POLITECNICO DI TORINO
Repository ISTITUZIONALE

Physics-Informed Temporal Convolutional Networks for State-of-Health Battery Estimation of Real-world Electric Vehicles

Original
Physics-Informed Temporal Convolutional Networks for State-of-Health Battery Estimation of Real-world Electric Vehicles / Nepi, M., Gallo, R., Zampolli, M., Tosco, P., Aliberti, A., Patti, E.. - (In corso di stampa). (22st IEEE International Conference on Intelligent Computer Communication and Processing (ICCP) Cluj-Napoca, Romania 22-24 October 2026).

Availability:
This version is available at: 11583/3017835 since: 2026-11-12T11:21:32Z

Publisher:
IEEE

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©9999 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

19 November 2026

Physics-Informed Temporal Convolutional Networks for State-of-Health Battery Estimation of Real-world Electric Vehicles

Mattia Nepi‡, Raimondo Gallo*, Marco Zampolli‡, Paolo Tosco‡, Alessandro Aliberti† and Edoardo Patti*
*Dept. of Control and Computer Engineering, Politecnico di Torino, Italy. Email: name.surname@polito.it
†Interuniversity Dept. of Regional and Urban Studies and Planning, Politecnico di Torino, Italy. Email: name.surname@polito.it
‡Edison S.p.A., Milano, Italy. Email: name.surname@edison.it

***Abstract*—Accurate State-of-Health (SOH) estimation for lithium-ion batteries in electric vehicles (EVs) under real-world driving conditions is challenged by fleet heterogeneity and the scarcity of labeled degradation data. Purely data-driven models are prone to shortcut learning, failing to generalize to batteries with nominal capacities unseen during training, i.e., the Out-of-Distribution (OOD) scenario. This paper proposes a two-stage hybrid framework based on Temporal Convolutional Networks (TCN). In the first Stage A, the model learns degradation dynamics from a simulator using capacity-agnostic features (C-rate, normalized voltage, temperature, and SOC). In the second Stage B, Domain Adaptation on heterogeneous real-world data is performed via Partial Freezing and a Physics-Informed loss function (PINN loss) that penalizes non-physical temporal jumps in the SOH estimate. Experiments on heterogeneous real-world EV battery datasets show that the proposed approach mitigates capacity-related overfitting, enabling stable real-time onboard estimation and yielding temporally smoother SOH estimates than purely data-driven baselines, and that it generalizes to a battery pack of unseen nominal capacity in on-road tests.**

***Index Terms*—State of Health, lithium-ion battery, temporal convolutional network, physics-informed neural network, transfer learning, domain adaptation, electric vehicles, Battery Management System**

I. Introduction

The widespread adoption of electric vehicles (EVs) has placed battery management systems (BMS) at the center of safety, efficiency, and long-term reliability concerns. Among all the quantities monitored by a BMS, the State-of-Health (SOH) is arguably the most critical: it quantifies the residual capacity of a lithium-ion cell relative to its nominal value, and its accurate estimation directly governs decisions on range prediction, warranty management, and second-life applications [1]. Despite considerable advances in battery diagnostics, computing SOH on-board in real time remains far more complex than in controlled laboratory conditions. Real driving profiles are characterized by rapid, high-amplitude current transients, aggressive accelerations, regenerative braking events, and thermal fluctuations, that push the battery far from the quasi-static assumptions underlying most electrochemical models [2], [3].

The problem becomes significantly harder when a single model must generalize across heterogeneous fleets of EVs in terms of manufacturers, nominal capacities and diverse battery chemistries. A BMS algorithm designed and validated on a battery with a 40 kWh pack cannot be naively deployed on a 78 kWh pack without retraining. Most data-driven SOH estimators implicitly encode the nominal capacity as a shortcut feature, leading to systematic biases and poor scalability across fleets. This Out-of-Distribution (OOD) failure is compounded by the fact that collecting and labeling real-world degradation data for every possible vehicle configuration is economically prohibitive in production fleets [4].

To address these limitations, this paper proposes a two-stage hybrid framework. In Stage A, a TCN is pre-trained on a physics-based simulator of a single vehicle battery pack using exclusively capacity-agnostic input features, C-rate (C_{rate}) and normalized voltage (V_{norm}) together with temperature and SOC, to prevent the network from anchoring its predictions on vehicle-specific parameters. In Stage B, the pre-trained model undergoes Domain Adaptation on real heterogeneous driving data via Partial Freezing of the convolutional layers and a Physics-Informed Neural Network (PINN) loss that encodes the slow electrochemical inertia of battery degradation without external post-processing filters.

The main contributions of this work are summarized as follows:

- A capacity-agnostic input representation (C_{rate} and V_{norm}) that prevents the model from exploiting absolute current and voltage magnitudes as a shortcut for pack size.
- A two-stage simulation-to-real transfer scheme for heterogeneous fleets, in which established partial-freezing strategies [5], [15] are combined with a temporal physics-informed regularization term [6], [7] to preserve source-domain representations while stabilizing on-board estimates.
- An experimental evaluation on five real EVs with nominal capacities from 33 to 78 kWh, complemented by on-road tests on two vehicles never seen during training.

The remainder of the paper is organized as follows. Section II reviews related work on data-driven SOH estimation, transfer learning for batteries, and physics-informed neu-

ral networks. Section III details the dataset, preprocessing pipeline, and model architecture. Section IV presents experimental results and discussion. Section V concludes the paper.

