

Advancing state of health estimation for electric vehicles: Transformer-based approach leveraging real-world data[☆]

Kosaku Nakano ^{a,*}, Sophia Vögler ^b, Kenji Tanaka ^{a,*}

^a Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, 1138656, Tokyo, Japan

^b School of Engineering and Design, Technical University of Munich, Arcisstrasse 21, Munich, 80333, Bavaria, Germany



ARTICLE INFO

Keywords:

Lithium-ion battery
Electric vehicle
State of health estimation
Deep learning
Transformer

ABSTRACT

The widespread adoption of electric vehicles (EVs) underscores the urgent need for innovative approaches to estimate their lithium-ion batteries' state of health (SOH), which is crucial for ensuring safety and efficiency. This study introduces SOH-TEC, a transformer encoder-based model that processes raw time-series battery and vehicle-related data from a single EV trip to estimate the SOH. Unlike conventional methods that rely on lab-experimented battery cycle data, SOH-TEC utilizes real-world EV operation data, enhancing practical application. The model is trained and evaluated on a real-world dataset collected over nearly three years from three EVs. This dataset includes reliable SOH labels obtained through periodic constant-current full-discharge tests using a chassis dynamometer. Despite the challenges posed by noisy EV real-world data, the model shows high accuracy, with a mean absolute error of 0.72% and a root mean square error of 1.17%. Moreover, our proposed pre-training strategies with unlabeled data, particularly SOH ordinal comparison, significantly enhance the model's performance; using only 50% of the labeled data achieves results nearly identical to those obtained with the full dataset. Self-attention map analysis reveals that the model primarily focuses on stationary or consistent driving periods to estimate SOH. While the study is constrained by a dataset featuring repetitive driving patterns, it highlights the significant potential of transformer for SOH estimation in EVs and offers valuable insights for future data collection and model development.

1. Introduction

Electric vehicles (EVs) play a pivotal role in decarbonizing the road transport sector, which accounts for over 15% of global energy-related emissions [1]. The growing awareness of EVs' environmental benefits has led to a significant increase in their adoption, with global sales projected to reach 35% of total vehicle sales by 2030 [2]. A critical component of EVs is their battery system, primarily lithium-ion batteries (LiBs) [3], chosen for their high energy and power density [4–6]. However, these batteries face a significant challenge: degradation in power and capacity over time [7]. This degradation, influenced by factors such as state of charge (SOC), temperature, voltage, and current, critically affects battery safety and performance [8–10]. Moreover, the fact that the degree of degradation cannot be directly measured [11] further complicates this challenge.

To quantify battery degradation, the State of Health (SOH) serves as a crucial metric [12]. SOH represents the current condition of a battery relative to its ideal, unused state. Accurate SOH estimation

in EVs is vital for several reasons. Firstly, it ensures vehicle safety and performance during operation [13]. Secondly, it enables efficient battery management, reducing ownership costs [13,14]. Thirdly, it optimizes the battery supply chain, informing decisions on recycling, reuse, or repurposing of retiring batteries [15]. This optimization is particularly important given the increasing number of retiring EVs [16]. Lastly, it aids in planning for sustainable sourcing of critical battery materials. This planning is critical in the clean energy transition [17], especially given the rising geopolitical competition over those materials [18]. Given these critical applications and the inherent difficulty in measuring degradation directly, developing accurate SOH estimation methods for EVs is essential for supporting the sustainable transition to electric mobility. Particularly, online SOH estimation, which does not disrupt regular battery use during the estimation process, is more suitable for EV applications as it allows for continuous monitoring without impacting vehicle operation.

Due to the critical nature of accurately estimating SOH for EVs, various methods have been proposed to estimate SOH online. These

[☆] The short version of the paper was presented at ICAE2023, Doha, Qatar, Dec 3–5, 2023. This paper is a substantial extension of the short version of the conference paper.

* Corresponding authors.

E-mail addresses: kosaku47@g.ecc.u-tokyo.ac.jp (K. Nakano), tanaka@tmi.t.u-tokyo.ac.jp (K. Tanaka).

Nomenclature	
Abbreviations	
BN	Batch normalization
CCCV	Constant current constant voltage
CNN	Convolutional neural network
DoD	Depth of discharge
EV	Electric vehicle
FFN	Feed-forward network
IC	Incremental capacity
IEA	International Energy Agency
LiB	Lithium-ion battery
LN	Layer normalization
LSTM	Long short-term memory
MAE	Mean absolute error
MHA	Multi-head attention
MLP	Multi-layer perceptron
MSE	Mean squared error
OCV	Open-circuit voltage
RMSE	Root mean squared error
RPT	Reference performance test
RUL	Remaining useful life
SA	Self-attention
SOC	State of charge
SOH	State of health
Symbols	
α	Ratio of labeled training data
γ	Learning rate decay factor
\mathbf{b}	Bias vector
\mathbf{P}	Positional encoding
\mathbf{W}	Weight matrix
\mathbf{x}	Input data
\mathbf{z}	Latent representation
D	Dataset
\mathcal{L}	Loss function
C	Number of data channels
d	Embedding dimension
E	Energy capacity
f	Model function
k	Kernel size of convolutional layers
L	Number of 1D-convolutional layers
M	Number of samples in dataset
m	Accumulated mileage
N	Number of transformer blocks
p	Probability
s	Stride of convolutional layers
T	Number of data points
y	SOH
Subscripts	
ℓ	Index for convolutional blocks
cls	CLS token
h	Attention head index
k_1, k_2	SOH reference points
k	Key in attention mechanism
o	Output of attention mechanism
q	Query in attention mechanism
v	Value in attention mechanism
Superscripts	
(i)	Index for samples in dataset
0	First token of sequence

methods include Coulomb counting methods, which integrate currents and open circuit voltage (OCV) based methods [19]. Furthermore, the field has seen substantial contributions due to the advancements

of machine learning, with numerous data-driven approaches being explored [20]. More recently, attention-based deep learning models, particularly Transformers, have inspired researchers to propose SOH estimation strategies utilizing attention mechanisms or Transformer architectures [21], and they show promising results, (which is reviewed in Section 2.1).

