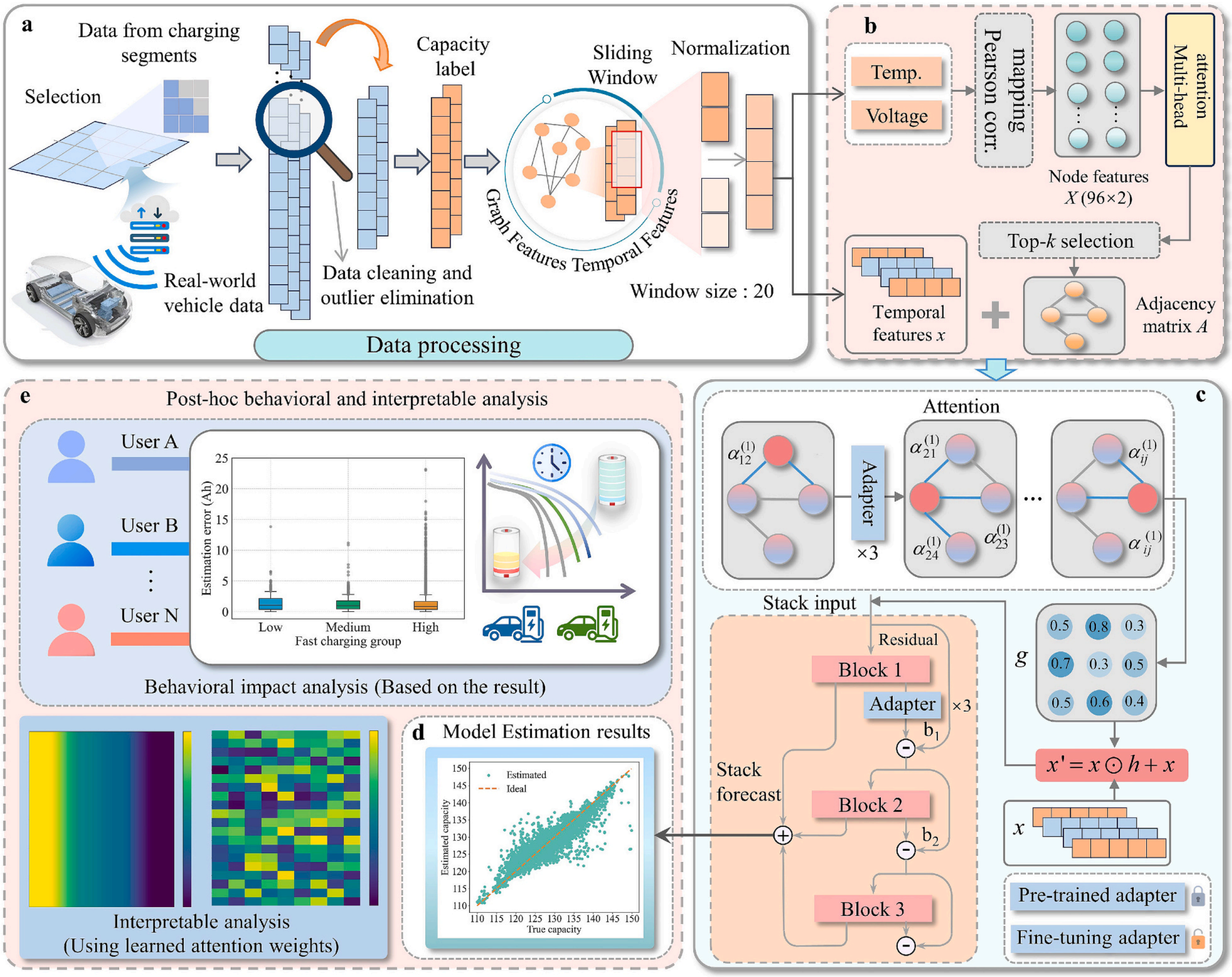

Fig. 4. An interpretable framework for battery capacity estimation using large-scale real-world EV operational data.

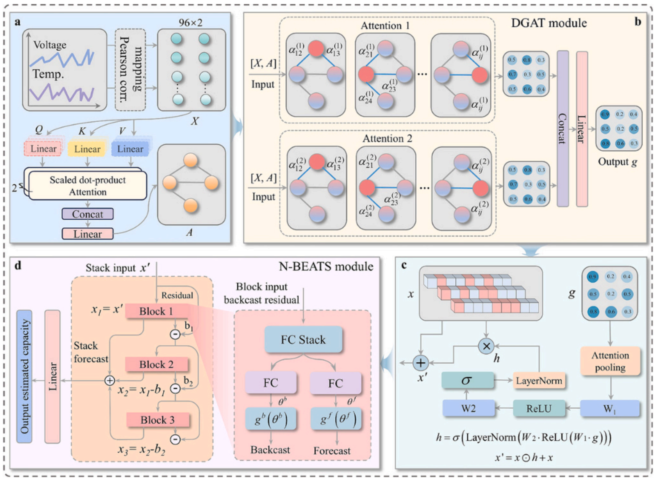

Fig. 5. The interpretable DGAT-N-BEATS architecture for the health diagnosis of EV batteries.

DGAT employs a multi-head attention mechanism to dynamically construct a directed graph based on per-cell voltage and aligned temperature inputs. By representing each battery cell as a node, attention weights are learned to quantify the influence of surrounding cells at each timestep, yielding time-dependent, sample-specific adjacency matrices. This dynamic behavior reflects the evolving nature of thermoelectric coupling; for instance, cells exhibiting similar overpotential behavior at high temperature may become increasingly correlated in their degradation pathways, as captured by strengthened edges. In contrast to static graphs that fix inter-cell relationships, the dynamic graph in DGAT continuously updates its adjacency structure based on the instantaneous electrothermal state. Consequently, the connectivity pattern adapts to charging conditions, temperature gradients, and aging progression, thereby facilitating the precise characterization of real-time interdependencies. The detailed configurations of the model architecture and its key hyperparameters used for implementation and scalability analysis are summarized in Table 3.

3.2. Dynamic graph attention network

To capture the underlying thermoelectric interactions between battery cells, this study constructs a dynamically inferred graph representation, realized through a graph attention network (GAT) [47]. Conventional GAT-based approaches typically rely on fixed adjacency matrices derived from offline preprocessing. In contrast, our proposed DGAT constructs graphs in real-time by directly computing attention-based edge weights from raw sensor inputs at each time step. Specifically, the graph comprises two heterogeneous node types: 96 voltage nodes (each corresponding to a single cell) and 32 temperature nodes (each corresponding to a physical temperature sensor). Since the number of temperature sensors is smaller than the number of battery cells, for each cell i, the temperature sensor k whose historical temperature signal $T_k^{(s)}$ exhibits the highest Pearson correlation with the cell's voltage signal V_i is selected. The corresponding sensor k is then mapped to the cell for alignment, as formally defined in Eq. (3),

$$k^* = \underset{k}{\operatorname{argmax}} \operatorname{corr}\left(V_i, T_k^{(s)}\right), \hat{T}_i = T_{k^*}^{(s)} \tag{3}$$

where each V_i denotes the voltage signal of the i-th battery cell, $T_k^{(s)}$ represents the temperature measured by the k-th physical sensor, corr(·) is the Pearson correlation coefficient computed over historical operating data, and $\hat{T}_i$ is the assigned temperature corresponding to the most correlated sensor. When several sensors show similar correlations, we take the average of the top 3 sensors. Following this mapping, each cell node is represented by a two-dimensional feature vector comprising its voltage and mapped temperature, thereby constructing the input matrix $\mathbf{X} \in \mathbb{R}^{96\times 2}$.

The DGAT module then applies a multi-head attention mechanism to learn the dynamic coupling strength among cells, forming a dynamic adjacency matrix that evolves with the battery's operating condition. The dynamic nature of this graph arises from the mechanism that both the neighborhood structure and the connection strengths between nodes evolve over time, conditioned on the instantaneous electrothermal states. For each node, only the top-k attention coefficients are retained to construct a sparse yet expressive dynamic graph. In contrast, edges associated with stable or thermally isolated cells are automatically down-weighted. This allows the DGAT module to capture non-stationary, condition-dependent thermoelectric coupling, going beyond static physical layouts. Through this adaptive graph construction, DGAT provides a physically grounded yet data-driven representation of evolving heat transfer paths and electrical imbalances in the pack, enhancing the model's ability to detect latent failure patterns and subtle degradation dynamics under real-world conditions. Specifically, the input matrix X, composed of voltage and temperature node features, is fed into the multi-head attention mechanism to learn the thermo-electric coupling relationships between each pair of nodes, as shown in Eq. (4),

$$\alpha_{ij}^{(h)} = \operatorname{softmax}_j\left(\frac{\mathbf{q}_i^{(h)} \cdot \left(\mathbf{k}_j^{(h)}\right)^\top}{\sqrt{d_k}}\right) \tag{4}$$

where $\mathbf{q}_i^{(h)} = \mathbf{W}_Q^{(h)}\mathbf{x}_i, \mathbf{k}_j^{(h)} = \mathbf{W}_K^{(h)}\mathbf{x}_j$ and $h \in \{1, \ldots, H\}$ indexes the attention heads. The adjacency matrix is thus dynamically computed at runtime, enabling the construction of a sample-specific, data-driven graph structure at each forward pass. To ensure sparsity and computational efficiency, the top-k attention scores per node are retained to form the edge set $\mathscr{E} \subseteq \mathscr{V} \times \mathscr{V}$. The resulting graph is passed through a stack of GAT layers to perform localized feature aggregation. Each layer updates nodes as shown in Eq. (4),

$$\mathbf{h}_i' = \sigma\left(\sum_{j\in\mathscr{N}(i)} \alpha_{ij}\cdot \mathrm{W}_V x_j\right) \tag{5}$$

where α is the learned attention weight, W_V is the value projection matrix, and σ denotes a non-linear activation. To improve representation capacity, multi-head attention is employed, and the outputs from all heads are concatenated or averaged. The final representation after L layers, as shown in Eq. (6),

$$\mathbf{H}^{(L)} = \mathrm{GAT}^{(L)}\left(\cdots \mathrm{GAT}^{(1)}(\mathbf{X}, \mathscr{E})\cdots\right) \tag{6}$$

where $\mathbf{H}^{(L)} \in \mathbb{R}^{N\times d_G}$ captures high-order spatial interactions across the battery cells. To bridge the graph encoder with the downstream N-BEATS predictor, an attention-based pooling mechanism is adopted to condense node-wise representations into a global embedding, as defined in Eq. (7),

Table 3
Model hyperparameter description.

Parameter	Description	Value	Role
d_T	Input temperature vector dimension	32	Determines the input dimension of thermal sensor data; supports modeling of spatial thermal effects and thermo-electrical interactions.
d_V	Cell voltage vector dimension	96	Specifies the spatial dimension across cell voltages; enables modeling of inconsistency-driven degradation relations.
W	Temporal window length (number of steps)	20	Determines the history length of the time series; critical for modeling aging trends and usage-dependent degradation.
d_F	Feature dimension per time step	9	Controls the expressive capacity per time step (e.g., voltage statistics, usage behavior); influences the fidelity of temporal modeling.
d_G	DGAT graph output dimension	64	Encodes spatial dependencies among battery cells driven by thermo-electrical coupling.
d_H	DGAT hidden dimension	64	Controls the capacity of the spatial encoder to model thermo-electrical coupling
L	Number of DGAT layers	3	Defines the depth of spatial feature extraction; deeper layers capture long-range dependencies among battery cells.
K	Number of N-BEATS blocks (trend)	3	Governs the decomposition of the SOH time series into trend components; supports long-horizon, interpretable forecasting.
K_L	Number of fully connected layers inside each trend block	4	Determines the non-linear mapping depth from modulated features to trend-basis coefficients
d	Degree of trend polynomial basis	3	Specifies the degree of polynomial basis for trend blocks in N-BEATS; balances fit accuracy and overfitting risk.
h	Hidden layer size in fully-connected blocks	256	Affects model capacity for nonlinear transformations; key for learning complex degradation trajectories.
p	Dropout rate	0.3	Regularization to prevent overfitting, especially important given the scale of the input data and model size.
lr	Learning rate	0.001	Controls training stability and convergence speed; tuned for large-scale EV time series.
a	Activation function	ReLU	Introduces nonlinearity across all fully connected layers; balances simplicity and model expressiveness.
B	Batch size of training samples	256	Larger batches stabilize updates and improve throughput, smaller batches enhance regularization but may yield noisier gradients.
E	Number of training iterations	100	Defines the maximum training horizon.
P	Early stopping patience	15	Prevent overfitting by stopping training when generalization stops improving.
top-k	Number of top k attention weights to retain	16	Sets graph sparsity and computational cost.
H	Number of attention heads in dynamic graph construction	2	Compute node features and adjacency matrices for GAT in real time.
H_G	Number of heads in GAT	2	Provides multi-channel aggregation over neighbors and improves spatial expressiveness

$$\mathbf{h}_G = \sum_{i=1}^{N} \beta_i \cdot \mathbf{h}_i, \beta_i = \frac{\exp(\mathbf{w}^\top \tanh(\mathbf{W}_h \mathbf{h}_i))}{\sum_j \exp(\mathbf{w}^\top \tanh(\mathbf{W}_h \mathbf{h}_j))} \tag{7}$$

where β_i denotes the learned importance of node *i,* and summarizes the spatial state of the battery system. This global feature serves as a modulation vector for time series estimation in the subsequent prediction stage. In summary, the proposed DGAT module distinguishes itself from traditional GAT-based architectures by performing fully dynamic, attention-driven graph construction at each time step, rather than relying on static or preprocessed topologies. This design ensures the model remains sensitive to real-time changes in battery state, enabling fine-grained, context-aware learning of spatial interactions—capabilities essential for accurate and interpretable SOH estimation in real-world EV deployments.