II. Related Works

Data-driven methods for battery SOH estimation from time-series measurements traditionally relied on Long Short-Term Memory (LSTM) networks, as demonstrated by Li et al. [8] for capturing aging dynamics. As an alternative to recurrent architectures, Bai et al. [9] showed that Temporal Convolutional Networks (TCN) can outperform recurrent networks on a wide range of sequence modeling tasks. Subsequently, Zhang et al. [10] successfully applied TCNs to battery diagnostics using energy-based features, showing superior performance on highly dynamic load profiles. Despite these architectural advances, purely data-driven models suffer from shortcut learning: as highlighted by Weiss et al. [11] and Wang et al. [12], when trained on a fixed nominal capacity and tested on different packs, prediction error increases sharply, exposing a severe Out-of-Distribution (OOD) vulnerability.

To address the scarcity of real-world data and bridge this domain gap, Chen et al. [13] proposed a transfer learning framework adapting laboratory pre-trained models to field conditions, while Zhang et al. [14] extended this concept to generalized multi-source frameworks. A critical challenge during this domain adaptation is catastrophic forgetting, as source-domain knowledge is overwritten [15]. To prevent this, continual learning strategies have proven highly effective [16], and freezing selected layers during fine-tuning is a well-established way to limit forgetting [5]. Consequently, our framework adopts Partial Freezing of the lower TCN blocks to preserve the representations learned on simulated data, allowing only the output head to adapt to the target fleet.

Finally, embedding physical knowledge acts as a powerful regularizer in low-data regimes. Ruan et al. [17] pioneered this approach for batteries by incorporating capacity fade differential equations as soft constraints. Focusing on broader applications, Li et al. [18] highlighted how physics-informed machine learning consistently outperforms purely data-driven models in forecasting tasks. Temporal consistency constraints on degradation estimates have also been explored in [6], [7]. Here we adopt a similar temporal regularizer and study its effect in a cross-capacity transfer setting on heterogeneous fleets, where it stabilizes the network output without relying on post-hoc filtering.

A comparative summary of these representative state-of-the-art approaches is provided in Table I. As highlighted, our methodology combines a TCN architecture, Transfer Learning, and a PINN loss, and explicitly evaluates cross-capacity generalization.

TABLE I
Comparison of representative SOH estimation approaches

Method	Arch.	TL	PINN	OOD
Li et al. [8]	LSTM	×	×	×
Zhang et al. [10]	TCN	×	×	×
Chen et al. [13]	CNN	✓	×	Partial
Ruan et al. [17]	PINN	×	✓	×
Zhao et al. [7]	TCN	×	✓	×
Proposed	TCN	✓	✓	✓*

TL = Transfer Learning; OOD = Out-of-Distribution generalization.
*Evaluated on two on-road vehicles unseen during training.

III. Materials and Methods

The proposed framework is illustrated in Fig. 1. It consists of a two-stage training pipeline operating on two distinct data domains, described in the following subsections.

A. Dataset and Physics-Informed Preprocessing

The development of the proposed capacity-agnostic framework relies on two distinct datasets: a realistic simulated dataset acting as Source Domain and a real-world heterogeneous dataset acting as Target Domain. Crucially, both datasets record the same multivariate time-series of physical quantities: pack voltage (V), current (I), internal temperature (T), and State of Charge (SOC).

The **Source Domain** dataset is generated using a high-fidelity electrochemical-thermal simulator parameterized for a 35.8 kWh lithium-ion battery pack (B_{sim}). The simulator produces pristine degradation trajectories spanning SOH values from 100% to 80% under standardized driving cycles. The key advantage of this synthetic dataset is the complete absence of sensor noise and the availability of continuous, ground-truth SOH labels at every time step, providing an ideal environment for the network to learn the fundamental dynamics of cell aging. The details of the simulator are described in [19].

The **Target Domain** dataset comprises real-world driving sessions collected from a heterogeneous EV fleet of five vehicles with different battery chemistries and nominal capacities (B_1–B_5, ranging from 33 to 78 kWh). The real-world data were acquired with proprietary fleet monitoring systems. These data present severe challenges: highly dynamic load profiles, sensor measurement noise, and missing data segments. Furthermore, the ground-truth SOH labels for the Target Domain were provided by a specialized third-party battery diagnostic company. These labels are derived from certified physical stress tests (full discharge cycles), ensuring a reliable baseline for training and validation, but do not cover all the discrete classes between 80% and 100% (where an SOH "class" is defined as a 1% resolution interval, e.g., 85%, 86%), resulting in a sparse labeling problem. The real-world dataset was provided by an industrial partner under a non-disclosure agreement and cannot be publicly released; the simulator configuration and the preprocessing and training code are available from the authors upon reasonable request.