The studies of SOH estimation techniques can be categorized into two clusters based on data types: lab-experimented and real-world operational data-based studies. In lab-experimented data-based SOH estimation studies, a variety of methods have been proposed [22,23]. These methods, though precise under controlled conditions, fall short in real-world EV applications for two primary reasons. First, they rely on cell-level data to estimate SOH, but such data is often inaccessible in EV application contexts [24]. Second, these methods typically rely on controlled lab settings with consistent charging and discharging conditions. They often assume uniform cycle intervals, depth of discharge (DoD), and temperature, sometimes even presuming full charge–discharge cycles. However, in real-world scenarios, these factors vary significantly and greatly influence battery degradation [25,26]. Moreover, while lab battery cycle tests can emulate driving patterns, synthetic driving cycles fail to accurately replicate the impact of real driving profiles on battery cell lifetime and degradation [27]. This mismatch between lab-derived data and real-world conditions can lead to SOH estimation models that are less effective when applied outside the laboratory, failing to accurately predict the battery’s health under real operational stresses. Although some laboratory-based studies have attempted to simulate real-world conditions through partial cycles or complex operations (reviewed in Section 2.1), their effectiveness in actual operational conditions of EVs has not been empirically validated. Meanwhile, studies based on real-world EV operational data have shown significant progress [28,29], yet still face challenges such as unreliable reference SOH and limited applicability across diverse real-world scenarios. A detailed literature review of these studies is provided in Section 2.2.

In this study, we introduce a **SOH** estimation Transformer for Electric Cars (**SOH-TEC**) to tackle these challenges by leveraging complex, noisy, real-world EV operation data. While there are several definitions of SOH [30], this study focuses on energy capacity-based SOH, which is a critical factor directly related to the cruising range in EVs. **SOH-TEC** incorporates a transformer encoder [31], inspired by recent high-performing SOH estimation works [21]. This architecture processes raw vehicle and battery data from individual trips, eliminating the need for feature engineering and enhancing applicability across diverse real-world driving scenarios.

SOH-TEC is trained and evaluated in a supervised manner using real-world EV operation data collected over a period of nearly three years from three EVs in Japan. This approach marks a significant departure from traditional lab-based studies, directly addressing the gap between controlled experimental conditions and the complexities of real-world applications. While labeling accurate SOH in real-world EVs is challenging in the lab environment [32], we address this by periodically conducting reference performance tests (RPTs) for each vehicle to obtain reliable reference SOH. This test involves a constant-current full-discharge test under 25 °C using a chassis dynamometer.

SOH-TEC utilizes vehicle dynamics and battery profiles during the driving session, or the discharging process, because the controlled charging process in our dataset may not fully capture uncontrolled diverse real-world scenarios. Additionally, the discharging process is less explored in existing research, primarily due to the challenges of replicating these conditions in a laboratory setting. This focus allows us to provide new insights into SOH estimation, filling a critical gap in the field. Moreover, SOH-TEC offers an optional pre-training phase using EV operation data without SOH labels to improve the model's predictive capabilities in a subsequent SOH estimation task. We propose two pre-training methods: masked token prediction and SOH ordinal comparison. This pre-training is crucial for practical applications, as acquiring SOH labels is time-consuming and expensive.

We evaluate SOH-TEC through three experiments: (1) using entire trips as input, (2) using partial trips with only urban driving segments, and (3) investigating pre-training strategies with unlabeled data followed by fine-tuning with limited labeled data. The first two experiments assess the model's foundational predictive capability without pre-training through the comparison with baseline models, while the third explores the effectiveness of pre-training. In addition to these experiments, our study provides insights into the model's development and dataset design. We analyze the self-attention map to identify the model's focus areas and conduct ablation studies to determine optimal settings. Additionally, we discuss strategies for future data collection to address current dataset limitations, such as route diversity and repetitiveness.

To the best of our knowledge, this is the first study to propose a transformer-based SOH estimation model for EVs using real-world operation data. Our main contributions are summarized as follows:

- We develop the transformer-based EV SOH estimation model, SOH-TEC, that processes raw EV operation data from a single trip (Section 3). We also introduce an optional pre-training phase using data without SOH labels. We confirm the effectiveness of the model and the pre-training through experiments (Sections 5 and 6).
- We create a real-world EV operation dataset with reliable reference SOH to evaluate SOH-TEC's performance in diverse driving conditions (Section 4). The reference SOH labels are obtained by conducting periodical RPTs.
- We reveal through a self-attention map analysis that the model focuses on specific driving conditions, such as stationary periods and consistent driving segments (Section 7.1).
- We analyze the impact of hyperparameters on SOH-TEC's performance and identify the optimal configurations through ablation studies (Section 7.2).
- We provide insights for future dataset creation, proposing strategies to enhance data diversity and systematic collection (Section 7.3).

2. Related work

This section provides a comprehensive review of recent SOH estimation studies, focusing on two key areas: studies that leverage attention mechanisms, and work that utilizes actual EV operational data.

2.1. Attention-based SOH estimation studies

Inspired by recent advancements in models utilizing attention mechanisms, including the Transformer, researchers have started applying these techniques to various battery-related tasks. These tasks include SOC estimation [33,34] and battery aging related tasks such as remaining useful life (RUL) prediction [35], early life prediction [36], end of discharge prediction [37] and degradation mode estimation [38].

In parallel, similar approaches have been developed for SOH estimation, highlighting the broad applicability and effectiveness of

attention-based models in battery monitoring and diagnostics. Several researchers approach SOH estimation as a time-series prediction task using attention-based models such as the combination of attention and mixture of expert [39], Transformer [40,41], and Temporal Fusion Transformer [42]. In this formulation, models forecast the SOH value (y_t) at time step t , based on historical data ($y_{1:t-1}$) and/or other relevant variables. However, this formulation assumes the availability of the past SOH or consistent usage pattern, which is not the case in the real-world scenario.

A lot of studies have succeeded in applying attention-based models to online SOH estimation under lab-controlled conditions. The most conventional approach is using the model to process features (or health indicators) extracted based on the domain knowledge [26,43–45]. This approach is effective for estimating the SOH of LiBs under controlled conditions, but it may not be directly applicable to real-world scenarios due to the difficulty of extracting features from complex real-world operation data.

Taking advantage of the attention-based models' ability to process sequences, various studies have proposed approaches that do not rely on domain knowledge-based features. Fan et al. [46] combined Transformer with Long Short-Term Memory (LSTM), while Zou et al. [47] introduced a Deep Cycle Attention Network for SOH estimation. Gao et al. [48] utilized PCC analysis for feature sequence selection before applying a Transformer. Chen et al. [49] adapted the Vision Transformer for SOH estimation, and Bai et al. [50] proposed a Convolutional Transformer-based multi-view information perception framework. Shi et al. [51] developed a Transformer-based model incorporating a dual-encoder architecture and a gating mechanism. These diverse approaches demonstrate the flexibility and potential of attention-based models in SOH estimation tasks. Notable among these studies are the works of Gu et al. [22] and Cai et al. [52]. Gu et al. [22] proposed a convolutional neural network (CNN)-Transformer model, processing feature sequences created with PCC and principal component analysis. Cai et al. [52] developed a CNN-attention model capable of simultaneously estimating SOC, SOH, and RUL, showcasing the potential for multi-task learning in battery diagnostics. Although these studies have made significant contributions, they face limitations in real-world applications due to their dependence on controlled laboratory conditions, including constant current discharge, uniform charging patterns, and complete charge-discharge cycles.