3.3. Interpretable temporal modeling via N-BEATS

To effectively capture the temporal evolution of battery degradation, this study adopts a structured and interpretable architecture based on N-BEATS [48]. This architecture fundamentally diverges from conventional black-box sequence modeling paradigms by systematically decomposing temporal signals into additive, interpretable basis functions. This approach establishes an explicit mathematical framework for characterizing long-term capacity degradation trajectories by incorporating three synergistic technical principles. The orthogonality of basis functions enables independent analysis of degradation trends, and the hierarchical residual structure further captures multi-scale temporal patterns across different operational phases. Crucially, the model achieves diagnostic interpretability by quantifying component-wise contributions, which associate specific basis functions with physical degradation mechanisms. This architectural design transforms capacity fade modeling from opaque extrapolation to mathematically traceable component interactions. It thereby provides both predictive accuracy and engineering-relevant degradation insights that are unavailable in traditional end-to-end approaches.

The input to the N-BEATS module is a fixed-length sliding window of operational features, denoted as $x \in \mathbb{R}^{T \times D}$,where T is the window length and D is the feature dimension. To integrate the structural information extracted by the DGAT module, it introduces a feature modulation mechanism based on the global graph embedding $g \in \mathbb{R}^{d_g}$, which captures cell-level interactions within the battery pack. This modulation applies a gated transformation to reweight the temporal input features, enhancing the ability of the model to adapt its temporal reasoning to the spatial state of the system. The modulation process applies a gated transformation to reweight the temporal features based on the structural representation extracted by DGAT, as shown in Eqs. (8) and (9),

$$h = \sigma(\mathrm{LayerNorm}(W_2 \cdot \mathrm{ReLU}(W_1 \cdot g))) \tag{8}$$

$$x' = x \odot h + x \tag{9}$$

where $W_1 \in \mathbb{R}^{d_h \times d_g}$and $W_2 \in \mathbb{R}^{D \times d_h}$ are learnable weight matrices, and $\odot$ denotes element-wise multiplication. The output $x' \in \mathbb{R}^{T \times D}$ is the modulated feature matrix, enriched by global spatial information. The temporal modeling is structured into a sequence of residual blocks. Each block receives the modulated feature vector and produces two outputs: a backcast component $\hat{x} \in \mathbb{R}^{T}$ used to reconstruct the observed input window and a forecast component $\hat{y} \in \mathbb{R}^{T'}$ to estimate the forward degradation trajectory. Formally, for block *l*, the outputs are as shown in Eqs. (10) and (11).

$$\left[\hat{x}^{(l)}, \hat{y}^{(l)}\right] = f^{(l)}\left(x^{(l)}\right) \tag{10}$$

$$x^{(l+1)} = x^{(l)} - \hat{x}^{(l)} \tag{11}$$

Each block is configured with interpretable basis functions designed to capture long-term degradation trends. The forecast $\hat{y}^{(l)}$ is expressed as a weighted sum over polynomial bases, as shown in Eq. (12),

$$\hat{y}^{(l)} = \sum_{k=0}^{K} \theta_k^{(l)} \cdot \phi_k(t), t \in [0, 1] \tag{12}$$

where $\varphi_k(t) = t^k$ denotes the *k*-th order polynomial basis, and $\theta_k^{(l)} \in \mathbb{R}$ is the learned coefficient. The use of explicit basis expansion allows the model to link its outputs to interpretable physical phenomena—e.g., linear fade, accelerated decay, or plateau phases. After *L* residual blocks, the final degradation prediction is obtained by aggregating the forecasts, as shown in Eq. (13).

$$\hat{y}_{\mathrm{f}} = \sum_{l=1}^{L} \hat{y}^{(l)} \tag{13}$$

The N-BEATS-based temporal module in the proposed framework serves as a transparent and interpretable learner of degradation dynamics.

3.4. Model training of the proposed framework

To quantify the discrepancy between the estimated battery capacity and the ground truth, the mean squared error (MSE) is employed as the objective function, as defined in Eq. (14),

$$\mathscr{L}_{\mathrm{MSE}} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{14}$$

where y_i and $\hat{y}_i$ are the actual and predicted values, respectively, and N denotes the number of samples used in the loss computation. In the proposed framework, the Adam optimizer [49] is employed to iteratively update the model weights to minimize this loss. The optimizer is configured with its default parameter settings. To ensure stable convergence, a dynamic learning rate schedule is implemented: the learning rate is initialized to the value specified in Table 3. It is decayed by a factor of 0.5 if the validation loss does not decrease for 5 consecutive epochs. Furthermore, to prevent overfitting and enhance generalization, a dropout regularization strategy is applied alongside an early stopping mechanism. As indicated in the training configuration, the training loop terminates automatically if the validation metric fails to improve after 15 epochs. Regarding the hyperparameter selection listed in Table 3, a two-fold strategy was adopted. The structural and optimization hyperparameters, including the graph embedding dimensions (d_G, d_H), network depth (L), N-BEATS configuration (K, K_L, d, h), attention heads (H, H_G), dropout rate (p), learning rate (lr), and batch size (B), were determined via a grid search algorithm on the validation set. In contrast, the parameters governing spatiotemporal complexity, specifically the sliding window length (W) and the graph sparsity parameter (top-k), were selected based on a rigorous trade-off analysis between computational cost and estimation accuracy, as detailed in the ablation study in Section 4.4. The pseudocode of the training logic is summarized in Table 4.

3.5. Transfer learning

To enhance cross-vehicle generalization under distribution shifts, this study employs a parameter-efficient transfer learning strategy based on adapter-based fine-tuning (Fig. 6) [50]. The DGAT-N-BEATS model is first pretrained on a large-scale source dataset (Dataset #1), which provides abundant historical records for a fleet of vehicles with a given battery configuration. The resulting backbone captures generic thermoelectric coupling patterns and temporal degradation dynamics at the pack level. The pretrained model is then adapted to a structurally distinct target domain (Dataset #2), which contains EVs with different

Table 4
Training procedures of the proposed framework.

Algorithm 1: DGAT-N-BEATS for EV battery capacity estimation

```
Input data: Per-window node voltages $\boldsymbol{V}_{(i)}$, node temperatures $\boldsymbol{T}^{(s)}$; temporal features $\mathbf{X}_{\text{temp}}$.
Dynamic graph construction (DGAT):
  (1) The correlation-based mapping constructs the input feature matrix $\mathbf{X} \in \mathbb{R}^{96\times2}$;
(2) Compute attention-based edge weights between nodes (Eq. (4));
(3) For each node, keep top-$k$ neighbors; set adjacency $\mathscr{E}$;
  (4) Apply stacked GAT layers on $(\mathbf{Z}, \mathscr{E})$ to obtain node embeddings $\mathbf{H}^{(L)}$;
(5) Attention-pool $\mathbf{H}^{(L)} \rightarrow$ global graph embedding $\mathbf{h}_{\mathbf{G}}$ (Eq. (7)).
Graph feature modulation:
  (1) Compute the gate vector for the global embedding $\boldsymbol{h}$ (Eq. (8));
  (2) Perform feature modulation using Eq. (9).
Interpretable temporal modeling (N-BEATS):
  (1) Initialize residual input with $x'$; set the forecast accumulator to zero.
  (2) For each residual block:
  (a) Produce a forecast for the window using a polynomial basis (Eq. (12));
  (b) Accumulate the forecasts across blocks.
  (3) Aggregate all block forecasts to obtain the final prediction (Eq. (13)).
  Training and validation:
  Initialize network parameters $\omega$.
  Input: optimizer, learning-rate policy, maximum epochs.
  for $i$ epoch do
  (1) Training loop: run DGAT → Modulation → N-BEATS to get $\widehat{y}_{\text{SOH}}$; compute $L_t = \text{MSE}\ (\widehat{y}_{\text{SOH}}, y_{\text{SOH}})$; backpropagation; update $\omega$.
  (2) Validation: forward on the validation set to obtain $L_v$;
  apply early stopping; save the best checkpoint $\omega^*$ if $L_v$ improves.
  end for
  Testing: Load $\omega^*$; forward on the test set to obtain$\widehat{y}_{\text{SOH}}$; report RMSE, MAPE, $\text{R}^2$.
  Output: Estimated capacity and the best model $\omega^*$.
```

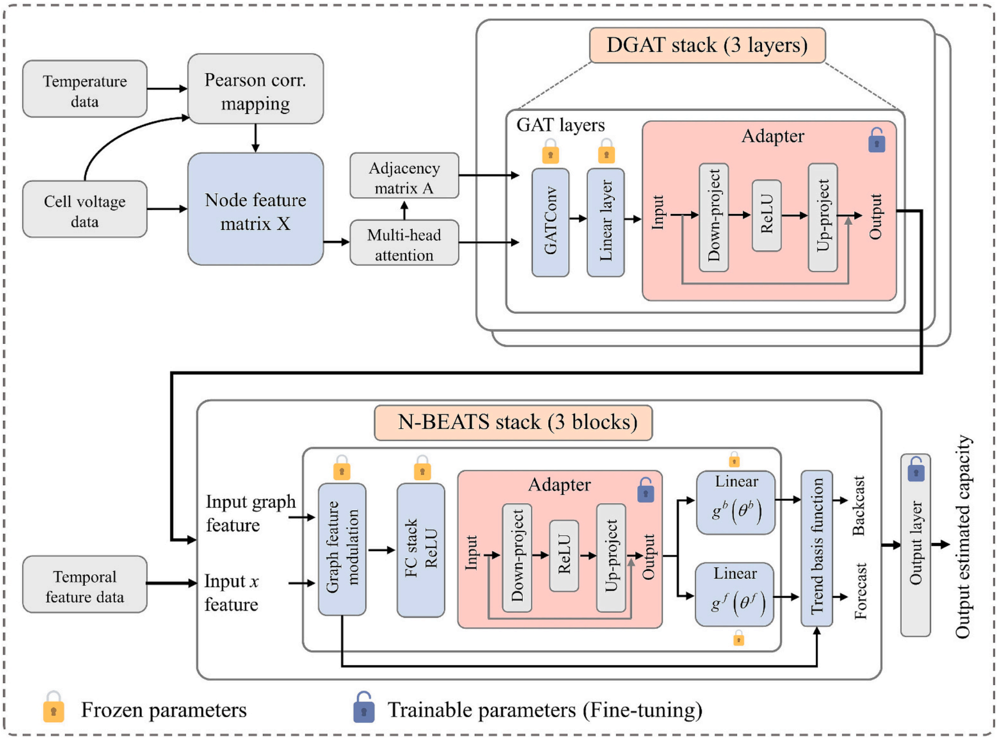

Fig. 6. Cross-vehicle transfer learning framework with dynamic graph construction and adapter-based fine-tuning.

battery pack configurations and usage profiles but only limited labeled data. Instead of fine-tuning all backbone parameters on Dataset #2, lightweight bottleneck adapters are inserted into selected DGAT layers and N-BEATS blocks, and the transfer stage primarily updates these adapter parameters. For an intermediate feature vector $x \in \mathbb{R}^d$, each adapter applies a low-dimensional transformation, defined by Eq. (15),

$$\text{Adapter}(x) = W_{\text{up}}\sigma(W_{\text{down}}x) + x \tag{15}$$

where $W_{\text{down}} \in \mathbb{R}^{d\times b}$ projects the features into a bottleneck space with $b \ll d$, $W_{\text{up}} \in \mathbb{R}^{d\times b}$ maps them back to the original dimension, and $\sigma(\cdot)$ denotes a nonlinear activation function. The residual connection ensures that the adapters learn domain-specific residual corrections on top of the pretrained representations, rather than overwriting them, which is particularly well-suited for capturing subtle distributional shifts across different vehicle fleets and pack configurations. During adaptation to Dataset #2, the entire network is not fine-tuned. In particular, only the bottleneck adapters inserted into the DGAT and N-BEATS blocks and the SOH prediction head are updated. At the same time, all remaining layers of the pretrained backbone are kept frozen. In this way, more than 90% of the original parameters are kept fixed, preserving the pretrained knowledge from Dataset #1 and mitigating catastrophic forgetting. In contrast, the small set of trainable modules is sufficient to capture target-domain-specific patterns. This parameter-efficient design substantially reduces the data requirements and computational overhead of transfer learning, enabling stable convergence even when only 10–30% of the target-domain samples are used for fine-tuning.