To rigorously evaluate the framework, the Target Domain dataset was partitioned into training (70%), validation (15%), and testing (15%) sets. To prevent temporal data leakage, this split was strictly session-based, ensuring that samples from the same driving session are never shared across different

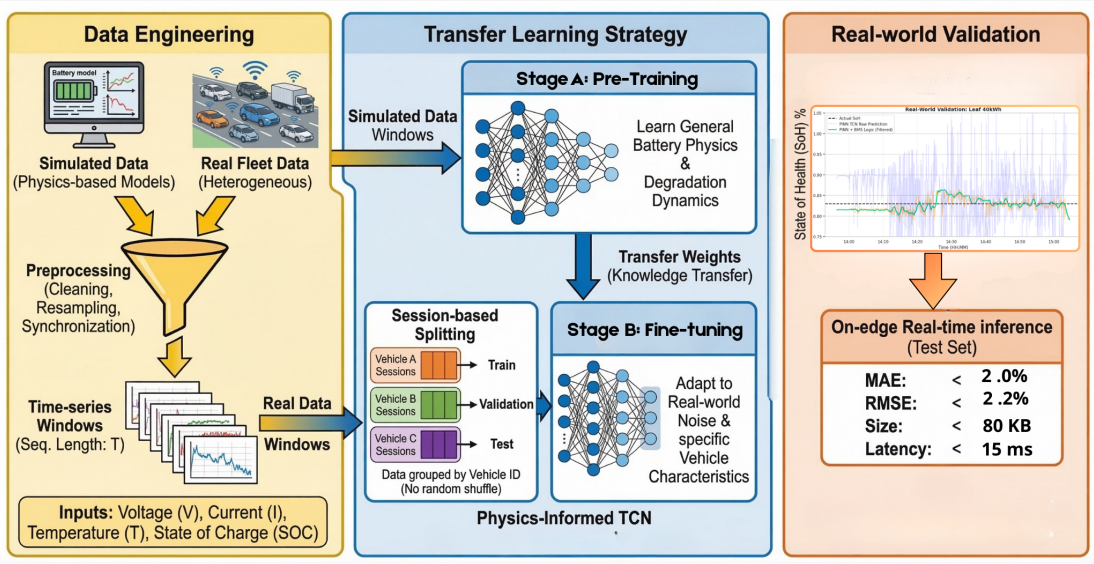


Fig. 1. Overview of the proposed two-stage hybrid framework. Stage A pre-trains the TCN on simulated data using capacity-agnostic features. Stage B performs Domain Adaptation on real heterogeneous EV data via Partial Freezing and the temporal PINN loss.

sets. Only partial discharge sessions lasting at least 4 min were retained, since shorter sessions do not provide enough context for a window-based estimate.

Additionally, a dedicated **On-field test set** was acquired through controlled on-road data collection sessions performed by the authors on two vehicles strictly unseen by the model during training: a 40 kWh pack ($B_{OOD,1}$) and a 62 kWh pack ($B_{OOD,2}$). These two vehicles share the same vehicle platform and battery cell chemistry but differ in nominal capacity by 55%. The nominal capacity of $B_{OOD,1}$ coincides with that of a training vehicle (B_2), although the vehicle itself and its degradation state were never observed, whereas the nominal capacity of $B_{OOD,2}$ does not appear in the training fleet. Both vehicles share the platform and cell chemistry of B_2; the OOD test therefore isolates the effect of unseen vehicles and pack capacities rather than of unseen chemistries.

As depicted in the *Data Pre-processing* block of Fig. 1, the raw data undergoes a rigorous preparation pipeline. First, to harmonize the heterogeneous logging frequencies and handle sensor anomalies, all time-series are cleaned, interpolated for small missing segments, and resampled to a strict 1 Hz resolution. Subsequently, to enable cross-capacity generalization and prevent shortcut learning—where the model infers SOH from the absolute magnitude of current or voltage rather than the normalized impedance response [11]—a physics-informed preprocessing scheme is applied. The raw inputs $I(t)$ and $V(t)$ are converted into capacity-agnostic quantities:

$$C_{rate}(t) = \frac{I(t)}{C_{nominal}} \tag{1}$$

$$V_{norm}(t) = \frac{V(t)}{V_{nominal}} \tag{2}$$

where $C_{nominal}$ is the nominal pack capacity [Ah] and $V_{nominal}$ is the nominal pack voltage [V]. This normalization masks vehicle identity: a 100 A discharge corresponds to approximately 0.9C for a 40 kWh (≈115 Ah) pack but only approximately 0.45C for a 78 kWh (≈220 Ah) pack. Note that $C_{nominal}$ and $V_{nominal}$ are nameplate values that are always available to the BMS; they are used only to normalize the inputs and are never fed to the network as features, so the model cannot condition its prediction on pack size.

Battery temperature T and State of Charge SOC are included as additional inputs without capacity-based rescaling: SOC is already a dimensionless relative quantity defined on the [0%–100%] interval, independent of pack size; T is an absolute thermodynamic scalar whose range is governed by ambient and thermal management conditions, not by pack capacity. The resulting input feature vector at any given time step t is therefore $\mathbf{x}(t) = [C_{rate}, V_{norm}, T, \text{SOC}]^\top$. The fleet specifications and capacities are summarized in Table II.

TABLE II
DATASET CHARACTERISTICS (TARGET FLEET VEHICLE IDENTITIES ANONYMIZED)

ID	Domain	Cap. (kWh)	SOH range
B_{sim}	Source	35.8	80–100%
B_1	Target	35.8	85–99%
B_2	Target	40.0	91–92%
B_3	Target	58.0	87–93%
B_4	Target	78.0	87–91%
B_5	Target	33.2	98–100%
$B_{OOD,1}$	OOD Test	40.0	83%
$B_{OOD,2}$	OOD Test	62.0	92%

OOD = Out-of-Distribution; never seen during training.