To tackle these challenges, recent studies have begun to focus on more realistic scenarios that better reflect the complexities of EV operation in the real world. For instance, Ke et al. [53] developed a temporal convolutional network with a self-attention mechanism to estimate SOH of battery cells cycled on driving profiles based on their charging process. Tang et al. [54] introduced a method capable of estimating SOH from partial charging profiles, addressing the common scenario where EVs rarely experience full charge-discharge cycles. Sun et al. [55] designed a dual-branch CNN transformer for estimating SOH from a fast-charging profile with an arbitrary SOC range. Chen et al. [56] proposed a self-attention knowledge domain adaptation network. Their model can estimate SOH from shallow cycle data by leveraging knowledge transfer from labeled full cycle data. While these studies on complex battery operations have primarily focused on the charging process, which is a valuable research direction, they still rely on lab-experiment data-based studies and their applicability to real-world EV users' uncontrolled charging behavior remains unconfirmed.

2.2. SOH estimation studies based on EV data

In this section, we review recent studies on SOH estimation based on EV data. While these studies have made valuable contributions, they still need improvement in terms of limited applicability to real-world scenarios and unreliable reference SOH values.

Some researchers have focused on estimating the resistance-based SOH of LiBs using EV operation data. For example, Giordano et al. [57] defined SOH as the 10-second discharge resistance and introduced a model-based estimation approach. Similarly, Wang et al. [15] predicted the internal resistance of EVs as a measure of SOH, treating it as a time-series prediction task. While these resistance-based methods show promise, they currently lack applicability in estimating capacity fade.

In contrast, many studies focus on estimating capacity-based SOH. Zhang et al. [58] utilized a dataset of 7,296 plug-in hybrid EVs to recursively estimate capacity fade at each time step. However, this recursive approach accumulates errors, which may limit its usefulness in practical applications. Further exploring capacity-based approaches, She et al. [59] developed a radial basis function neural network to predict the incremental capacity (IC) curve peak value, which is a key indicator in battery aging. Zhao et al. [60] devised an extreme learning machine-based SOH estimation framework where the target SOH was modeled as a function of accumulated mileage. Similarly, Song et al. [61] estimated SOH using a feed-forward network (FFN) and evaluated their method on a dataset from 700 EVs. Although these studies achieved high performance and offer valuable insights, they rely on mileage as an input feature. This reliance may not be suitable for vehicles that do not drive consistently, particularly those for which calendar degradation is the predominant mode of deterioration.

Another notable approach is presented by Wen et al. [28], who proposed the NGBoost model to estimate the reference capacity computed by ampere-time integration divided by depth of charge. However, this method's reliability is impacted by SOC inaccuracies and temperature effects. Addressing these issues, Liu et al. [29] pointed out the problems of this ampere-time integration and introduced an improved integration method. Their method focused on identifying segments where SOC estimation is most accurate, enhancing reliability. Despite its accuracy, this method requires specific conditions for application, such as certain EV rest durations and SOC ranges, which may not be universally applicable in real-world scenarios. Lastly, Zhao et al. [62] employed machine learning models to predict cell-level capacity and pack-level RUL of EV batteries, utilizing features derived from IC and differential voltage analysis. While their approach demonstrates robustness, the authors noted that their method lacks applicability in uncontrolled charging scenarios, including partial and fast charging, which are common in everyday EV usage.

3. SOH-TEC: SOH estimation transformer for electric cars

The SOH-TEC model estimates the SOH using EV operation data from a single trip, as depicted in Fig. 1. The problem formulation and the SOH-TEC architecture are provided in Section 3.1 and Section 3.2, respectively. The model is trained in a supervised manner for the SOH estimation task (Section 3.3). Additionally, SOH-TEC can be optionally pre-trained on datasets without SOH labels to further improve its performance (Section 3.4). The training pipeline that includes pre-training and supervised learning for SOH estimation is explained in Section 3.5.

3.1. Problem formulation

We first define the task of estimating SOH using data from a single EV trip. A single trip is defined as the consecutive period during which the EV is in drive mode. The data from a single trip is represented as $\mathbf{x} \in \mathbb{R}^{C \times T}$. Here, C is the number of data channels, which can include vehicle and battery-related data, and T is the number of data points or time steps. In this study, the data channels include vehicle speed, battery voltage, current, and SOC, thus $C = 4$. The data is recorded at one-second intervals, so T corresponds to the trip's duration in seconds. The SOH $y \in [0, 100]$, which is the estimation target, is defined as:

$$y = \frac{E_{\text{current}}}{E_{\text{initial}}} \times 100\%, \quad (1)$$

where E_{current} is the current energy capacity in kWh and E_{initial} is the initial energy capacity in kWh. The objective is to develop a model $f : \mathbb{R}^{C \times T} \rightarrow \mathbb{R}$ that maps the input data \mathbf{x} to the SOH y .

3.2. Network architecture

To accurately deduce the SOH using data from a single EV trip, we develop SOH-TEC, a neural network architecture that incorporates the transformer encoder. The motivation behind leveraging the transformer encoder is its competency in recognizing the dependencies inherent within sequential data. A visual representation of the proposed model architecture can be found on the right side of Fig. 1.

3.2.1. Input embedding

Tokenization or patching is essential for Transformers to effectively process sequential data, as demonstrated in natural language processing [63], computer vision [64], and audio processing [65]. This approach is similarly utilized in our model, where raw time-series data $\mathbf{x} \in \mathbb{R}^{C \times T}$ is converted by the input embedding module into a high-dimensional latent space. The input embedding module of SOH-TEC is composed of a sequence of L convolutional blocks. Each block comprises a 1D-convolutional layer, a batch normalization (BN) [66], and a ReLU activation function. The convolutional layer of the first block applies a kernel size k_1 and stride s_1 , expanding the data's dimensionality C to the embedding dimension d (Eq. (2)). The convolutional layers of the subsequent blocks apply a kernel size k and stride s , and the embedding dimension is unchanged (if $L > 2$, Eq. (3)). The final block returns logits without applying BN or activation functions (Eq. (4)). In this study, a number of embedding layers, kernel sizes, strides, and embedding dimensions are set as $L = 2$, $k_1 = 4$, $k = 3$, $s_1 = 2$, $s = 2$, and $d = 256$, respectively.