A key challenge in cross-vehicle transfer is that different EV platforms may contain battery packs with varying numbers of cells and sensor channels, leading to graphs of various sizes. The proposed DGAT-N-BEATS framework naturally accommodates such structural variations because the underlying graph construction and message-passing mechanisms are node-invariant. Specifically, the dynamic graph attention module constructs the adjacency matrix at runtime solely from voltage–temperature feature embeddings, without relying on any fixed pack topology. Regardless of pack size, each graph node consistently represents a single cell, and its feature vector is constructed in the same manner from voltage- and temperature-derived signals. Temperature channels are associated with individual cells via a Pearson–correlation–based mapping, ensuring that the node feature definition remains consistent across datasets, even when the number of cells differs. The adjacency structure is not inherited from the source vehicle's physical topology; instead, it is reconstructed dynamically on the target vehicle using multi-head attention. For any node count N, the dynamic graph attention module computes an $N \times N$ interaction matrix and selects the strongest relations to form a sparse dynamic graph. Because DGAT message-passing weights are shared across nodes and do not depend on N, the pretrained encoder can operate seamlessly on graphs with different sizes. Finally, an attentive pooling operator aggregates node-level embeddings into a fixed-dimensional pack representation, eliminating dependence on node count before the temporal forecasting stage. This design enables architecture-preserving transfer from a 96-cell pack to a 91-cell pack without requiring explicit one-to-one node correspondence, while adapting the model to new thermoelectric patterns through lightweight adapter updates.

Conventional full-model fine-tuning updates all parameters in both the DGAT encoder and the N-BEATS temporal blocks, allowing unrestricted changes throughout the entire representation hierarchy. Although this can increase flexibility, it also makes optimization highly unstable when only 10–30% of target-domain data are available, leading to overfitting and destructive drift in the pretrained graph and temporal representations. Low-rank adaptation (LoRA) [51] offers a more parameter-efficient alternative by injecting trainable low-rank matrices into the query, key, and value projections of the attention modules, thereby modifying the weight matrices directly at the parameter level. However, LoRA is tightly coupled with Transformer-style attention and does not intervene at the feature-layer level; as a result, it cannot simultaneously modulate both the graph-structure inference and the downstream temporal forecasting pathway. Gradual unfreezing instead begins by training only the top layers of the backbone and progressively releases deeper layers during optimization [52]. While this strategy reduces catastrophic forgetting relative to full fine-tuning, it still requires updating a large portion of the backbone parameters, leading to substantially higher variance and a greater risk of overfitting on small target-domain datasets. By contrast, the proposed adapter-based strategy performs domain adaptation through lightweight, bottleneck residual modules inserted into both the DGAT and N-BEATS feature blocks. These adapters implement compact, low-rank feature transformations without modifying the pretrained backbone parameters, ensuring that the learned thermoelectric interaction patterns and degradation dynamics remain intact while allowing domain-specific corrections to be learned efficiently. This mechanism enables stable adaptation with more than 90% of parameters frozen, jointly influencing graph construction, feature propagation, and temporal prediction in a manner neither LoRA nor gradual unfreezing can achieve. As demonstrated in Section 4.2, this yields consistently lower estimation errors and narrower error distributions under identical target-domain data budgets, confirming that the proposed mechanism offers a more suitable and robust transfer learning paradigm for cross-vehicle capacity estimation.

4. Results and discussion

4.1. Capacity estimation results of the proposed method

To verify the effectiveness of the DGAT-N-BEATS model in capacity estimation for real-world vehicles, this study evaluates its predictive performance using field-operational data. Here, the dataset is split using a vehicle-level protocol. Specifically, vehicles are divided into 70% for training, 10% for validation, and 20% for testing. The dataset is partitioned using a fixed random seed to ensure reproducibility. All models are implemented using the PyTorch deep learning framework (Python 3.9) and executed on a hardware platform featuring an Intel Core i5-14600KF processor with 32 GB of random access memory, supplemented by an NVIDIA RTX 4070 graphics processing unit (GPU) equipped with 12 GB of dedicated video memory. To demonstrate the advantages of the proposed DGAT-N-BEATS for battery-capacity estimation, this study compares it with typical models, including CNN, Transformer, recurrent neural network (RNN), extreme gradient boosting (XGBoost), and support vector regression (SVR), as well as three strong task-specific baselines from the literature: N-CatBoost model [41], CNN-Transformer model [31], and deep fusion transfer learning network (DFTN) [32]. To ensure a fair and consistent benchmarking protocol, all compared models, including tree-based methods, are trained and evaluated under identical data splits, preprocessing procedures, and evaluation metrics. During training, all baseline models use the same optimizer, learning rate, hidden layer dimension, and batch size configurations. The number of training epochs is fixed at 100, with early stopping applied based on validation performance. No model-specific hyperparameter tuning is performed beyond standard default values, and the training framework remains consistent with that of the proposed DGAT-N-BEATS model to isolate the effect of each architectural modification. This controlled and unified protocol ensures that inconsistencies do not influence performance comparisons in data processing, training procedures, or evaluation criteria.

Fig. 7(a) shows that the predictions of the proposed DGAT-N-BEATS model closely follow the ideal diagonal reference line. This indicates excellent agreement between the estimated and true capacities. By comparison, baseline methods show larger estimation deviations. The CNN model, shown in Fig. 7(b), and the Transformer model, shown in Fig. 7(c), demonstrate evident scatter and systematic discrepancies, particularly at higher capacities. Similarly, predictions from the RNN model, shown in Fig. 7(d), deviate significantly from true values, whereas classical methods such as XGBoost, shown in Fig. 7(e), and SVR, shown in Fig. 7(f), reveal notable prediction dispersion, reflecting difficulties in accurately capturing complex degradation behaviors. Furthermore, Fig. 7(g)–(i) illustrate the results of several recent hybrid and ensemble methods. The DFTN model (Fig. 7(g)) achieves moderate accuracy but still underestimates mid-range capacities. The N-CatBoost model (Fig. 7(h)) reduces some of the variance but struggles with high-capacity samples. The CNN-Transformer hybrid model (Fig. 7(i)) improves local trend fitting yet introduces notable overfitting artifacts, especially near the tail of the distribution. These comparative results underscore the superior balance of accuracy and generalization achieved by the DGAT-N-BEATS architecture. Fig. 8 presents the estimated versus actual capacity trajectories for six representative vehicles from the test set. These results highlight the model's ability to accurately track the capacity degradation trends across diverse degradation patterns and aging behaviors. Despite variations in battery aging rates and usage

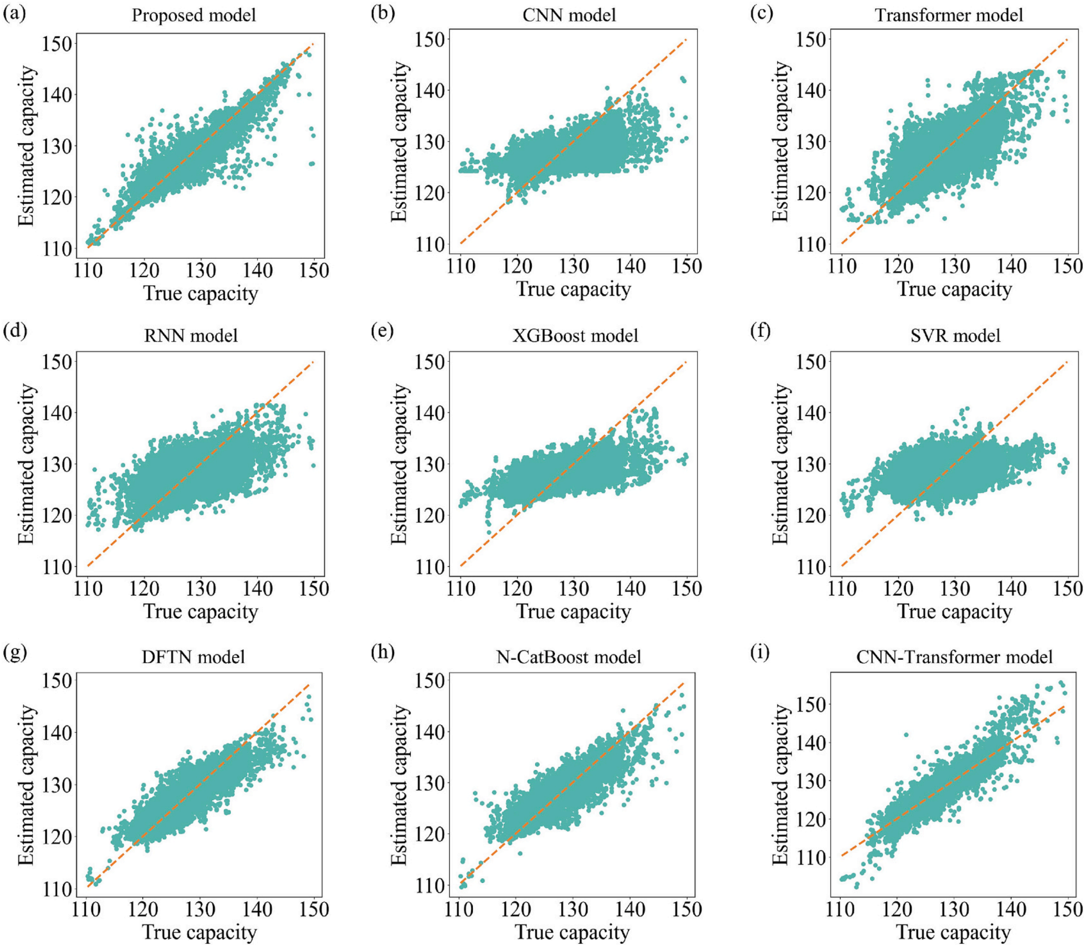

Fig. 7. Comparison of the proposed method and baseline models for EV battery capacity estimation.

dynamics, the DGAT-N-BEATS model consistently provides smooth, accurate estimates throughout the full lifecycle, confirming its robustness and transferability to unseen vehicles. To quantitatively assess model performance, three widely used regression evaluation metrics are adopted: RMSE, coefficient of determination (R^2), and MAPE. These metrics are defined as Eqs. (16)–(18),

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{16}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}^i - y_i)^2}{\sum_{i=1}^{n}(y^i - \bar{y})^2} \tag{17}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{18}$$

where y_i and $\hat{y}_i$ denote the true and predicted values, respectively, $\bar{y}$ is the mean of the true values, and n is the total number of samples.