TABLE III
WIDE-TCN ARCHITECTURE (INPUT: 30 TIME STEPS × 4 FEATURES)

Layer	Kernel	Dilation	Filters / Units	Params	Stage B
Residual block 1	3	1	64–128		frozen
Residual block 2	3	2	64–128	≈29,000	frozen
Residual block 3	3	4	64–128		frozen
Global Avg. Pooling	–	–	–	0	–
Dense (ReLU)	–	–	64	≈8,200	trainable
Dense (linear)	–	–	1		trainable
Total				≈37,000	

B. TCN Architecture and Pre-Training (Stage A)

The core sequence modeling architecture employed in this work is a Temporal Convolutional Network (TCN). Unlike LSTMs, which process data sequentially and are prone to vanishing gradients over long sequences, TCNs exploit 1D causal convolutions with exponentially increasing dilation factors, maintaining a large receptive field while guaranteeing strictly causal inference [9]. This property is essential for real-time onboard deployment, where future measurements are unavailable by definition.

The proposed architecture, denoted as Wide-TCN to emphasize its reliance on a high number of parallel filters per layer rather than a deep sequential structure, consists of three stacked residual blocks. Each block applies a 1D causal convolution with kernel size $k = 3$ and a broad filter configuration (64 to 128 filters), followed by weight normalization and ReLU activation. The dilation rate d increases exponentially across blocks ($d = 1, 2, 4$), enabling the network to capture both rapid high-frequency transients (e.g., sudden regenerative braking) and slower thermal diffusion effects within the observation window ($SEQ_LEN = 30\,\text{s}$). The convolutional output is aggregated via Global Average Pooling, followed by a fully connected layer (64 units, ReLU) and a linear output neuron predicting the SOH. The total number of trainable parameters is approximately 37,000, confirming the suitability of the model for deployment on automotive-grade embedded hardware. The architecture is depicted in Fig. 2 and detailed in Table III.

In **Stage A**, the Wide-TCN is pre-trained exclusively on the B_{sim} simulated dataset. The network learns to map the capacity-agnostic feature vector $\mathbf{x}(t) = [C_{rate}, V_{norm}, T, \text{SOC}]^\top$ to the simulated ground-truth SOH, using an Adam optimizer with an initial learning rate of 10^{-3} and a cosine annealing schedule. This stage is intended to initialize the convolutional filters as feature extractors for the dynamic voltage response, before any exposure to real-world noise.

C. Domain Adaptation via Partial Freezing and Temporal PINN Loss (Stage B)

While Stage A yields a model trained on physically consistent simulated data, directly deploying it on real EV data leads to suboptimal performance due to the residual domain gap (unmodeled electrochemical dynamics, sensor inaccuracies, and pack-level heterogeneity). Stage B closes this gap through Domain Adaptation on the heterogeneous real-world dataset (i.e., B_1–B_5 in Table II).

To prevent catastrophic forgetting of the representations learned in Stage A, a *Partial Freezing* strategy is adopted in **Stage B**: the weights of the three convolutional blocks (approximately 29,000 parameters, ≈80% of the total) are frozen, preserving the feature extractors learned in Stage A. Only the fully connected output head (≈8,200 trainable parameters) is unfrozen and fine-tuned with a reduced learning rate (10^{-4}) to adapt to the specific sensor characteristics and data distributions of the real-world fleet. The freezing boundary is explicitly indicated in Fig. 2.

Applying standard regression losses to highly dynamic driving data results in physically implausible short-term oscillations in the estimated SOH, since battery degradation is a slow electrochemical process that cannot change appreciably within a single driving session. To encode this prior, a Physics-Informed Neural Network (PINN) loss function is introduced [6], [7] extending the Wide-TCN (hereinafter referred to as PINN-Wide-TCN). The total loss $\mathcal{L}_{total}$ combines a Huber loss (robust to sparse-label outliers) with a temporal regularization penalty:

$$\mathcal{L}_{total} = \mathcal{L}_{Huber}(y, \hat{y}) + \lambda \frac{1}{B-1} \sum_{t=1}^{B-1} (\hat{y}_{t+1} - \hat{y}_t)^2 \tag{3}$$

where y and $\hat{y}$ are the ground-truth and predicted SOH respectively, B is the batch size, and λ is a regularization hyperparameter. For the temporal term, mini-batches are built from B consecutive, non-shuffled windows belonging to the same driving session, so that $\hat{y}_t$ and $\hat{y}_{t+1}$ always refer to adjacent time instants; shuffling is applied at the session level only. The value $\lambda = 0.02$ was selected via grid search over the set $\{0.001, 0.005, 0.01, 0.02, 0.05, 0.1\}$ on the held-out validation sessions of the Target Domain, balancing temporal smoothness against point-wise accuracy. The second term penalizes large inter-sample variations in consecutive SOH predictions, acting as a gradient-level constraint during training. Unlike post-hoc moving-average filters—which introduce inference latency and may mask genuine long-term degradation trends—this formulation intrinsically stabilizes the network's output without requiring any downstream processing in the BMS software stack.