$$\mathbf{x}_1 = \text{ReLU}(\text{BN}(\text{Conv}(\mathbf{x}; k_1, s_1, d))) \quad (2)$$

$$\mathbf{x}_\ell = \text{ReLU}(\text{BN}(\text{Conv}(\mathbf{x}_{\ell-1}; k, s, d))), \quad \ell = 2, \dots, L-1 \quad (3)$$

$$\mathbf{x}_L = \text{Conv}(\mathbf{x}_{L-1}; k, s, d). \quad (4)$$

3.2.2. CLS token and positional encoding

The embedding process incorporates a learnable CLS token, analogous to those used in BERT [63] and Vision Transformer [64]. The CLS token $\mathbf{x}_{\text{cls}} \in \mathbb{R}^d$ is prepended to the output sequence \mathbf{x}_L from the input embedding module. This token aggregates and represents the sequence's information for tasks such as SOH estimation. Additionally, learnable positional encoding $\mathbf{P} \in \mathbb{R}^{d \times (T'+1)}$ is added to maintain the sequence order, where T' is the length of \mathbf{x}_L . Both the CLS token and positional encoding are initialized with a truncated normal distribution (mean $\mu = 0$, standard deviation $\sigma = 0.2$, truncation limits $[-2, 2]$) to stabilize early training.

$$\mathbf{z}_0 = ([\mathbf{x}_{\text{cls}}; \mathbf{x}_L] + \mathbf{P})^\top \quad (5)$$

3.2.3. Transformer encoder

The Transformer encoder in SOH-TEC, depicted on the right side of Fig. 1, consists of $N = 4$ transformer blocks. Each transformer block consists of two sub-layers: a Multi-head Attention (MHA) layer (Eq. (6)) and a FFN layer (Eq. (7)). Each sub-layer includes residual connections and employs a pre-norm architecture with Layer Normalization (LN) [67] where normalization precedes the sub-layers. Additionally, DropPath is applied to the output of each sub-layer to enhance the model's generalization and prevent overfitting. Drop-path involves randomly dropping the output of sub-layers while training, in which case the input from the residual connection is the only contribution passed to the next sub-layer. We employ a drop-path rate of $p_{\text{drop-path}} \times (i/N)$ for i th transformer block, implementing a stochastic depth approach. In this study, we set $p_{\text{drop-path}} = 0.1$.

$$\mathbf{z}'_i = \mathbf{z}_{i-1} + \text{DropPath}(\text{MHA}(\text{LN}(\mathbf{z}_{i-1}))) \quad (6)$$

$$\mathbf{z}_i = \mathbf{z}'_i + \text{DropPath}(\text{FFN}(\text{LN}(\mathbf{z}'_i))). \quad (7)$$

MHA in the first sub-layer is composed of multiple self-attention (SA) heads, and MHA is calculated by concatenating the output of

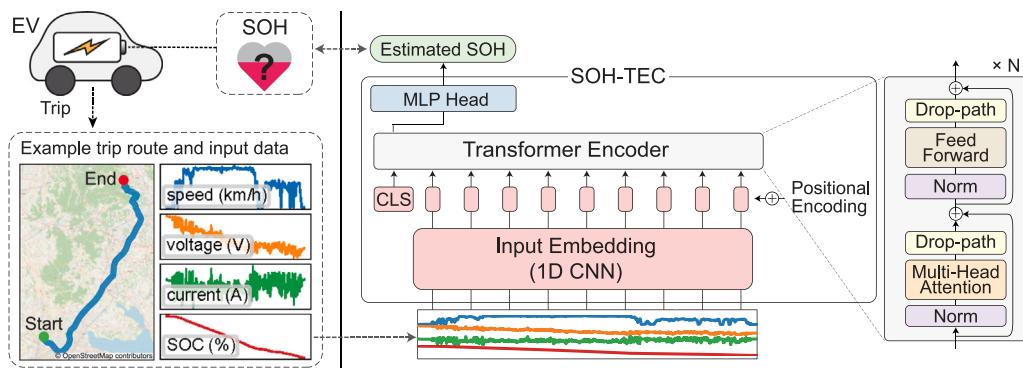


Fig. 1. An overview of the SOH estimation by SOH-TEC. The left side shows an example of trip routes where a vehicle drives and the corresponding vehicle and battery-related data obtained during the trip.¹ This sequence is used as an input to the model. The right side illustrates the architecture of the transformer-based SOH estimation model. The input sequence is embedded by an input embedding module and then processed by the transformer encoder. The latent representation corresponding to the CLS token is fed into an MLP head, resulting in an estimated SOH.

$H = 16$ SA heads and passing them through a linear layer. (Appendix A.1 provides more details about MHA). To address variable sequence lengths and avoid overfitting the input sequence, we apply a key mask to MHA. The key mask works by setting the attention score of masked positions before the softmax operation to $-\infty$ in the SA calculation, which effectively removes the contribution of the masked positions to the SA calculation. In addition to masking, we also apply attention dropout with dropout rate $p_{\text{attn_dropout}} = 0.1$.

The FFN in the second sub-layer is a simple two-layer MLP with a GeLU activation function. We employ the dimension of the hidden layer as $d_{\text{FFN}} = 4d$. While the FFN can apply dropout, the dropout rate is set to $p_{\text{ffn_dropout}} = 0.0$.

3.2.4. Multilayer perceptron (MLP) head

The latent representation corresponding to the CLS token is used for the regression task of SOH estimation. This representation is denoted as $\mathbf{z}_{\text{cls}} \in \mathbb{R}^d$. Specifically, \mathbf{z}_{cls} is the first token of the last transformer block's output, as $\mathbf{z}_{\text{cls}} = \mathbf{z}_N^0$. The MLP head for SOH estimation consists of two fully connected layers with a ReLU activation function in the hidden layer. The hidden layer dimension is set to $d_{\text{MLP}} = d/2$. The output layer has a dimension of 1 and does not employ an activation function.

3.3. Supervised learning for SOH estimation with SOH-labeled dataset

The SOH-TEC model is trained in a supervised manner utilizing a dataset $D = \{(\mathbf{x}^{(i)}, y^{(i)})\}_{i=1}^M$, where M is the number of samples in the dataset. The mean squared error (MSE) loss function is used for the training process. The model is trained for 300 epochs. The learning rate is set to 5e-5 at the beginning and 1e-5 after 150 epochs. The purpose of this learning rate decay is to help convergence. AdamW [68] is employed as the optimizer. Gradient clipping is applied with a max norm of 10 to the gradients to prevent the problem of exploding gradients.