Quantitative evaluation metrics are summarized in Table 5, further confirming these visual observations. The proposed DGAT-N-BEATS method achieves significantly superior results, exhibiting the lowest RMSE of 1.22%, the highest R^2 of 0.88, and the lowest MAPE of 0.98%. Comparatively, the SVR model underperforms substantially, yielding the highest RMSE (3.32%), the lowest R^2 (0.11), and the highest MAPE (3.37%), suggesting poor prediction reliability and limited generalization capabilities. Transformer and RNN models, while outperforming CNN, still lag behind DGAT-N-BEATS with higher RMSEs of 2.17% and 2.71%, respectively, and notably lower R^2 values. XGBoost and CNN exhibit markedly inferior performance, with RMSE values of 2.81% and 2.98%, respectively, and corresponding R^2 values that are substantially lower than those of the proposed method. Among the advanced task-specific models, DFTN achieves relatively strong performance with an RMSE of 1.49% and an R^2 of 0.83, while N-CatBoost and CNN-Transformer achieve MAPEs of 0.82% and 0.94%, respectively. Nevertheless, DGAT-N-BEATS consistently outperforms all of them across multiple metrics, indicating both higher predictive accuracy and improved generalizability under real-world conditions. The superior performance of the DGAT-N-BEATS model stems primarily from its hybrid structure. This design seamlessly integrates graph attention mechanisms to capture spatial interactions with polynomial-based trend decomposition for temporal modeling. Overall, the results clearly demonstrate that the proposed DGAT-N-BEATS model outperforms both traditional and state-of-the-art baselines, highlighting its strong potential for reliable and accurate battery capacity estimation in real-world battery management systems.

To evaluate the feasibility of real-time deployment and investigate the trade-off between accuracy and computational cost, a comprehensive comparison of inference latency, floating-point operations (FLOPs), and parameter size across deep learning models was conducted, summarized in Table 6. While the proposed DGAT-N-BEATS model exhibits higher inference latency (43.04 ms) than the CNN-Transformer (11.26 ms), this increase is primarily due to the computational overhead of dynamic graph construction and the architecture's sequential computation. However, a deeper analysis reveals that the proposed framework offers significant advantages in storage and energy efficiency. As shown in Table 6, the proposed model requires only 1.69 M parameters and

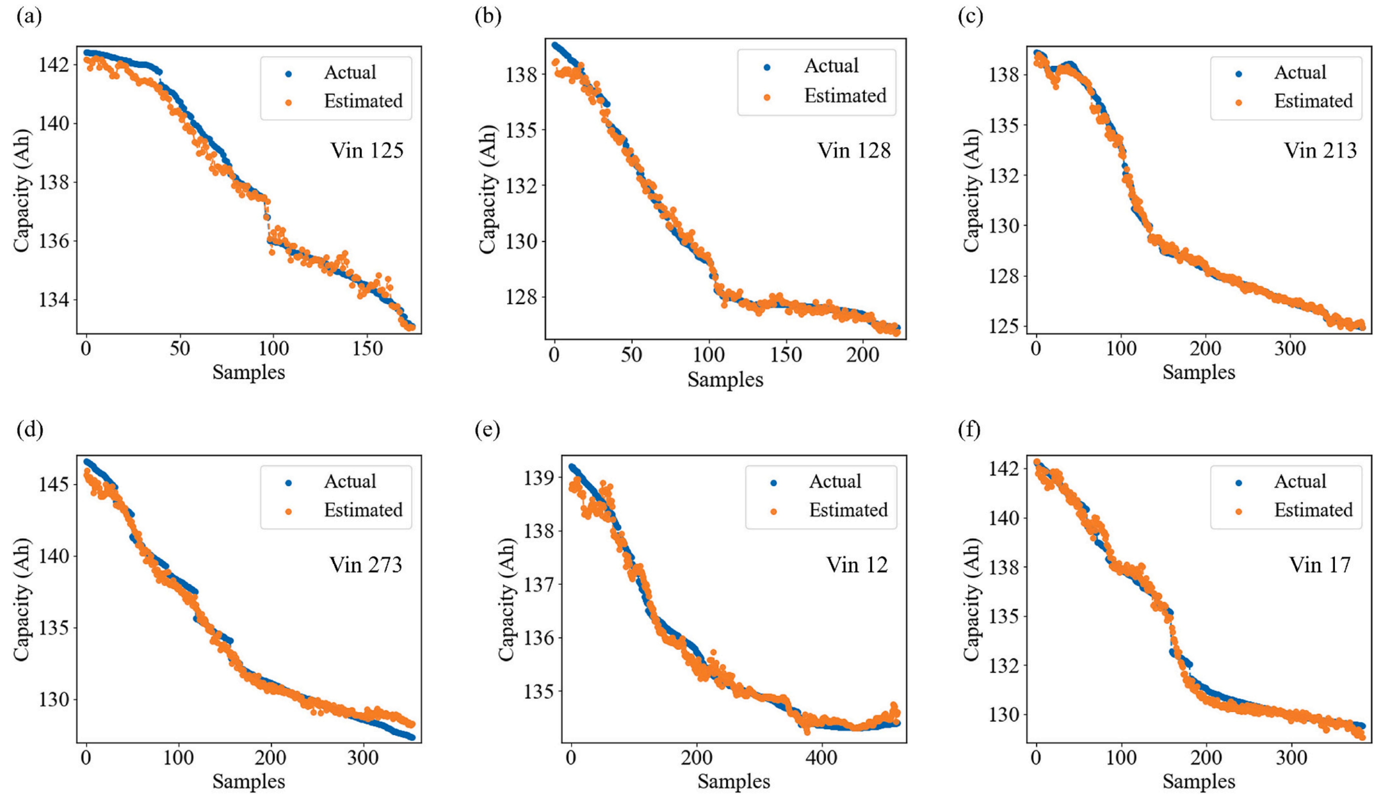

Fig. 8. The capacity estimation results of six representative vehicles in the test set.

Table 5
Performance comparison of different models for EV battery capacity estimation.

Models	RMSE (%)	R^2	MAPE (%)
DGAT-N-BEATS (our proposed)	1.22	0.88	0.98
CNN	2.98	0.28	2.90
Transformer	2.17	0.61	1.98
RNN	2.71	0.41	2.59
XGBoost	2.81	0.36	2.75
SVR	3.32	0.11	3.37
DFTN	1.57	0.79	1.45
N-CatBoost	1.69	0.76	1.54
CNN-Transformer	1.46	0.82	1.25

Table 6
Comparison of computational cost between the proposed DGAT-N-BEATS and other deep learning models.

Models	Parameters (M)	FLOPs (G)	Inference time (ms)
DGAT-N-BEATS (our proposed)	1.69	1.23	43.04
CNN	0.40	2.05	3.25
Transformer	5.53	8.51	4.69
RNN	0.33	1.70	3.78
DFTN	2.53	7.11	2.31
CNN-Transformer	8.17	40.51	11.26

1.23 GFLOPs, which correspond to approximately one-fifth of the parameter size and one-thirtieth of the computational load of the CNN-Transformer (8.17 M parameters and 40.51 GFLOPs). This indicates that although the inference time is slightly longer, the proposed architecture is remarkably lightweight and energy-efficient, making it highly suitable for storage-constrained edge BMS where energy consumption is a critical concern. Furthermore, unlike black-box deep learning models, the N-BEATS backbone in the proposed architecture inherently offers interpretability by decomposing the degradation trajectory into an interpretable trend component, providing valuable physical insights for battery health management.

Beyond standard accuracy, we further investigated the robustness of the proposed architecture to sensor noise. Fig. 9 illustrates the performance degradation of different models under Gaussian noise ranging from 0% to 10%. The results demonstrate that high-complexity models such as the CNN-Transformer and Transformer experience a sharp performance decline (with RMSE spiking from 1.46% to 3.35%), as their attention mechanisms are highly sensitive to noise and often amplify high-frequency disturbances. Conversely, while traditional machine learning models (e.g., SVR) and simple CNNs are less sensitive to noise, their baseline accuracy is limited by their inability to capture subtle degradation patterns. In contrast, the DGAT-N-BEATS model achieves a superior balance, maintaining exceptional stability (RMSE increasing only to 1.76% at 10% noise). This confirms that the proposed framework not only achieves high precision but also possesses the critical robustness required for reliable operation in complex, noisy electromagnetic environments.

4.2. *Cross-vehicle capacity estimation performance*

To evaluate the cross-vehicle adaptability of the proposed DGAT-N-BEATS model, a lightweight adapter-based fine-tuning strategy is applied to transfer a pretrained model (on Dataset #1) to target vehicles with different battery configurations (Dataset #2) for SOH estimation. During transfer, only the parameters within the inserted adapter modules are updated, while the backbone network remains frozen. This approach greatly reduces training overhead and enhances adaptation under limited data conditions. Fig. 10(a) – (g) present the comparison between estimated and true battery capacities under different proportions of fine-tuning data, ranging from 10% to 70%. As the fine-tuning ratio increases, the predicted values align more closely with the ideal diagonal, indicating a significant improvement in estimation

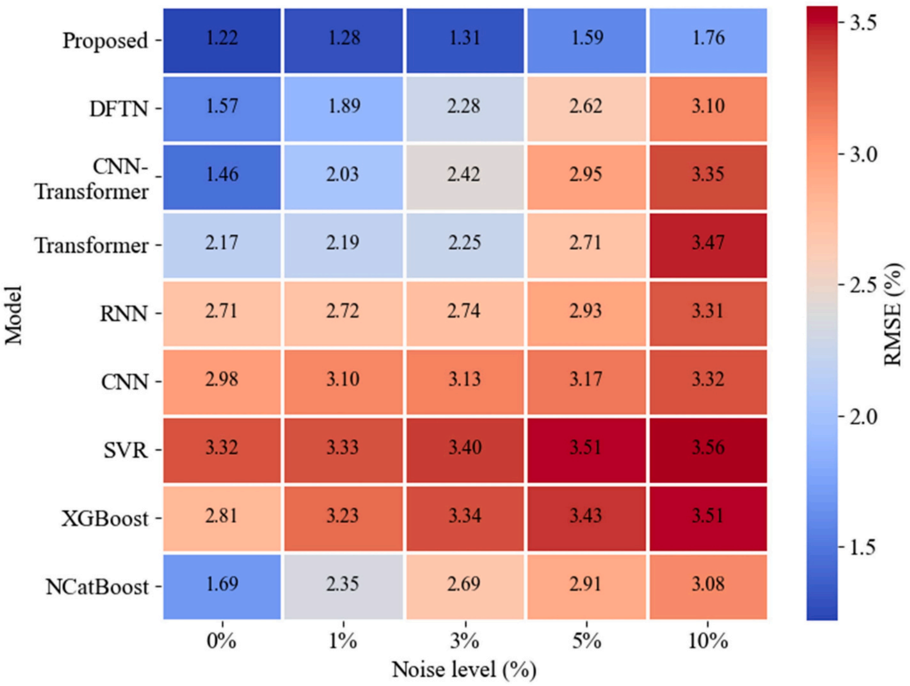

Fig. 9. Robustness evaluation of SOH estimation models under varying noise.

accuracy. Fig. 10(h) shows the results without any fine-tuning, where clear underestimation and systematic deviations are observed. Fig. 10(i) further summarizes the distribution of estimation errors under varying fine-tuning ratios. Even with only 10% of the target data, the adapted model significantly reduces estimation variance and bias, while higher ratios further enhance prediction stability.

Table 7 presents quantitative evaluation metrics including RMSE, R^2, and MAPE for different fine-tuning ratios. The results show a clear trend of performance improvement as the fine-tuning ratio increases, with diminishing gains and convergence behavior observed once the ratio exceeds 60%. As the fine-tuning sample ratio increases, model performance continuously improves and tends to converge when the ratio exceeds 60%. Specifically, with 70% data for fine-tuning, the model achieves optimal performance (RMSE $=$ 1.23%, $R^2 = 0.90$, MAPE $=$ 0.94%). In comparison, the error of the model without fine-tuning is significantly higher on the same test set (RMSE $=$ 6.89%, $R^2 = -2.09$, MAPE $=$ 5.74%). In summary, the proposed adapter-based strategy achieves a favorable balance between data efficiency and generalization performance. With as little as 10%–30% of the target domain data, the model achieves satisfactory accuracy, while additional data further refines the performance. These results confirm the transferability of the proposed model across battery systems with varying pack configurations and demonstrate its potential for scalable deployment in real-world EV fleets. Furthermore, the adapter-based fine-tuning strategy is benchmarked against other parameter-efficient transfer methods, as shown in Fig. 11, which compares its performance with LoRA and gradual unfreezing using 10% of the target domain data. As shown in Fig. 11(d), the adapter method achieves the lowest MAPE and a narrower error distribution, indicating superior estimation accuracy and stability under limited-data conditions. This comparison further validates the effectiveness of the proposed strategy in achieving high transfer performance with minimal target domain samples.