IV. RESULTS AND DISCUSSION

This section evaluates the proposed framework from three complementary perspectives. After introducing the evaluation metrics (Section IV-A), the PINN-Wide-TCN is benchmarked against a BiLSTM and a TCN-Transformer on the five vehicles of the Target Domain, all trained with the same two-stage strategy, to assess cross-capacity accuracy (Section IV-B). The temporal stability of the SOH estimates under dynamic driving conditions is then analyzed (Section IV-C). Finally, the deployed model is validated on-road on two vehicles never

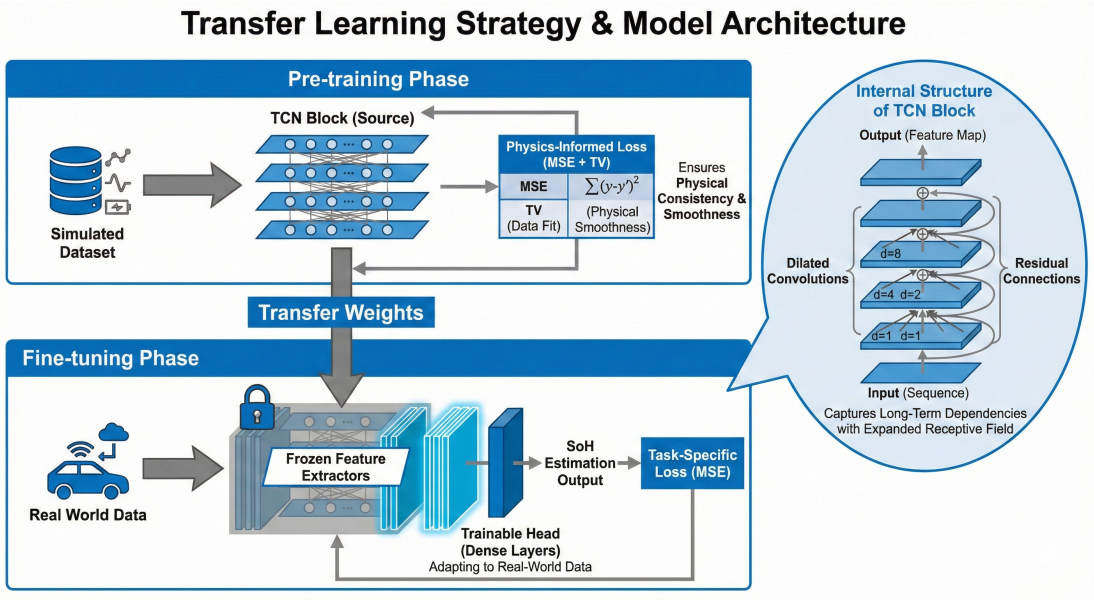


Fig. 2. Transfer Learning strategy: pre-training on simulated data (Stage A) and fine-tuning of the output head on real-world data with frozen convolutional blocks (Stage B).

seen during training, in order to test its behavior under a genuine out-of-distribution scenario (Section IV-D).

A. Evaluation Metrics

To quantitatively evaluate the performance of the proposed framework, three standard regression metrics are employed: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). They are defined as follows:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{4}$$

$$\text{MAPE} = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{5}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{6}$$

where y_i is the ground-truth SOH, $\hat{y}_i$ is the SOH predicted by the neural network, and N is the total number of samples evaluated.

B. Cross-Capacity Evaluation: Transfer Learning Impact

To rigorously evaluate the proposed framework, a benchmarking protocol was established against two state-of-the-art sequence modeling architectures: a Bidirectional LSTM (BiLSTM) and a hybrid TCN-Transformer network [20]. To ensure a fair and unbiased comparison, all three models were trained using the exact same two-stage Transfer Learning strategy, pre-training on the simulated Source Domain (Stage A) and fine-tuning on the real-world Target Domain (Stage B), utilizing the same capacity-agnostic features.

Table IV reports the MAE, MAPE, and RMSE metrics aggregated per drive cycle across all five vehicles in the Target Domain test set. The three benchmarked models are compared: the BiLSTM baseline, the TCN-Transformer, and the proposed physics-informed PINN-Wide-TCN.

The proposed framework outperforms the BiLSTM baseline in MAE on all five vehicles. The largest gain is observed on B_4 (78 kWh), which also represents the largest capacity gap with respect to the source domain (35.8 kWh): the proposed model reduces MAE by 43.9% (from 3.85% to 2.16%) and RMSE by 34.3% (from 4.28% to 2.81%). On B_5 (33 kWh), whose capacity is close to that of the source domain, all models show comparatively high errors and the BiLSTM obtains a marginally lower RMSE (3.91% vs. 3.95%). We attribute this to the near-pristine SOH range of this vehicle (98–100%), which lies at the upper edge of the labeled distribution. Averaged across all five vehicles, the proposed model achieves a 15.3% reduction in MAE, a 15.3% reduction in MAPE, and a 10.9% reduction in RMSE with respect to the BiLSTM baseline.

It is worth noting that on B_2 (40 kWh), the TCN-Transformer achieves a marginally lower MAE (1.27% vs. 1.47%). This vehicle has a nominal capacity close to the source domain (35.8 kWh), a condition under which the distribution shift is minimal, and the attention mechanism of the Transformer provides a slight edge over the PINN regularization. Overall, these results indicate that the capacity-agnostic normalization combined with the temporal regularization reduces the dependence of the model on absolute pack size across the tested fleet, although the advantage is not uniform across vehicles.