For data augmentation purposes, the input sequence undergoes random cropping during training. The lengths of these cropped sequences are sampled from a discrete uniform distribution $U(\lambda S, S)$. Here, λ represents the minimum length ratio and S denotes the original sequence length. For evaluation, the sequence length is fixed at S . In this study, the values are set as $\lambda = 0.9$ and $S = 1800$. This random cropping helps simulate varying driving conditions and durations. Additionally, we apply random key masking (explained in Section 3.2.3) to the input sequence while training. The random masking ratio, p_{mask} , is sampled from a uniform distribution $U(0.1, 0.3)$ for each input. The mask is

applied starting from a randomly chosen point $t \in \{0, \dots, T-1\}$ (where T is the sequence length) and extends for $\lfloor p_{\text{mask}} \cdot T \rfloor$ positions in the sequence. This masking technique helps the model avoid overfitting by preventing it from relying too heavily on specific parts of the input data.

3.4. Pre-training with unlabeled data

Creating SOH-labeled datasets for EVs is notably resource-intensive because it requires years of driving data collection coupled with periodic RPTs to accurately measure the SOH. The transformer models have demonstrated exceptional capabilities across various applications through the utilization of pre-training paradigms, particularly with unlabeled data [63,64,69]. In the context of battery state estimation, the use of pre-training and transfer learning protocols has been identified as a particularly promising approach [70]. Building on this progression, this study proposes two pre-training methods and verifies their effectiveness: namely masked token prediction and SOH ordinal comparison. These methods enhance the SOH-TEC model's understanding of SOH progression, improving predictive accuracy in downstream SOH estimation tasks. The pre-training methods are illustrated in Fig. 2.

3.4.1. Masked token prediction

In the masked token prediction pre-training, the SOH-TEC model is trained to predict the original input tokens from masked versions of the input sequences (see the left side of Fig. 2). This pre-training guides the model in learning meaningful representations of the input data by effectively predicting the masked tokens. The process unfolds as follows: initially, the input sequence undergoes embedding, positional encoding, and the addition of a CLS token. Subsequently, each token in the sequence, except for the CLS token, is replaced with a learnable mask token with a probability of 50%. This masked sequence is then fed into the transformer encoder, whose output tokens are processed by a two-layer MLP token-wise. The MLP head consists of two linear layers with a hidden dimension of d and a ReLU activation function in between. The loss is calculated as the MSE between the masked original tokens and their corresponding reconstructed tokens. For the training process, the same hyperparameters are used as in the SOH estimation task, which is described in Section 3.3.

3.4.2. SOH ordinal comparison

In the SOH ordinal comparison pre-training phase, the model performs an ordinal comparison by predicting which of the two input trips has a higher SOH (see the right side of Fig. 2). The pair of input trips are sampled from the same EV, and which trip has a higher SOH is determined based on their accumulated mileage. This SOH comparison through the mileage is feasible because the SOH decreases

¹ Map data from OpenStreetMap (openstreetmap.org/copyright).

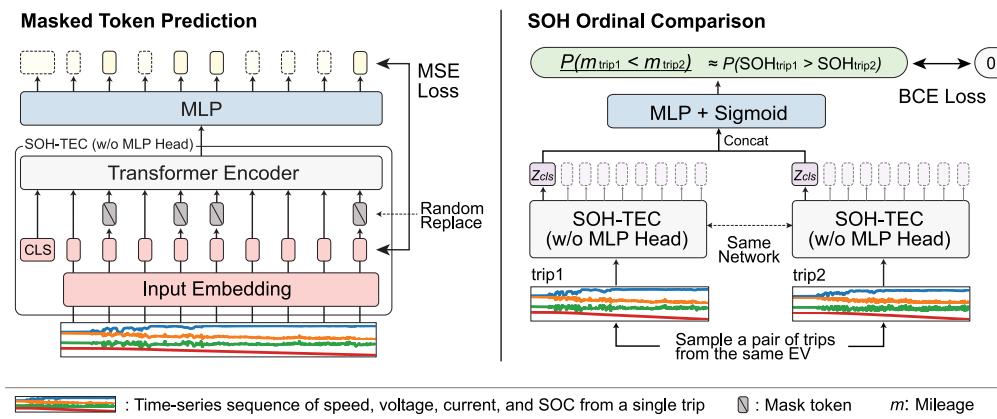


Fig. 2. Overview of the pre-training methods for SOH-TEC. (Left) Masked token prediction: the model learns to predict original input tokens from masked versions of the input sequences. (Right) SOH ordinal comparison: the model processes pairs of input sequences from the same EV and compares their SOH values based on mileage.

monotonically with mileage. The pre-training employs a Siamese network architecture, where both input sequences are processed by the same network. This pre-training is designed to help the model learn better input representations for SOH estimation without requiring SOH labels.

To process the input sequences, we use the SOH-TEC model without the MLP head used for the SOH estimation task. After the process by the SOH-TEC model, the embedded tokens corresponding to the CLS tokens from both input sequences are concatenated and fed into a two-layer MLP head. This MLP head consists of a linear layer followed by a ReLU activation function, another linear layer, and a sigmoid activation function. The input dimension of the MLP head is $2d$, the hidden dimension is d , and the output dimension is 1. The output represents the probability that the mileage of the first input sequence is smaller than that of the second one. This probability corresponds to the likelihood that the SOH of the first input trip is greater than that of the second, given the inverse relationship between mileage and SOH. The binary cross-entropy (BCE) loss is used as the loss function.

For the training process, all valid trip pairs in the dataset are used. For instance, if the dataset includes n trips for a vehicle, $\binom{n}{2}$ pairs of trips are used. In this study, due to the quadratic growth of the dataset size with the number of trips, the SOH ordinal comparison pre-training is limited to 5 epochs. Other hyperparameters remain the same as those used in SOH estimation.

3.5. Training pipeline

The SOH-TEC model can learn the SOH estimation task in a supervised manner using an SOH-labeled dataset, or it can first be pre-trained on a dataset without SOH labels and then fine-tuned on a dataset with SOH labels. Fine-tuning is the process of training the model for the SOH estimation task after the pre-training process. This SOH-TEC training pipeline is illustrated in Fig. 3. The first two training steps are pre-training, which is optional and can be skipped. Here, the model uses the SOH-unlabeled dataset only for pre-training. The pre-training process does not employ any dataset split; instead, the model is trained to minimize the loss on the entire unlabeled dataset. The final step is the supervised learning for SOH estimation, or fine-tuning if the pre-training process is conducted. In this process, the model uses the dataset with SOH labels. The dataset is split to separate the training, validation, and test datasets. In the transition between these training steps, only the MLP head is replaced for the subsequent training step, while the input embedding module and transformer encoder remain the same. For instance, the model is trained for SOH ordinal comparison. After this phase, the MLP head is replaced, and the model is then trained for SOH estimation.