4.3. *Interpretability of the proposed model*

Interpretability is crucial for the practical deployment and reliability assurance of data-driven battery health models. In contrast to traditional black-box models that rely on post-hoc interpretability, the DGAT-N-BEATS architecture is designed with intrinsic interpretability built into its computational framework. Specifically, the model leverages attention-based spatial graph learning and polynomial-based temporal decomposition, providing direct insights into critical battery cells, influential features, and underlying degradation patterns without requiring external interpretability approximations. This structural interpretability ensures model transparency, aligns with physical intuition, and facilitates practical diagnostics and decision-making in the BMS of EVs.

At the spatial level, the DGAT learns dynamic interdependencies among battery cells by utilizing attention mechanisms. Fig. 12(a) illustrates the learned attention weight matrix, quantitatively revealing directional influences between pairs of battery cells. The distribution of attention weights clearly indicates that certain nodes have a significant impact on battery performance degradation. Notably, nodes that consistently receive high attention scores across multiple samples suggest that the model has learned to focus on cells that are more critical to the degradation process, which aligns with known mechanisms of non-uniform degradation. This explicit attention-driven graph structure directly encodes interpretable spatial interactions, enabling immediate identification of problematic or critical cells within the battery pack. The attention heatmap shown in Fig. 12(b) further demonstrates feature-level interpretability, highlighting the relative importance of input features at each timestep within a sliding window. Specific features—particularly those related to extreme temperatures and charging indicators—receive higher attention weights. Specifically, the model consistently assigns higher importance to maximum temperature, temperature difference, and the spread between cell-level voltage extremes. Together, these features account for 57.6% of the attention mass, underscoring the sensitivity of the model to thermal and electrical heterogeneity—two well-established factors that drive non-uniform battery aging. Fig. 12(c) visualizes the graph structure learned from the attention module, providing another intuitive perspective on interpretability. Nodes representing battery cells are color-coded based on their computed attention-derived importance scores. Notably, certain nodes exhibit higher centrality, indicating that specific battery cells are more dominant in driving degradation trends.

Interpretability is further enhanced by the polynomial basis

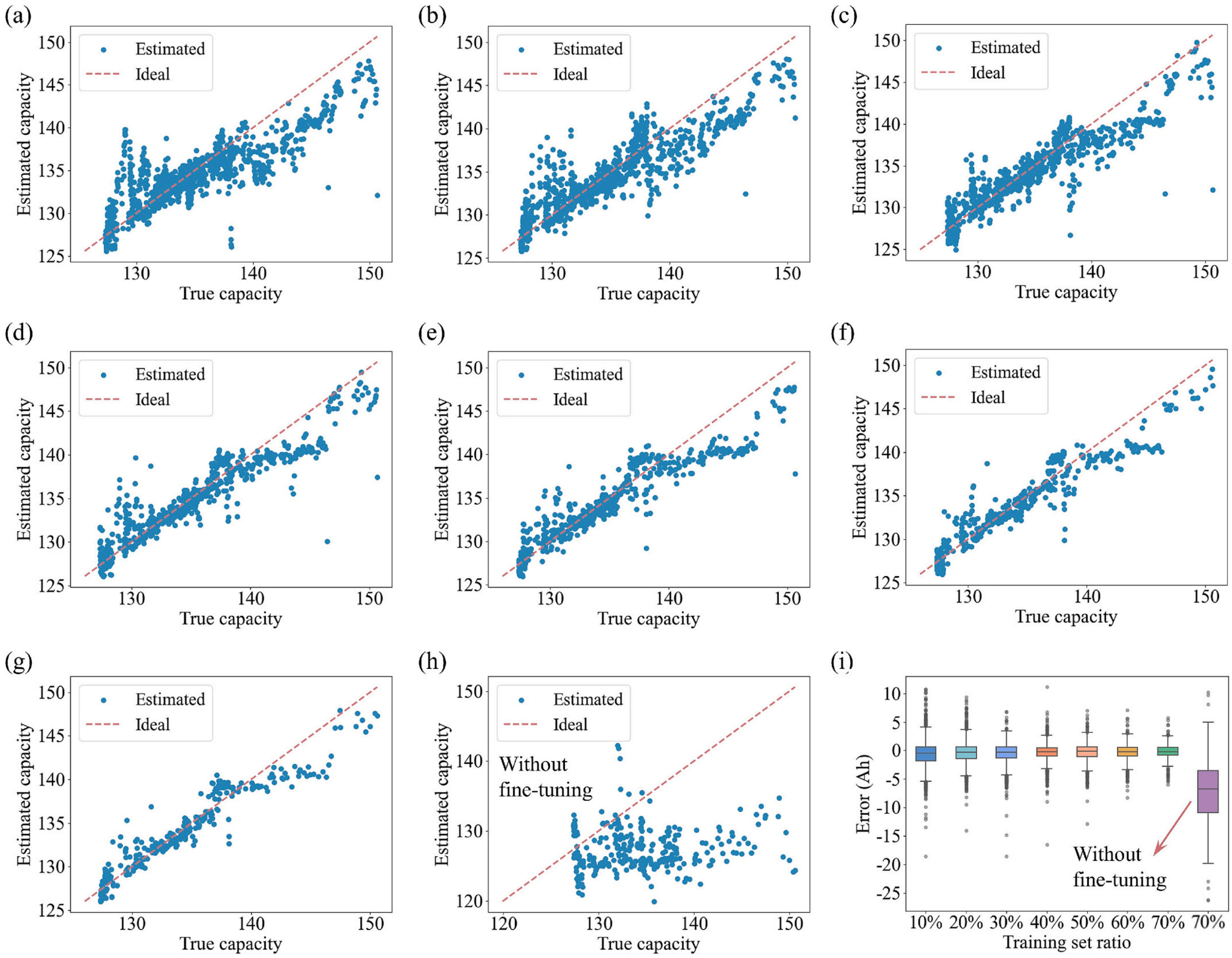

Fig. 10. Evaluation of transfer learning performance using adapter-based fine-tuning on target vehicles.

Table 7
Model performance for capacity estimation using different methods.

Method	RMSE (%)	R^2	MAPE (%)
Adapter fine-tuning (10%)	2.10	0.69	1.43
Adapter fine-tuning (20%)	1.80	0.77	1.25
Adapter fine-tuning (30%)	1.69	0.80	1.11
Adapter fine-tuning (40%)	1.53	0.84	0.98
Adapter fine-tuning (50%)	1.55	0.84	1.04
Adapter fine-tuning (60%)	1.34	0.88	0.90
Adapter fine-tuning (70%)	1.23	0.90	0.84
Without fine-tuning (70%)	6.89	−2.09	5.70

decomposition embedded within the N-BEATS module. Fig. 12(d)-(g) illustrate the transparent temporal modeling strategy of the model, showing distributions and evolutions of polynomial coefficients across multiple samples. Fig. 12(d) shows that the temporal trends are predominantly driven by lower-order polynomial terms (constant and linear), reflecting the inherent monotonic or mildly nonlinear characteristics commonly observed in battery capacity-fade mechanisms, such as solid-electrolyte interphase growth and electrode material degradation. In contrast, the forecast branch shown in Fig. 12(e) incorporates slightly more contributions from higher-order terms, suggesting the ability of the model to capture localized nonlinearities or deviations potentially induced by operational variability, such as irregular charging behaviors or transient events. As illustrated in Fig. 12 (f) and (g), the standard deviation of the linear term across 300 samples is 0.014 for backcast and 0.017 for forecast, indicating highly consistent modeling of global degradation trends. In contrast, the cubic term exhibits greater variability (std = 0.045), suggesting that higher-order terms primarily capture localized deviations or behavioral perturbations. Such distinctions are physically meaningful, as long-term capacity fade is generally smooth, whereas high-frequency fluctuations may reflect noise, external stressors, or short-term user-induced dynamics.

Additionally, the average trend curves reconstructed from multiple stacked N-BEATS blocks, as shown in Fig. 12(h), further demonstrate the interpretability advantage of the proposed method. A decomposition analysis indicates that Block 0 predominantly captures the global monotonic trend (accounting for 72.5% of the overall trend magnitude), whereas Blocks 1 and 2 primarily capture secondary, nonlinear components and localized fluctuations. This block-wise separation reflects the physical hierarchy of aging signals, ranging from slow, irreversible capacity fade to short-term voltage or temperature-induced variations. Analyses of polynomial coefficient distributions and trend decomposition further reinforce model transparency, explicitly aligning with known battery aging mechanisms and operational realities. Importantly, such layer-wise transparency is difficult to obtain from traditional black-box models or post-hoc explanation tools, underscoring the value of embedding interpretability into the model architecture. In summary, by embedding interpretability directly into attention-based spatial graph

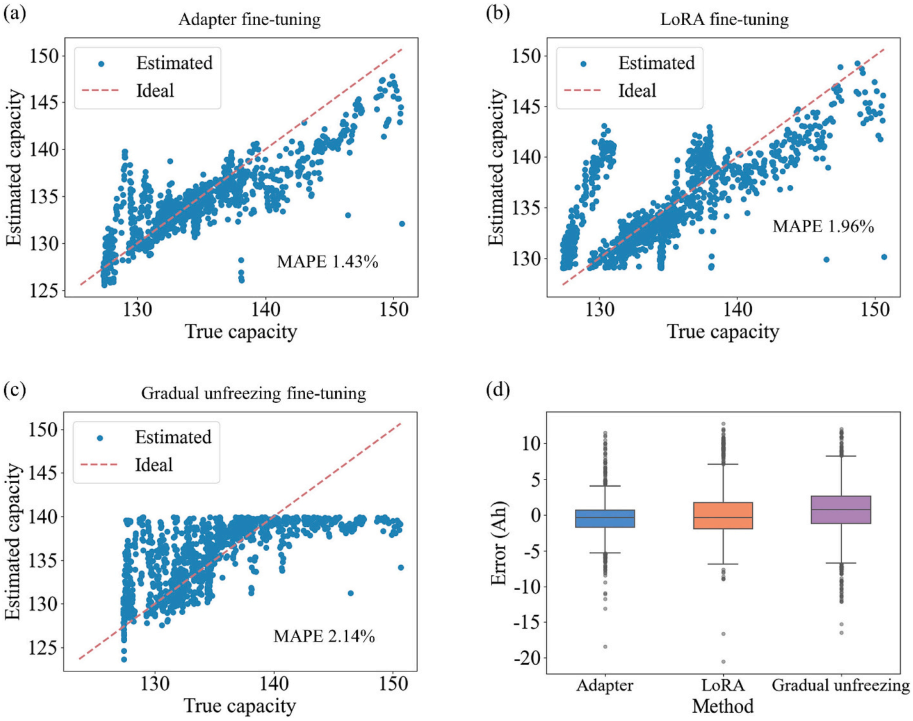

Fig. 11. Performance comparison of the adapter and other fine-tuning strategies with 10% target data.

modeling and polynomial-driven temporal forecasting components, the proposed DGAT-N-BEATS model achieves substantial interpretability advantages over traditional post-hoc interpretability techniques.

4.4. Ablation experiments

To isolate the contribution of each component, we compare five variants trained under the same data split, preprocessing, optimizer, batch size, and early-stopping protocol: Fig. 13(a) the proposed DGAT-N-BEATS, Fig. 13(b) DGAT only, Fig. 13(c) N-BEATS only, Fig. 13(d) static GAT, and Fig. 13(e) GAT-N-BEATS. Fig. 13(a)–(e) reports parity plots of estimated versus true capacity, and Fig. 13(f) summarizes the error distributions. Fig. 13(g)–(h) and Table 8 analyze the effect of the sliding-window length W and the top-k. GPU memory and inference time are measured on the same hardware to assess scalability. DGAT learns and aligns inter-cell relations by recomputing multi-head attention on node features within each window. It then forms a time-varying adjacency matrix by retaining the top-k neighbors for each node. The attention scores function as a content-based relation learner, aligning voltage–temperature dependencies and preserving the most informative couplings at each operating state. In contrast, the GAT baseline uses a fixed adjacency before attention reweighting, which only rescales pre-existing edges and does not infer new connections. This limitation weakens its ability to capture nonstationary thermoelectric interactions, contributing to the accuracy gap. DGAT produces tighter clouds around the identity line and narrower error distributions in Fig. 13(f).