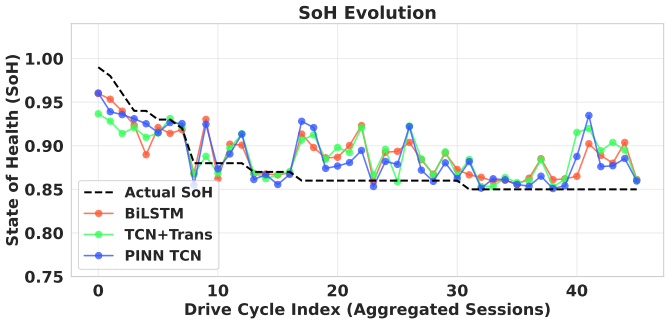
TABLE IV
SOH ESTIMATION METRICS PER VEHICLE AND MODEL (AGGREGATED OVER DRIVE CYCLES). BEST RESULT PER VEHICLE IN **BOLD**.

Vehicle (Cap.)	Model	MAE (%)	MAPE (%)	RMSE (%)
B_1 (35.8 kWh)	BiLSTM	2.31	2.65	2.84
	TCN-Transformer	2.51	2.87	3.25
	PINN-Wide-TCN	**2.08**	**2.39**	**2.82**
B_2 (40 kWh)	BiLSTM	1.57	1.72	1.80
	TCN-Transformer	**1.27**	**1.39**	**1.41**
	PINN-Wide-TCN	1.47	1.60	1.67
B_3 (58 kWh)	BiLSTM	2.25	2.45	2.46
	TCN-Transformer	2.09	2.30	2.24
	PINN-Wide-TCN	**1.96**	**2.13**	**2.13**
B_4 (78 kWh)	BiLSTM	3.85	4.35	4.28
	TCN-Transformer	3.62	4.09	4.14
	PINN-Wide-TCN	**2.16**	**2.46**	**2.81**
B_5 (33 kWh)	BiLSTM	3.27	3.28	**3.91**
	TCN-Transformer	3.98	4.00	4.49
	PINN-Wide-TCN	**3.16**	**3.18**	3.95
Avg. improvement vs. BiLSTM		*−15.3%*	*−15.3%*	*−10.9%*

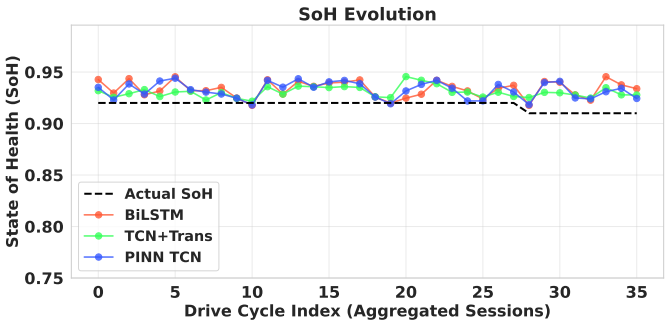
C. Temporal Stability and Physics-Informed Regularization

Beyond point-wise accuracy, a key requirement for on-board BMS deployment is the temporal stability of the SOH estimate. Raw regression models applied to dynamic driving data tend to produce physically implausible short-term oscillations during high-C-rate events such as aggressive acceleration or regenerative braking, where the voltage response is dominated by ohmic and charge-transfer overpotentials rather than by actual capacity fade.

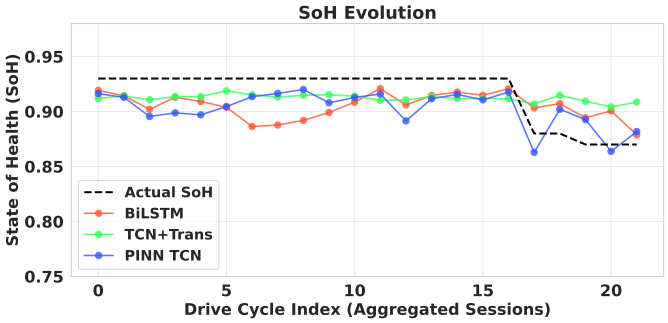
Figs. 3 and 4 show the SOH trajectories predicted by the three models over aggregated drive cycles for all five vehicles. The proposed model tracks the ground-truth SOH trend with visibly smaller inter-cycle fluctuations in most cases. This behavior is consistent with the effect of the temporal PINN penalty in Eq. (3): during training, the network is penalized for inter-sample SOH jumps, which encourages it to reflect the slow electrochemical inertia of degradation.



(a) B_1 (35.8 kWh)



(b) B_2 (40 kWh)



(c) B_3 (58 kWh)

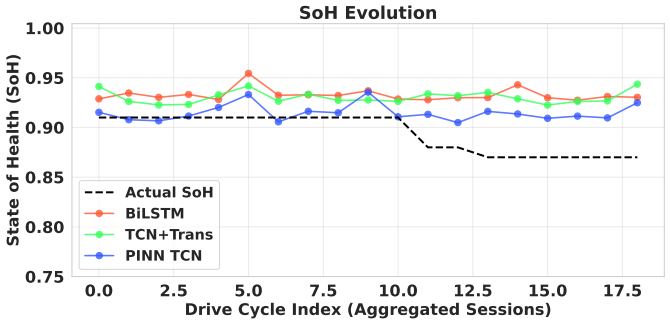
Fig. 3. SOH evolution over aggregated drive cycles for B_1–B_3: ground truth (dashed black), BiLSTM (orange), TCN-Transformer (green), and PINN-Wide-TCN (blue).