SOH-TEC Training Pipeline

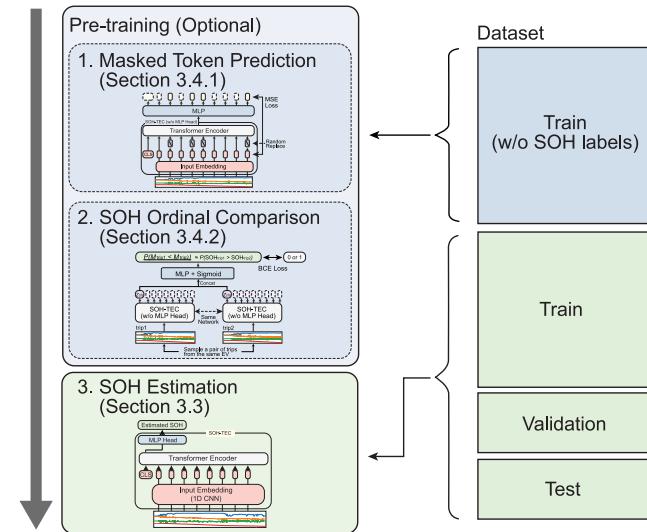


Fig. 3. Training pipeline for SOH estimation. The SOH-TEC model can be optionally pre-trained on a dataset without SOH labels. If pre-trained, the model is then fine-tuned on a dataset with SOH labels for SOH estimation. If no pre-training is conducted, the weight-initialized model is directly trained on the dataset with SOH labels for SOH estimation.

4. Dataset

This section details the dataset $D = \{(x^{(i)}, y^{(i)})\}_{i=1}^M$ used in this study. Data was collected from three EVs of the same mass-market model by the Japan Automobile Research Institute. The collection period spanned from August 2011 to June 2014, with missing data from September 2012 to March 2013. The data collection consists of two parts: collecting EV real-world operation data (Section 4.1) and obtaining ground truth SOH from RPTs (Section 4.2).

4.1. EV operation data collection

The driving tests were designed to promote battery degradation while collecting comprehensive data across various conditions. Each vehicle followed predefined schedules comprising six main routes: three highway routes (Highway 1a, 1b, 2) and three urban routes (Urban 1a, 1b, 2). Routes not matching these categories were labeled as “Others”, which included driving within the research facilities. To ensure data quality, only trips lasting 10–90 min and covering 10–90 km

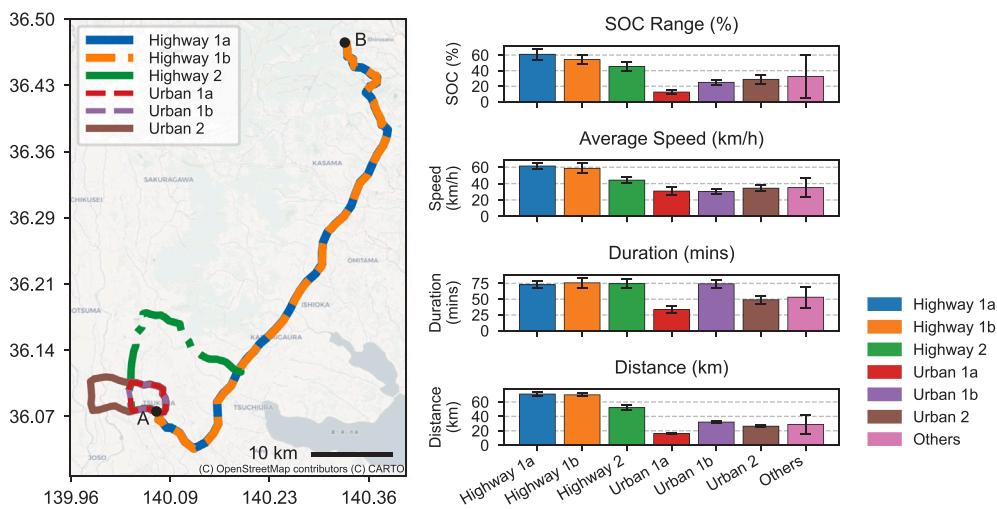


Fig. 4. Overview of trip routes.² The trips in the dataset are labeled according to their route type. Highway routes (1a, 1b, 2) include travel on highways, with 1a and 1b covering the same path in opposite directions. Urban routes (1a, 1b, 2) stay within the city, with 1a being a single loop and 1b a double loop. Trips not fitting these categories are labeled “Others”. The map on the left shows the trip routes, with Highway 1a starting at point A and ending at B, Highway 1b going in the opposite direction, and all other routes both starting and ending at point A. The bar chart on the right displays the mean and standard deviation of key metrics for each route type, including SOC range, average speed, trip duration, and total distance traveled.

were included in the dataset. The testing protocol involved distinct phases, each lasting from several months to over two years. A phase was composed of one or two designated routes. During the phase, a vehicle drove these routes one or several times a day, depending on the route length. After completing a phase, the vehicle switched to different routes for the next phase. Each vehicle experienced several phases throughout the testing period, following a unique combination. Vehicles underwent either fast or normal charging using the constant current constant voltage (CCCV) protocol.

Fig. 4 provides a comprehensive overview of the trip routes. (Individual route visualizations and trip counts are included in Appendix A.2). The left side of the figure displays a map of the six primary routes. Highways 1a and 1b traverse the same path in opposite directions, with 1a starting at point A and ending at B. The other routes are round trips, beginning and ending at point A. Urban 1a is a single loop, while Urban 1b is a double loop on the same route. The right side of Fig. 4 presents a bar chart showing the mean and standard deviation of key metrics for each route type: SOC range, average speed, trip duration, and total distance traveled. Highway routes generally exhibit a higher SOC range, average speed, and total distance, while urban routes show a lower SOC range, lower average speed, and shorter duration. “Others” routes display larger variances across all metrics, as they comprise trips not fitting the highway or urban categories.

During the drive, speed, voltage, and current — used as inputs for the SOH estimation task — are recorded every second. SOC, however, is logged every minute. For consistency with the second-interval input data x required by SOH-TEC, the SOC data is linearly interpolated to match the one-second frequency. Fig. 5 illustrates examples of the input data x for each route. Highway driving sections, characterized by higher (> 75 km/h) and more consistent speeds, exhibit steeper voltage and SOC decreases over time. In contrast, urban driving sections show lower speeds and less pronounced voltage and SOC decreases. While more than half of Highway 1a and 1b trips consist of highway driving segments, Highway 2 trips are mostly urban, with only about 10 min of highway driving (refer to speed profiles in Fig. 5).