A second benefit of DGAT is improved transfer across vehicles with different physical layouts. Because the graph is reconstructed from features rather than hardwired topology, DGAT can realign node relations when the spatial arrangement or sensor mapping changes. The model therefore preserves key cross-cell dependencies without manual remapping, which supports adaptation to new packs and fleets. This property explains why the full DGAT-N-BEATS variant retains accuracy after transfer, while static GAT variants degrade when the underlying layout differs from the source configuration. The N-BEATS backbone alone (Fig. 13(c)) captures smooth degradation trends but misses spatial interactions, leading to greater dispersion in atypical regimes. GAT-NBEATS (Fig. 13(e)) improves over either part alone, yet the proposed DGAT-N-BEATS (Fig. 13(a)) produces the narrowest cloud around the identity line and the smallest error interquartile range. These results confirm that both temporal trend decomposition and dynamic graph learning are necessary. Fig. 13(g) illustrates the distribution of capacity estimation errors for five different sliding-window lengths. Extending W from 10 to 20 samples reduces both median error and dispersion. This suggests that a longer window can better average short-term fluctuations while still preserving the trend. However, increasing the window beyond 20 samples yields diminishing returns, as the 25- and 30-sample windows offer only marginal accuracy gains but noticeably increase computational cost. A 30-sample window improves RMSE to 0.97% but requires roughly 2959 MB and 52 ms. Balancing accuracy and efficiency, the study adopts a 20-sample window as the default configuration.

Fig. 13(h) illustrates the distribution of capacity estimation errors analyzed under five different top-k settings. Error decreases as the neighborhood expands, since more cross-cell interactions are captured. However, memory and latency grow quickly because attention and message passing scale with the number of active edges. When the top-k value is set to 64, the model demonstrates the highest accuracy, achieving an RMSE of 0.81%, a MAPE of 0.64%, and an R^2 score of 0.94. This setting, however, incurs a computational cost of approximately 5855 megabytes of GPU memory and 74 milliseconds per inference. In comparison, using a top-k value of 16 yields an RMSE of 1.22%, a MAPE of 0.98%, and an R^2 of 0.88. It also significantly reduces memory usage to around 2593 MB and inference time to 43 ms. Given deployment on resource-constrained platforms, the top-k value is set to 16 by default, preserving most of the accuracy benefit while roughly halving compute and memory usage relative to larger neighborhoods. The ablations demonstrate that the dynamic construction in DGAT is essential for modeling nonstationary thermoelectric coupling. The temporal trend decomposition from N-BEATS complements spatial reasoning, and the moderate window length and neighborhood size provide the best compromise between accuracy and efficiency.

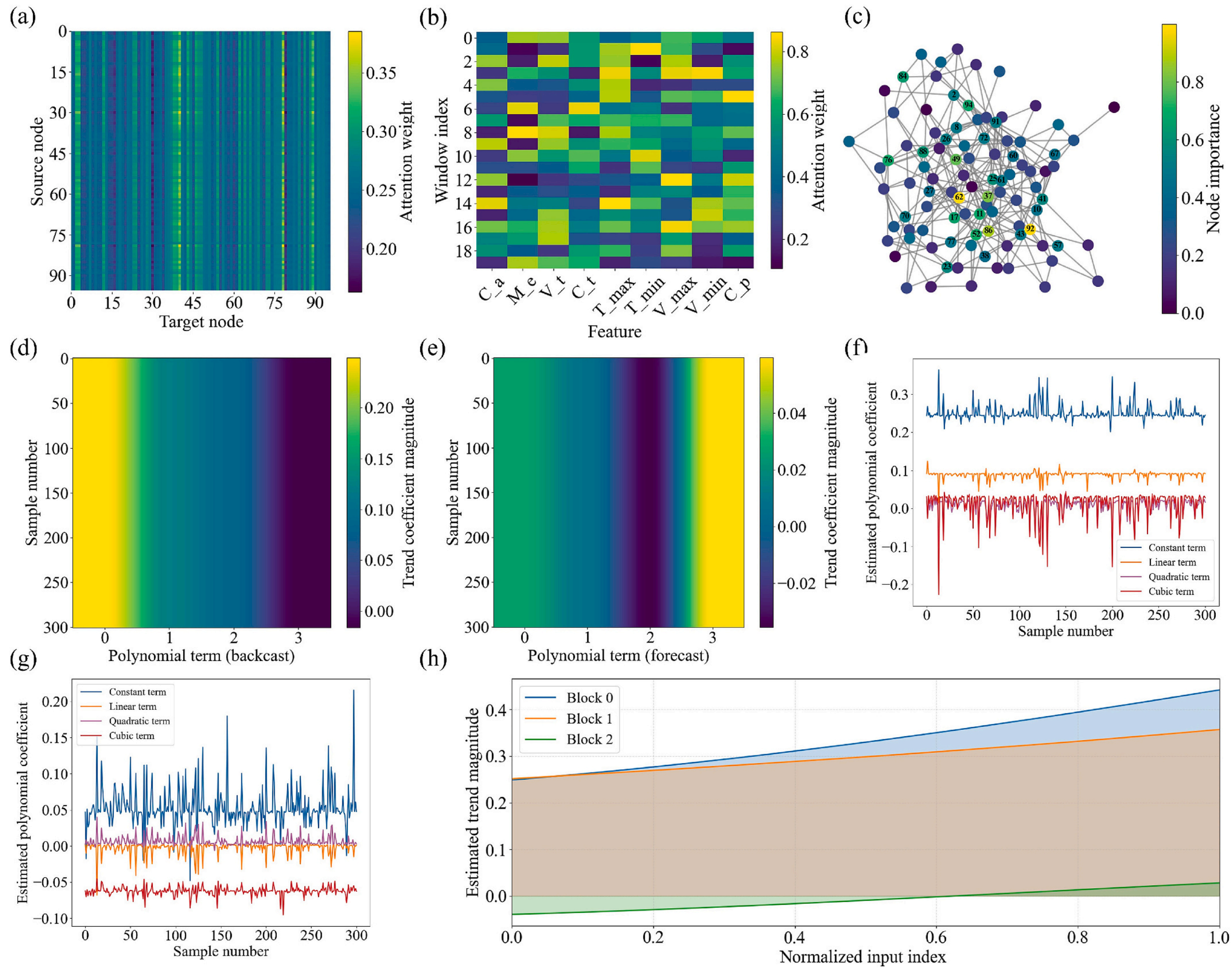

Fig. 12. Interpretability analysis of the proposed DGAT-N-BEATS model for EV battery health diagnostics.

4.5. Behavior-driven degradation and performance analysis

To further evaluate the applicability of the proposed DGAT-N-BEATS model across diverse usage scenarios, the potential impact of differences in user charging behavior on battery capacity degradation and model performance is analyzed using the model's predictions. Specifically, user behavioral profiles are constructed based on vehicle fast-charging ratios. A unified indicator, termed relative charging preference (RCP), represents the dominance of fast versus slow charging habits and is introduced to characterize behavioral tendencies. The RCP is defined by Eq. (19),

$$r = \frac{N_f}{N_f + N_s} \in [0,1], \quad \mathrm{RCP} = 2r - 1 \in [-1,1] \tag{19}$$

where, N_f denotes the number of fast-charging segments, and N_s denotes the number of slow-charging segments for each vehicle. The RCP serves as a symmetric index, representing a monotonic re-parameterization of r that preserves the original ordering and threshold relationships; RCP $>$ 0 indicates fast-charging dominance, while RCP $<$ 0 indicates slow-charging dominance. Fast charging and slow charging are defined based on engineering experience and domain knowledge. Charging segments with a peak power greater than 30 kW are classified as fast charging, whereas segments with a peak power equal to or below 30 kW are classified as slow charging. By incorporating the RCP indicator into the evaluation framework, the proposed method facilitates behavior-aware SOH assessment. It highlights its strong potential to adapt to variations in user charging behavior, thereby enabling personalized battery health management. Fig. 14 presents the main analytical results. As shown in Fig. 14(a), users are categorized into three groups according to their fast-charging ratio: low ($\leq$ 0.3), medium (0.3–0.6), and high ($>$ 0.6). The distribution of capacity fade (ΔQ) demonstrates a clear stratification across these groups. Notably, the median ΔQ in the high group reaches 0.21, significantly greater than that of the Low group (0.14). This is accompanied by more extreme degradation outliers (ΔQ $>$ 0.4), indicating that fast-charging-dominant behavior likely accelerates battery performance deterioration. Fig. 14(b) further reveals a nonlinear relationship between fast-charging ratio and ΔQ. The highest density cluster appears around a fast-charging ratio of approximately 0.9 and ΔQ near 0.2, indicating that while many users frequently engage in fast charging, their degradation levels remain moderate. Nevertheless, localized clusters with ΔQ $>$ 0.4 imply a heightened risk of severe degradation in some cases.

On the other hand, behavior also appears to affect the accuracy of capacity estimation. As illustrated in Fig. 14(c), the high fast-charging group not only exhibits a slightly higher median estimation error, but also a notably elongated error distribution tail with maximum errors exceeding 20%. This suggests that extreme user behaviors may compromise the stability of health estimation models. Fig. 14(d) further

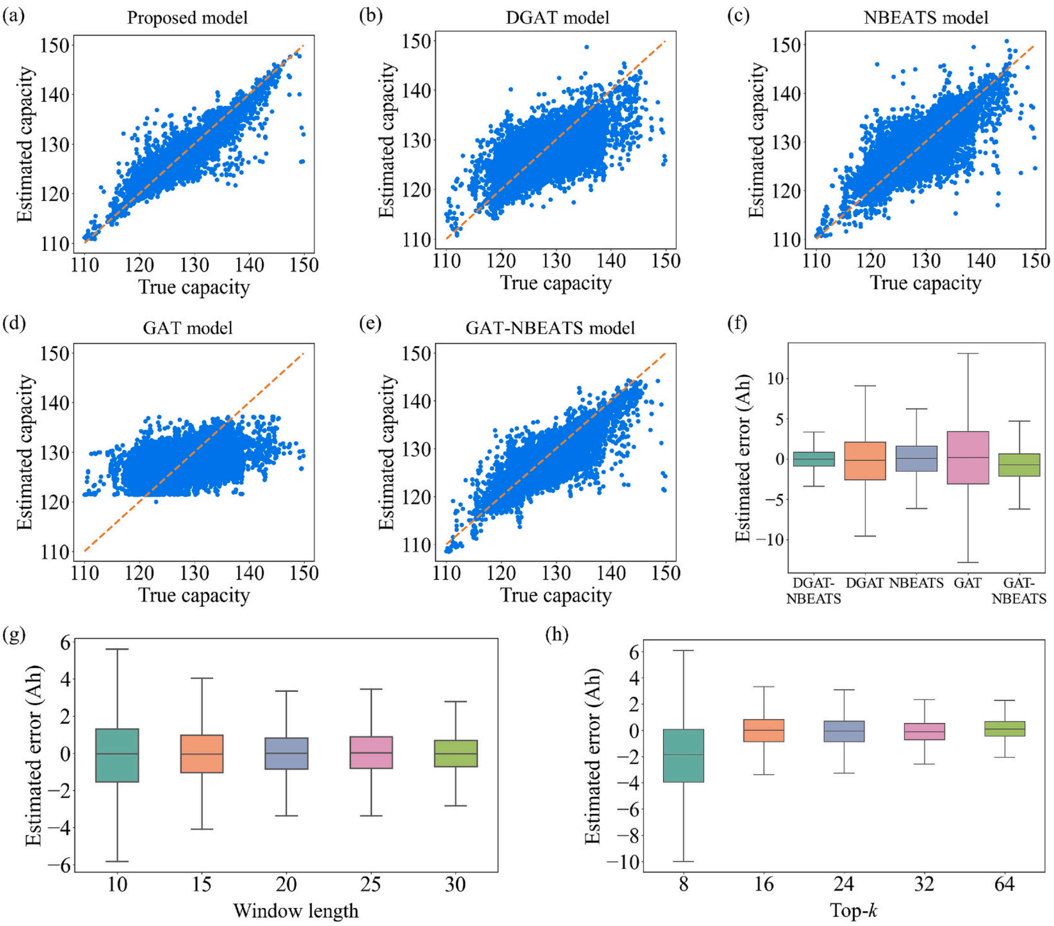

Fig. 13. Component comparisons and effects of sliding-window length and top-*k* on model performance.