D. Out-of-Distribution Real-Time Validation

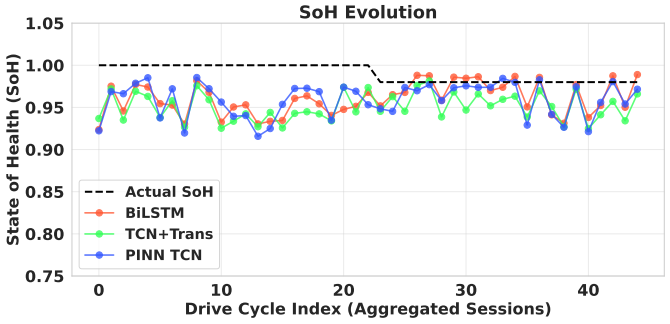
The most critical evaluation of the proposed framework is its behavior on $B_{OOD,1}$ (40 kWh) and $B_{OOD,2}$ (62 kWh): two vehicles sharing the same platform and chemistry but differing in nominal pack capacity by 55%, neither of which was seen at any stage of training or fine-tuning. This experiment directly tests the capacity-agnostic generalization claim of the framework.

Fig. 5 shows the real-time on-board SOH estimate generated by the deployed model during on-road data collection sessions. On $B_{OOD,2}$ (62 kWh, SOH $\approx$ 92%), whose nominal capacity does not appear in the training fleet, the estimate remains physically plausible and stable, with an MAE below 1.2%. On $B_{OOD,1}$ (40 kWh, SOH $\approx$ 83%), the estimate stabilizes within the first few minutes but does not reach the same accuracy, failing to fully capture the more advanced degradation of this pack. This vehicle lies below the SOH range covered by the Target Domain labels (85–100%), so this result highlights the limited extrapolation capability of the model outside the labeled degradation range. The filtered trace in Fig. 5 is shown for reference only, as it would be displayed by the BMS; all reported metrics refer to the raw network output. In both cases, the nominal capacity is used only for input normalization and is not available to the network: the SOH level is estimated solely from the normalized load response (C_{rate}, V_{norm}, T, SOC).

These results indicate that the combination of Partial Freezing and PINN regularization produces a model that is robust to differences in pack size within the tested fleet and temporally stable under the high-frequency disturbances typical of real-world driving, as long as the target SOH lies within the labeled range. However, the OOD evaluation is limited to two



(a) B_4 (78 kWh)



(b) B_5 (33 kWh)

Fig. 4. SOH evolution over aggregated drive cycles for B_4–B_5: ground truth (dashed black), BiLSTM (orange), TCN-Transformer (green), and PINN-Wide-TCN (blue).

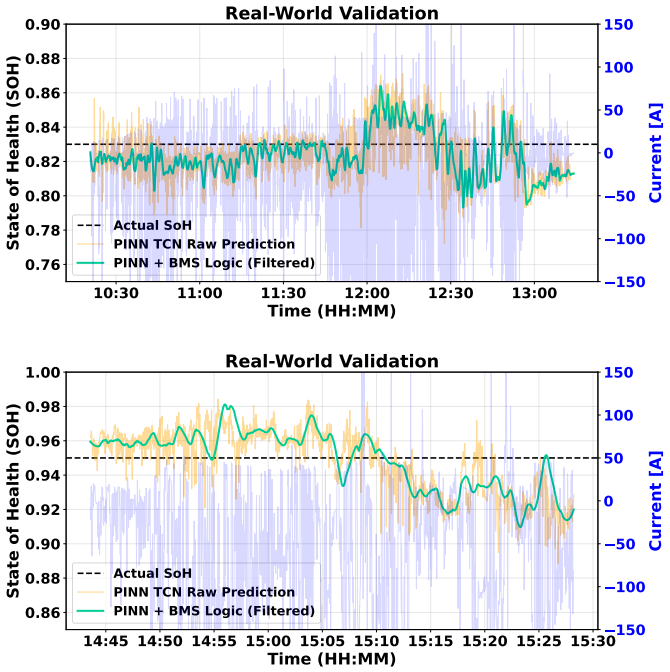


Fig. 5. Real-time SOH on $B_{OOD,1}$ (40 kWh) (top) and $B_{OOD,2}$ (62 kWh) (bottom).

vehicles sharing the same platform and chemistry; a larger multi-manufacturer OOD campaign is required to fully assess cross-capacity and cross-chemistry generalization.

V. Conclusion

This paper presented a two-stage hybrid framework for capacity-agnostic, real-time State-of-Health (SOH) estimation in heterogeneous electric vehicle fleets. By combining capacity-agnostic feature normalization (C_{rate}, V_{norm}), Partial Freezing transfer learning, and a temporal PINN loss, the proposed model generalizes to a battery pack of unseen nominal capacity in on-road tests. Experimental validation demonstrated an average MAE reduction of 15.3% over a baseline BiLSTM, peaking at a 43.9% improvement on B_4 (78 kWh), the vehicle with the largest capacity gap from the source domain. Dedicated on-board tests on two unseen vehicles showed that the framework yields physically consistent SOH estimates solely from normalized load responses, while revealing limited extrapolation below the labeled SOH range.