4.2. Ground truth SOH

To obtain the dependent variable $y^{(i)}$, which represents the SOH of the battery, periodic RPTs were conducted for each vehicle alongside the collection of the operation data. The RPT involved a constant-current full discharge with a C-rate of $C/3$ under a controlled temperature of 25 °C. The C-rate is defined as the current rate relative to the battery’s nominal capacity, meaning that a $C/3$ discharge would fully deplete a non-degraded battery in three hours. Before the full discharge test, the vehicle was charged to 100% SOC and rested for at least a few hours at about 25 °C. These controlled environments ensure consistent battery capacity measurements. These tests were performed using a chassis dynamometer, which enables vehicle operation while stationary. Fig. 6 illustrates this setup, where the vehicle’s wheels are placed on rollers simulating road conditions. During the RPT, the vehicle was operated to maintain a constant current of $C/3$. The test continued until the vehicle could no longer sustain the $C/3$ discharge. RPTs were conducted approximately every six months to evaluate battery degradation throughout the data collection period.

The SOH is calculated based on the current and voltage records from the RPT. However, since RPTs are conducted only approximately every six months, linear interpolation is necessary to estimate SOH values for trips occurring between these tests. This interpolation is applied based on accumulated mileage obtained from the vehicle’s odometer. The method is reasonably robust due to the monotonic decrease of SOH over time and cycle, as well as the repetitive nature of the test routes and consistent driving patterns.

To construct the dataset $D = \{(x^{(i)}, y^{(i)})\}_{i=1}^M$, each trip is paired with the accumulated mileage from the vehicle’s odometer at the start of the trip. For each trip of a particular vehicle, we have corresponding input data denoted as $x^{(i)}$. The associated SOH $y^{(i)}$ for this trip is determined through linear interpolation using the RPT results of the vehicle and the odometer reading for that specific trip. Specifically, the RPTs provide SOH values at various mileages for each vehicle, and the interpolation is conducted between the two nearest reference points surrounding the trip’s odometer reading. For example, if the trip’s odometer reading is $m^{(i)}$, and the nearest reference points are m_{k1} (with SOH of y_{k1}) and m_{k2} (with SOH of y_{k2}), the SOH for the trip is estimated as Eq. (8). This method ensures that the interpolated approximation is realistic and grounded in the actual vehicle usage between RPTs.

$$y^{(i)} = y_{k1} + \left(\frac{m^{(i)} - m_{k1}}{m_{k2} - m_{k1}} \right) \times (y_{k2} - y_{k1}). \quad (8)$$

² Map data from OpenStreetMap (openstreetmap.org/copyright) and CARTO (<https://carto.com/>).

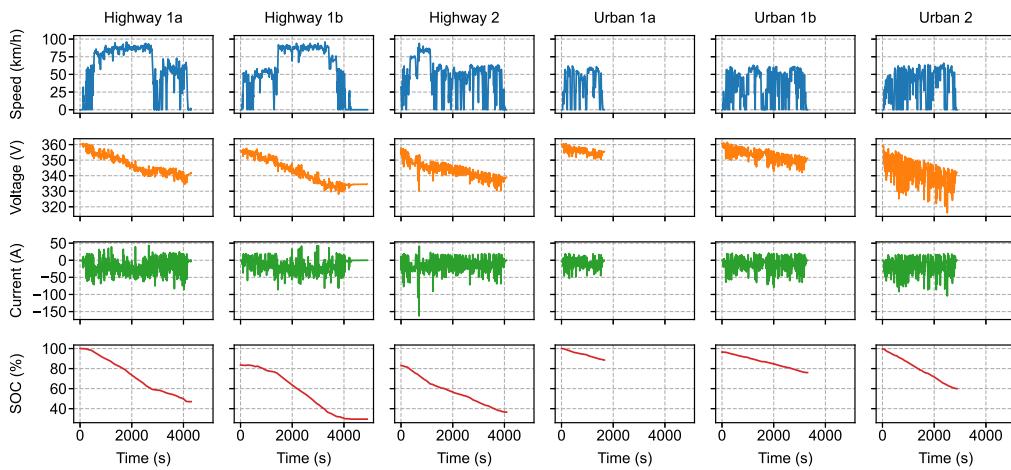


Fig. 5. Speed (km/h), voltage (V), current (A), and SOC (%) profiles over time for examples from six trip routes—Highway 1a, Highway 1b, Highway 2, Urban 1a, Urban 1b, and Urban 2.

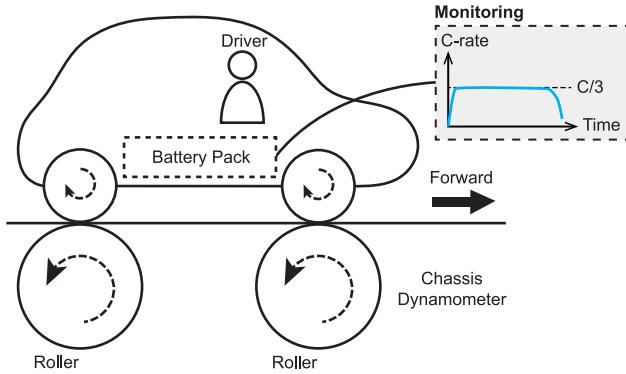


Fig. 6. Illustration of the full discharge test with a chassis dynamometer. A driver drives the vehicle to maintain the constant current of $C/3$. The vehicle is driven as long as it can keep the $C/3$ discharge rate.

5. Experimental setup

To evaluate the effectiveness of SOH-TEC, we employ the following nine most fundamental models as comparative baselines: Dummy Regressor, Route-aware Dummy Regressor, MLP, CNN, LSTM, Linear Regression, Ridge Regression, Random Forest, and LightGBM. The detailed descriptions of these models are provided in [Appendix A.3](#). While recent SOH estimation models have shown significant advancements, they are primarily designed for specific contexts such as controlled lab settings or charging processes. Thus, they may not align well with the real-world discharging processes central to our study. Adapting these models without considering their original context could lead to skewed results and misinterpretations, which is why they were not included as baselines.

For the model training process, we randomly split the dataset into training, validation, and test sets, with a ratio of 8:1:1. Throughout the whole experiment, the same data split is used. The model is trained using the training dataset, and its performance is assessed with the validation dataset at the end of each epoch. The best model checkpoint, determined by the lowest validation loss, is then evaluated on the test dataset to assess its effectiveness. We use the root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination R^2 as evaluation metrics. The four deep learning-based models (SOH-TEC, MLP, CNN, and LSTM) and two tree-based models (Random Forest and LightGBM) are trained with eight different random seeds to ensure the reproducibility of the results. The models' hyperparameters and their searching process are detailed in [Appendix A.4](#).