Table 8
Effect of window length and graph sparsity on model accuracy and efficiency.

Parameter	Value	RMSE (%)	MAPE (%)	R^2	GPU memory usage (MB)	Inference time (ms)
top-*k*	8	2.24	2.10	0.58	2061	30.56
	16	1.22	0.98	0.88	2593	43.04
	24	1.03	0.88	0.91	3159	54.95
	32	0.94	0.76	0.92	3943	64.06
	64	0.81	0.64	0.94	5855	73.69
W	10	2.05	1.69	0.65	2118	33.88
	15	1.49	1.21	0.82	2364	37.92
	20	1.22	0.98	0.88	2593	43.04
	25	1.14	0.96	0.89	2705	45.91
	30	0.97	0.80	0.92	2959	51.53

highlights a continuous relationship between fast-charging ratio and estimation errors, exhibiting a distinct nonlinear surge in errors beyond a fast-charging ratio of 0.85. Additionally, a regression analysis with RCP as the horizontal axis is performed against ΔQ (Fig. 14(f)), indicating a positive correlation between behavioral dominance and capacity fade. The fitted regression line shows a moderate upward trend, suggesting that as fast-charging preference intensifies, overall degradation risk also increases. However, inter-individual variability significantly influences the upper and lower bounds of the trend. To further quantify this relationship, the mean and standard deviation of ΔQ within different RCP intervals are analyzed. Specifically, users with strong fast-charging preference (RCP∈[0.8, 1.0]) exhibit an average ΔQ of 0.203, which is 38.1% higher than the 0.147 average ΔQ observed in the slow-charging group (RCP∈[−0.5, 0.0]). This highlights the adverse impact of aggressive fast-charging behaviors on long-term battery health.

To test whether the RCP-quantified behavioral cue stabilizes the estimation under fast-charging–dominant usage, an RCP prior is injected into the first input window of N-BEATS at inference only. It is defined by Eq. (20),

$$\widetilde{x} = x + \gamma\alpha(r)\widetilde{\mathbf{K}}x \tag{20}$$

where x denotes the original input sequence, and $\widetilde{x}$ represents the enhanced input with a behavioral prior. The matrix $\widetilde{\mathbf{K}}$ applies a temporal weighting kernel based on exponential decay. The scaling factor $\alpha(r)$ depends on the vehicle's fast-charging preference, activating only for high-RCP users ($r > 0.6$). γ is a global scaling hyperparameter controlling the strength of the behavioral prior. It is fixed across all users. This RCP prior leaves the training data, loss, and parameters unchanged. It is enabled only during evaluation and acts as a controllable robustness knob. Specifically, for users with a high fast-charging preference, the $\widetilde{\mathbf{K}}$ slightly smooths the input window over time before it is processed by N-BEATS. This smoothing reduces spurious sensitivity to local irregularities that occur more frequently under aggressive fast-charging conditions.

Fig. 14(e) compares SOH error across RCP groups with/without the behavior-aware prior. In the high group, the upper-tail error is clearly reduced, while the median changes marginally. Low/Medium groups remain virtually unchanged. This pattern is consistent with our behavioral analysis: high RCP induces short-horizon volatility that inflates tail risk, and a small kernel mixture dampens such volatility without affecting central tendency. The result confirms that RCP is inherently connected to DGAT-N-BEATS via the temporal smoothness of its input

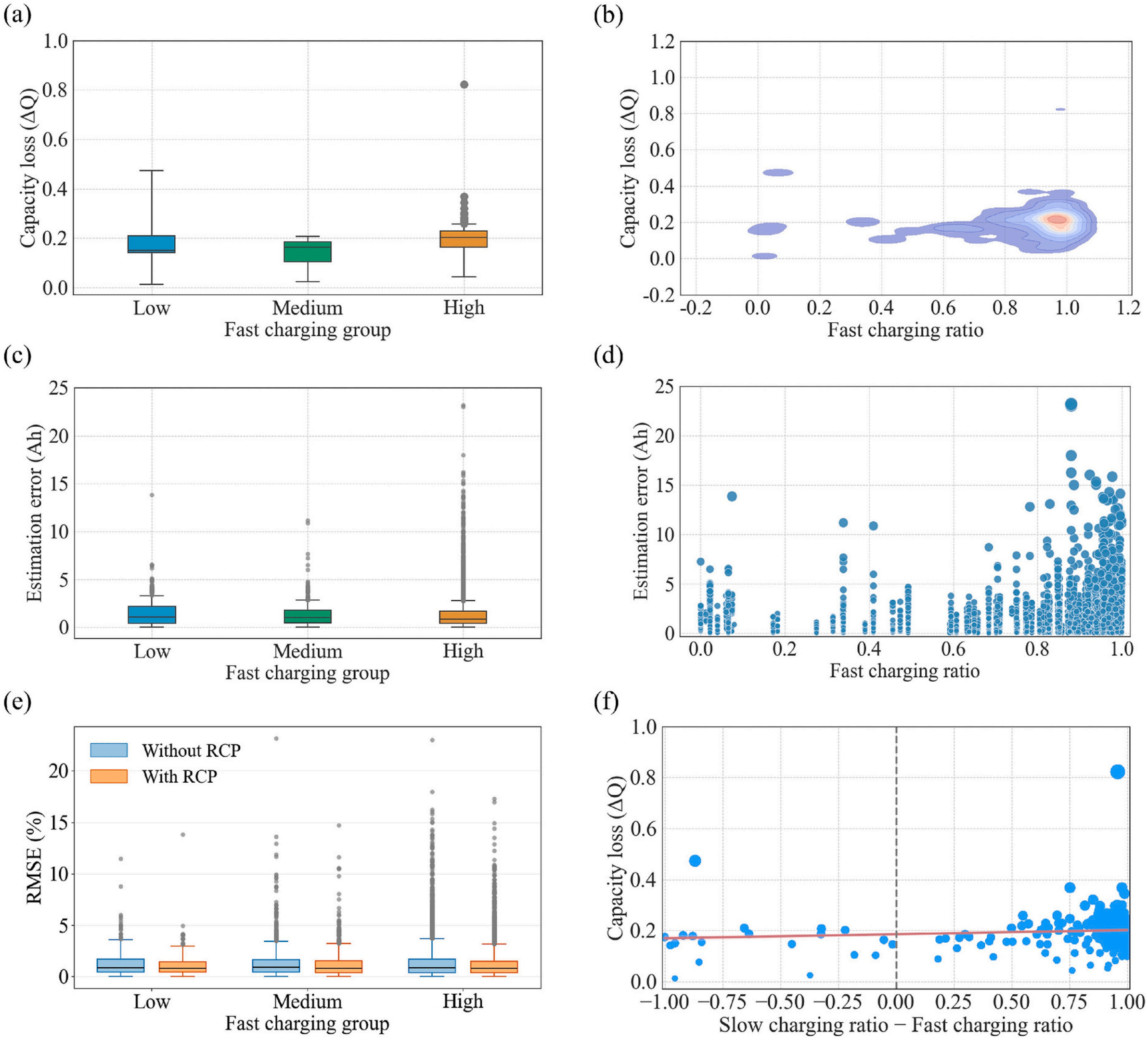

Fig. 14. Analysis of user charging behavior, battery degradation, and model estimation error.

window and demonstrates a deploy-time, plug-and-play way to enhance robustness under extreme user behavior. In summary, the results indicate a significant association between user charging behavior diversity and battery degradation patterns. In particular, users with strong fast-charging preferences face greater challenges in maintaining capacity retention and achieving stable SOH estimation. This highlights the crucial importance of developing health management models tailored to behavioral variability, thereby facilitating personalized battery health management and proactive prediction of degradation risk.

5. Conclusions

This study presents a unified and interpretable framework for EV battery SOH estimation that integrates a DGAT with the N-BEATS forecasting model. The main conclusions and contributions are summarized as follows:

(1) The proposed architecture effectively captures the spatiotemporal evolution of battery degradation by dynamically constructing thermoelectric graphs from cell-level voltage and temperature data and by decomposing long-term SOH trajectories into interpretable polynomial trends. This design overcomes limitations in interpretability and in modeling dynamic inter-cell dependencies.

(2) Extensive validation using operational data from 300 electric vehicles demonstrates that the DGAT-N-BEATS framework significantly outperforms conventional baselines, achieving an RMSE of 1.22%, MAPE of 0.98%, and R^2 of 0.88.

(3) The adapter-based transfer learning strategy confirms strong cross-system adaptability, maintaining reliable accuracy with only 10% of the target-domain data.

(4) The proposed RCP metric reveals the significant impact of user charging behaviors on degradation. Specifically, fast-charging users ($RCP > 0.8$) exhibit up to 38% faster capacity fade, providing quantitative support for behavior-informed maintenance strategies.

(5) The interpretability analysis shows that voltage- and temperature-related features contribute over 50% to the prediction. These findings align closely with known electrochemical degradation mechanisms, validating that the model captures meaningful physical trends.

Future research will extend the proposed framework to a broader range of battery chemistries, including solid-state, sodium-ion, and high-nickel cathode systems, to evaluate its adaptability across next-generation technologies. To facilitate practical deployment, efforts will focus on optimizing the computational efficiency and memory footprint of the framework, enabling real-time inference on embedded battery management hardware. Furthermore, acknowledging that the temperature measured at the battery terminal may deviate from the internal battery state due to aging-induced changes in thermal contact resistance, future work will investigate methods to integrate thermal resistance correction to further improve diagnostic precision. Finally, integration with predictive maintenance and cloud–edge collaborative architectures will be pursued to improve the scalability and application potential of the framework in large-scale BMS.

CRediT authorship contribution statement

Zhilong Lv: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis. **Shiqi (Shawn) Ou:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis. **Hao Jing:** Visualization, Data curation. **Haobo Dong:** Writing – review & editing. **Dapai Shi:** Writing – review & editing. **Jingyuan Zhao:** Writing – review & editing.

Code availability

The code related to the modeling framework is available at https://github.com/Translab-SCUT/Interpretable-Thermoelectric-Graph-SOH-Estimation-Framework or from the corresponding author upon reasonable request.

Declaration of competing interest

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

Acknowledgements

This work is supported by the National Key Research and Development Program of China (2024YFE0115800) and the Introduced Innovative R&D Team of Guangdong (2023ZT10L145). All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of sponsors.

Data availability

Dataset #1 used in this study is publicly available at http://ivstskl.changan.com.cn/?p=2697. Dataset #2 is confidential, and the authors do not have the rights to share it.

References

[1] G. Luderer, S. Madeddu, L. Merfort, F. Ueckerdt, M. Pehl, R. Pietzcker, E. Kriegler, Impact of declining renewable energy costs on electrification in low-emission scenarios, Nat. Energy 7 (1) (2022) 32–42, https://doi.org/10.1038/s41560-021-00937-z.

[2] H.E. Melin, M.A. Rajaeifar, A.Y. Ku, A. Kendall, G. Harper, O. Heidrich, Global implications of the EU battery regulation, Science 373 (6553) (2021) 384–387, https://doi.org/10.1126/science.abh1416.

[3] G. Crabtree, The coming electric vehicle transformation, Science 366 (6464) (2019) 422–424, https://doi.org/10.1126/science.aax0704.

[4] R. Schmuch, R. Wagner, G. Hörpel, T. Placke, M. Winter, Performance and cost of materials for lithium-based rechargeable automotive batteries, Nat. Energy 3 (4) (2018) 267–278, https://doi.org/10.1038/s41560-018-0107-2.

[5] N.A. Sepulveda, J.D. Jenkins, A. Edington, D.S. Mallapragada, R.K. Lester, The design space for long-duration energy storage in decarbonized power systems, Nat. Energy 6 (5) (2021) 506–516, https://doi.org/10.1038/s41560-021-00796-8.

[6] C. Xu, P. Behrens, P. Gasper, K. Smith, M. Hu, A. Tukker, B. Steubing, Electric vehicle batteries alone could satisfy short-term grid storage demand by as early as 2030, Nat. Commun. 14 (1) (2023) 119, https://doi.org/10.1038/s41467-022-35393-0.