A key practical outcome of the proposed PINN formulation is the intrinsic smoothness of its predictions, which encourages the network to reflect the electrochemical inertia of degradation. This reduces the need for downstream smoothing such as low-pass filtering or Kalman smoothing. Combined with a lightweight TCN architecture of only 37,000 parameters (8,200 trainable during fine-tuning), the framework is well suited for deployment on automotive-grade microcontrollers with strict memory budgets.

Despite these encouraging results, some limitations remain. The reliance on a relatively narrow SOH band (85%–100% in the labeled data) due to real-world data scarcity limits the validity of the model for more degraded packs, as shown by $B_{OOD,1}$; the OOD evaluation involves only two vehicles of the same platform; and the real-world dataset cannot be publicly released. Future work will focus on modeling deeply aged packs (SOH $<$ 80%) by leveraging generative AI techniques to synthesize physically plausible degradation trajectories, and on extending the on-road validation to vehicles from multiple manufacturers. Additionally, architectural enhancements, such as self-attention mechanisms, and broader thermal features will be explored to further improve robustness under extreme driving and climatic conditions.

References

[1] A. Kumar, S. Sharma, and P. Bhatt, "Cycle based state of health estimation of lithium-ion cells using deep learning architectures," *Sci. Rep.*, vol. 15, no. 1, p. 20995, 2025.

[2] L. Zhang, Z. Mu, and C. Sun, "State of health estimation of lithium-ion batteries with a temporal convolutional network," *Appl. Energy*, vol. 329, p. 120330, Jan. 2023.

[3] W. Liu, M. He, and J. Zhao, "Sensitivity of lithium-ion battery SOH estimates to sensor errors," in *Proc. IEEE EEEIC / I&CPS Europe*, 2025.

[4] J. Li, X. Luo, and Y. He, "Lithium-ion power battery state of health evaluation for electric vehicles based on multi-source transfer learning," *Energy*, 2025.

[5] R. Patel and S. Kumar, "Mitigating catastrophic forgetting in small language models through strategic layer freezing," *Int. J. Adv. Eng. Manage.*, 2025.

[6] H. Huang, Y. Luo, and J. Zhang, "A time series physics-informed neural network framework for battery degradation modeling," *J. Energy Storage*, Art. no. 119549, 2025.

[7] Z. Zhao, X. Chen, and S. Xia, "Physics-aware attentive temporal convolutional network for battery degradation," arXiv:2512.11332, Dec. 2025.

[8] F. Li, Y. Xu, and J. Zhou, "LSTM-based estimation of lithium-ion battery SOH using data characteristics and spatio-temporal attention," *PLOS ONE*, vol. 19, no. 12, p. e0312856, Dec. 2024.
[9] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," arXiv:1803.01271, Apr. 2018.
[10] Y. Zhang, R. Xiong, H. He, and M. G. Pecht, "State of health estimation of lithium-ion batteries based on energy-based features and temporal convolutional network," *IEEE Access*, vol. 10, pp. 77–86, 2022.
[11] M. Weiss, R. Ruiz-Morales, and A. Jossen, "Transfer learning-based state of health estimation for lithium-ion batteries using partial discharge curves," arXiv:2409.00141, Sep. 2024.
[12] Z. Wang et al., "A temporal convolutional network model with attention mechanisms for state-of-health estimation of lithium-ion batteries," *Appl. Energy*, 2026.
[13] Y. Chen, Y. Zhang, X. Zhang, and D. Gu, "A transfer learning-based method for personalized state of health estimation of lithium-ion batteries," *IEEE Trans. Ind. Informat.*, vol. 18, no. 7, pp. 4710–4719, Jul. 2022.
[14] H. Zhang, R. Zhao, and L. Wang, "Transfer learning based generalized framework for state of health estimation of lithium-ion batteries," *Sci. Rep.*, vol. 12, p. 12574, Jul. 2022.
[15] B. Thompson, H. Khayrallah, A. Anastasopoulos, A. Dalmia, G. Neubig, R. Sedoc, and P. Koehn, "Overcoming catastrophic forgetting during domain adaptation of neural machine translation," in *Proc. NAACL-HLT*, 2019, pp. 2062–2068.
[16] Z. Liu et al., "Continual learning for online state of charge estimation across different cycling conditions," *J. Energy Storage*, 2025.
[17] T. Ruan, A. Bhatt, I. D. Gates, and J. A. W. Elliott, "Physics-informed neural network for lithium-ion battery degradation stable training and interpretable characterization," *Nat. Commun.*, vol. 15, p. 4048, May 2024.
[18] Y. Li, D. Che, and X. Zhao, "Physics-informed machine learning for battery degradation forecasting," arXiv:2404.04429, 2024.
[19] R. Gallo, T. Monopoli, M. Zampolli, R. J. P. Jaboeuf, P. Tosco, E. Patti, and A. Aliberti, "A novel procedure for real-time SOH estimation of EV battery packs based on time series extrinsic regression," *IEEE Access*, vol. 13, pp. 326–340, 2025.
[20] Y. Chen, D. Li, X. Huang, J. Hong, C. Mu, L. Wu, and K. Li, "Exploring life warning solution of lithium-ion batteries in real-world scenarios: TCN-transformer fusion model for battery pack SOH estimation," *Energy*, vol. 335, Art. no. 138053, 2025.