[7] Z. Zhu, T. Jiang, M. Ali, Y. Meng, Y. Jin, Y. Cui, W. Chen, Rechargeable batteries for grid scale energy storage, Chem. Rev. 122 (22) (2022) 16610–16751, https://doi.org/10.1021/acs.chemrev.2c00289.

[8] M.R. Palacín, A. de Guibert, Why do batteries fail? Science 351 (6273) (2016) 1253292 https://doi.org/10.1126/science.1253292.

[9] W.M. Dose, W. Li, I. Temprano, C.A. O'Keefe, B.L. Mehdi, M.F. De Volder, C. P. Grey, Onset potential for electrolyte oxidation and Ni-rich cathode degradation in lithium-ion batteries, ACS Energy Lett. 7 (10) (2022) 3524–3530, https://doi.org/10.1021/acsenergylett.2c01722.

[10] J. Zhao, Z. Lv, D. Li, X. Feng, Z. Wang, Y. Wu, A.F. Burke, Battery engineering safety technologies (BEST): mechanisms, modes, metrics, modelling and mitigation, ETransportation (2024) 100364, https://doi.org/10.1016/j.etran.2024.100364.

[11] X. Liu, W. Ai, M.N. Marlow, Y. Patel, B. Wu, The effect of cell-to-cell variations and thermal gradients on the performance and degradation of lithium-ion battery packs, Appl. Energy 248 (2019) 489–499, https://doi.org/10.1016/j.apenergy.2019.04.108.

[12] J.P. Pender, G. Jha, D.H. Youn, J.M. Ziegler, I. Andoni, E.J. Choi, C.B. Mullins, Electrode degradation in lithium-ion batteries, ACS Nano 14 (2) (2020) 1243–1295, https://doi.org/10.1021/acsnano.9b04365.

[13] S. Vignesh, H.S. Che, J. Selvaraj, K.S. Tey, J.W. Lee, H. Shareef, R. Errouissi, State of Health (SoH) estimation methods for second life lithium-ion battery—Review and challenges, Appl. Energy 369 (2024) 123542, https://doi.org/10.1016/j.apenergy.2024.123542.

[14] Y. Tian, C. Lin, H. Li, J. Du, R. Xiong, Detecting undesired lithium plating on anodes for lithium-ion batteries–a review on the in-situ methods, Appl. Energy 300 (2021) 117386, https://doi.org/10.1016/j.apenergy.2021.117386.

[15] W. Liu, P. Liu, D. Mitlin, Review of emerging concepts in SEI analysis and artificial SEI membranes for lithium, sodium, and potassium metal battery anodes, Adv. Energy Mater. 10 (43) (2020) 2002297, https://doi.org/10.1002/aenm.202002297.

[16] W. Duan, S. Song, F. Xiao, Y. Chen, S. Peng, C. Song, Battery SOH estimation and RUL prediction framework based on variable forgetting factor online sequential extreme learning machine and particle filter, J. Energy Storage 65 (2023) 107322, https://doi.org/10.1016/j.est.2023.107322.

[17] S. Khaleghi, D. Karimi, S.H. Beheshti, M.S. Hosen, H. Behi, M. Berecibar, J. Van Mierlo, Online health diagnosis of lithium-ion batteries based on nonlinear autoregressive neural network, Appl. Energy 282 (2021) 116159, https://doi.org/10.1016/j.apenergy.2020.116159.

[18] A. Thelen, Y.H. Lui, S. Shen, S. Laflamme, S. Hu, H. Ye, C. Hu, Integrating physics-based modeling and machine learning for degradation diagnostics of lithium-ion batteries, Energy Storage Mater. 50 (2022) 668–695, https://doi.org/10.1016/j.ensm.2022.05.047.

[19] S. Yang, C. Zhang, J. Jiang, W. Zhang, L. Zhang, Y. Wang, Review on state-of-health of lithium-ion batteries: characterizations, estimations and applications, J. Clean. Prod. 314 (2021) 128015, https://doi.org/10.1016/j.jclepro.2021.128015.

[20] J. Lu, R. Xiong, J. Tian, C. Wang, C.W. Hsu, N.T. Tsou, J. Li, Battery degradation prediction against uncertain future conditions with recurrent neural network enabled deep learning, Energy Storage Mater. 50 (2022) 139–151, https://doi.org/10.1016/j.ensm.2022.05.007.

[21] H. Li, X. Xie, X. Zhang, A.F. Burke, J. Zhao, Battery state estimation for electric vehicles: translating AI innovations into real-world solutions, J. Energy Storage 115 (2025) 116000, https://doi.org/10.1016/j.est.2025.116000.

[22] Z. Deng, L. Xu, H. Liu, X. Hu, Z. Duan, Y. Xu, Prognostics of battery capacity based on charging data and data-driven methods for on-road vehicles, Appl. Energy 339 (2023) 120954, https://doi.org/10.1016/j.apenergy.2023.120954.

[23] Z. Lv, J. Zhao, Resource-efficient artificial intelligence for battery capacity estimation using convolutional FlashAttention fusion networks, eTransportation 23 (2025) 100383, https://doi.org/10.1016/j.etran.2024.100383.

[24] J. Tian, X. Liu, S. Li, Z. Wei, X. Zhang, G. Xiao, P. Wang, Lithium-ion battery health estimation with real-world data for electric vehicles, Energy 270 (2023) 126855, https://doi.org/10.1016/j.energy.2023.126855.

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

[26] Z. Wang, F. Yang, Q. Xu, Y. Wang, H. Yan, M. Xie, Capacity estimation of lithium-ion batteries based on data aggregation and feature fusion via graph neural network, Appl. Energy 336 (2023) 120808, https://doi.org/10.1016/j.apenergy.2023.120808.

[27] X.Y. Yao, G. Chen, M. Pecht, B. Chen, A novel graph-based framework for state of health prediction of lithium-ion battery, J. Energy Storage 58 (2023) 106437, https://doi.org/10.1016/j.est.2022.106437.

[28] H. Liu, C. Li, X. Hu, J. Li, K. Zhang, Y. Xie, Z. Song, Multi-modal framework for battery state of health evaluation using open-source electric vehicle data, Nat. Commun. 16 (1) (2025) 1137, https://doi.org/10.1038/s41467-025-56485-7.

[29] L. Shen, J. Li, L. Meng, L. Zhu, H.T. Shen, Transfer learning-based state of charge and state of health estimation for Li-ion batteries: a review, IEEE Trans. Transp. Electrif. 10 (1) (2023) 1465–1481, https://doi.org/10.1109/TTE.2023.3293551.

[30] Q. Wang, M. Ye, X. Cai, D.U. Sauer, W. Li, Transferable data-driven capacity estimation for lithium-ion batteries with deep learning: a case study from laboratory to field applications, Appl. Energy 350 (2023) 121747, https://doi.org/10.1016/j.apenergy.2023.121747.

[31] J. Zhao, Z. Wang, Specialized convolutional transformer networks for estimating battery health via transfer learning, Energy Storage Mater. 71 (2024) 103668, https://doi.org/10.1016/j.ensm.2024.103668.

[32] H. Liu, Z. Deng, Y. Che, L. Xu, B. Wang, Z. Wang, X. Hu, Big field data-driven battery pack health estimation for electric vehicles: a deep-fusion transfer learning

approach, Mech. Syst. Signal Process. 218 (2024) 111585, https://doi.org/10.1016/j.ymssp.2024.111585.

[33] Y.Y. Soo, Y. Wang, H. Xiang, Z. Chen, A novel transfer learning model for battery state of health prediction based on driving behavior classification, J. Energy Storage 111 (2025) 115409, https://doi.org/10.1016/j.est.2025.115409.

[34] Y. Fan, Z. Huang, H. Li, M.B. Kaleem, Y. Wu, State-of-health estimation for battery packs of real-world electric vehicles with cell-to-pack transfer learning, Energy (2025) 138328, https://doi.org/10.1016/j.energy.2025.138328.

[35] I. Jarraya, S.B. Atitallah, F. Alahmed, M. Abdelkader, M. Driss, F. Abdelhadi, A. Koubaa, SOH-KLSTM: A hybrid Kolmogorov-Arnold network and LSTM model for enhanced lithium-ion battery health monitoring, J. Energy Storage 122 (2025) 116541, https://doi.org/10.1016/j.est.2025.116541.

[36] A. Tang, Y. Xu, Y. Hu, J. Tian, Y. Nie, F. Yan, Q. Yu, Battery state of health estimation under dynamic operations with physics-driven deep learning, Appl. Energy 370 (2024) 123632, https://doi.org/10.1016/j.apenergy.2024.123632.

[37] H. Tu, S. Moura, Y. Wang, H. Fang, Integrating physics-based modeling with machine learning for lithium-ion batteries, Appl. Energy 329 (2023) 120289, https://doi.org/10.1016/j.apenergy.2022.120289.

[38] K. Zhang, Z. Jiao, J. Ma, X. Zhao, S. Wang, M. Yu, S. Li, SOH estimation of battery based on the BIGRU-transformer algorithm considering the hierarchical voltage interval characterization of the discharge curve, J. Energy Storage 132 (2025) 117770, https://doi.org/10.1016/j.est.2025.117770.

[39] Y. Lu, D. Guo, G. Xiong, Y. Wei, J. Zhang, Y. Wang, M. Ouyang, Towards real-world state of health estimation: part 2, system level method using electric vehicle field data, eTransportation 22 (2024) 100361, https://doi.org/10.1016/j.etran.2024.100361.

[40] R. Sun, J. Chen, C. Piao, Battery health features extraction and state of health estimation based on real-time online vehicle driving data, J. Power Sources 645 (2025) 236784, https://doi.org/10.1016/j.jpowsour.2025.236784.

[41] S. Wen, N. Lin, S. Huang, X. Li, Z. Wang, Z. Zhang, Lithium battery state of health estimation using real-world vehicle data and an interpretable hybrid framework, J. Energy Storage 96 (2024) 112623, https://doi.org/10.1016/j.est.2024.112623.

[42] A.H. Tan, D.S. Ong, M. Foo, Estimation of battery internal resistance using built-in self-scaling method, J. Energy Storage 59 (2023) 106481, https://doi.org/10.1016/j.est.2022.106481.

[43] Y. Liu, L. Wang, D. Li, K. Wang, State-of-health estimation of lithium-ion batteries based on electrochemical impedance spectroscopy: a review, Prot. Control Mod. Power Syst. 8 (3) (2023) 1–17, https://doi.org/10.1186/s41601-023-00314-w.

[44] S. Wen, N. Lin, S. Huang, X. Li, Z. Wang, Z. Zhang, Lithium battery state of health estimation using real-world vehicle data and an interpretable hybrid framework, J. Energy Storage 96 (2024) 112623, https://doi.org/10.1016/j.est.2024.112623.

[45] F. Xia, Y. Yu, J. Chen, SOH and RUL prediction of lithium batteries based on fusions of RLOESS filtered electrochemical and thermal features by bidirectional gated recurrent unit network, J. Energy Storage 102 (2024) 114134, https://doi.org/10.1016/j.est.2024.114134.

[46] H. Jing, J. Hu, S.S. Ou, Z. Lv, R. Lyu, J. Zhao, Scalable and generalizable deep learning for battery state of health estimation in on-road electric vehicles, J. Energy Chem. 110 (2025) 823–841, https://doi.org/10.1016/j.jechem.2025.07.020.

[47] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph Attention Networks, arXiv preprint arXiv:1710.10903., 2017, https://doi.org/10.48550/arxiv.1710.10903.

[48] B.N. Oreshkin, D. Carpov, N. Chapados, Y. Bengio, N-BEATS: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting, arXiv preprint arXiv:1905.10437., 2019, https://doi.org/10.48550/arxiv.1905.10437.

[49] D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, arXiv preprint arXiv:1412.6980., 2014, https://doi.org/10.48550/arXiv.1412.6980.

[50] N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. De Laroussilhe, A. Gesmundo, S. Gelly, Parameter-efficient transfer learning for NLP, in: International Conference on Machine Learning, PMLR, 2019, May, pp. 2790–2799. https://arxiv.org/abs/1902.00751.

[51] E.J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, W. Chen, Lora: Low-rank adaptation of large language models, ICLR 1 (2) (2022) 3, https://doi.org/10.48550/arXiv.2106.09685.

[52] J. Howard, S. Ruder, Universal Language Model Fine-tuning for Text Classification, arXiv preprint arXiv:1801.06146, 2018, https://doi.org/10.48550/arXiv.1801.06146.