6. Experiments

We conducted three experiments to evaluate the SOH-TEC model. The base setting experiment utilized full trips from the dataset for training and testing ([Section 6.1](#)). The urban driving setting focused on urban driving segments to assess performance under typical urban conditions ([Section 6.2](#)). Both these experiments compared SOH-TEC against nine baseline models for SOH estimation, using only labeled data in a supervised manner. The third experiment explored the effectiveness of masked token prediction and SOH ordinal comparison pre-training strategies. This experiment used partially unlabeled data for pre-training, followed by fine-tuning on the remaining labeled data ([Section 6.3](#)). Detailed results and analysis for each experiment are presented in the following subsections.

6.1. Experiment 1: base setting

In this experiment, we used the full trips from the dataset for the training and testing. This experiment aims to evaluate the basic estimation performance of SOH-TEC and compare it with the baseline models. The comprehensive results of the experiment are shown in [Table 1](#). (The training and validation loss curve can be found in [Appendix A.5](#).) [Table 1](#) provides a quantitative performance outcome through three metrics: RMSE, MAE, and R^2 . Lower RMSE, MAE, and higher R^2 indicate better performance. As indicated in [Table 1](#), the SOH-TEC model outperforms all other models significantly across all metrics (RMSE, MAE, and R^2). Specifically, SOH-TEC achieved an RMSE of 1.17 (± 0.17) and an MAE of 0.72 (± 0.04). In contrast, the best-performing baseline model, LSTM, recorded an RMSE of 1.80 (± 0.13) and an MAE of 1.12 (± 0.05). The effectiveness of SOH-TEC's transformer encoder-based approach for EV SOH estimation is evident from these results, underscoring its advanced capability over traditional models.

[Fig. 7](#) shows scatter plots (true SOH vs. estimated SOH) for the top four models: SOH-TEC, LSTM, CNN, and LightGBM. (Full scatter plots are available in [Appendix A.6](#).) In the scatter plot, each point represents a test sample, color-coded by trip route and the black dashed line represents perfect estimation. Here, we observe that SOH-TEC's predictions are tightly clustered around this line. In contrast, other baseline models show broader point dispersion. Notably, SOH-TEC accurately estimates SOH for challenging trips that other models struggle with, such as the trip in the 'Others' cluster with true SOH around 67%. These results indicate the SOH-TEC's superiority in terms of high accuracy and consistency across various driving conditions. The 3% margin delineated in [Fig. 7](#) further highlights SOH-TEC's precision; most predictions fall within this narrow error margin, clearly demonstrating the model's reliability.

Table 1

Comparison of SOH estimation performance for different models in Experiment 1: base setting using the full dataset. The best-performing model is highlighted in bold.

Model	RMSE↓	MAE↓	$R^2 \uparrow$
Dummy Regressor	6.822	5.459	-0.004
Route Aware Dummy Regressor	4.454	3.492	0.572
Linear Regression	3.446	1.964	0.744
Ridge Regression	3.058	1.942	0.798
Random Forest	2.262 (± 0.03)	1.554 (± 0.02)	0.890 (± 0.00)
LightGBM	1.984 (± 0.02)	1.327 (± 0.01)	0.915 (± 0.00)
MLP	2.168 (± 0.08)	1.621 (± 0.07)	0.899 (± 0.01)
CNN	1.992 (± 0.05)	1.451 (± 0.03)	0.914 (± 0.00)
LSTM	1.801 (± 0.13)	1.124 (± 0.05)	0.930 (± 0.01)
SOH-TEC (Ours)	1.167 (± 0.17)	0.719 (± 0.04)	0.970 (± 0.01)

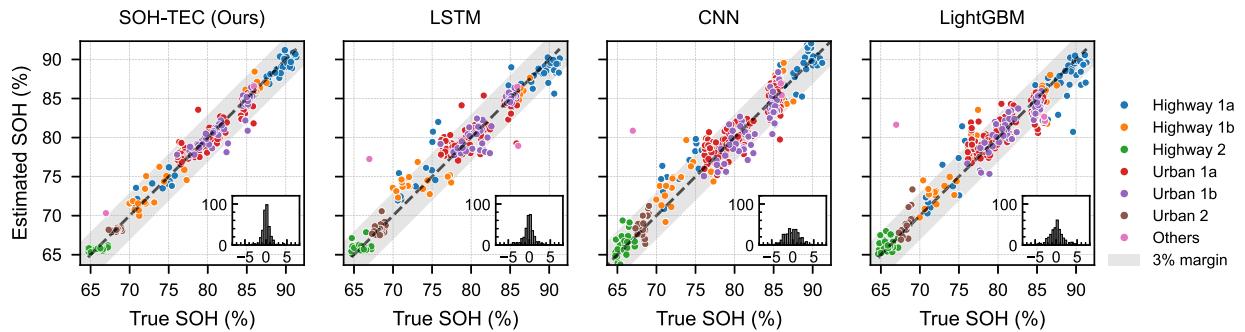


Fig. 7. The estimation results of Experiment 1: base setting. The scatter plot illustrates true SOH vs. estimated SOH by the best-performing models for the test dataset. Marker colors represent different trip routes, and the gray region illustrates the 3% error margin. The insets show the error distributions for each model.

6.2. Experiment 2: urban driving setting

In addressing the challenge posed by repeated and limited trip routes in the dataset, our approach involved random masking and cropping techniques to foster a more generalized learning process. However, recognizing the potential over-representation of highway driving characteristics and their correlation with the SOH estimations, our second experiment narrowed the focus to urban driving conditions. This choice was driven by two reasons: to reduce the risk of overfitting to specific highway routes and to assess model performance in more challenging and diverse urban driving scenarios, which are the most common in daily life, such as commuting and shopping.

Our methodology isolates urban driving segments from the collected data. First, we identify the longest continuous segment in each trip where the vehicle's speed (smoothed over 180 s) remains below 70 km/h. This threshold reflects the highway speeds recorded in the dataset. These segments are then refined by cropping to the points most representative of urban driving: starting where the vehicle's speed drops to zero and ending where it does not exceed 60 km/h. This 60 km/h threshold effectively distinguishes urban from highway driving. Examples of these cropped urban driving segments are shown in Fig. 8.

The quantitative results of Experiment 2 are shown in Table 2. (The training and validation loss curve can be found in Appendix A.5). Our proposed SOH-TEC model demonstrates a remarkable performance and outperforms all other models significantly across all metrics. Fig. 9 illustrates the comparison of the SOH estimations between the proposed SOH-TEC model and the top three baseline models (LSTM, CNN, and LightGBM) in the urban driving setting. Full scatter plots of the test results can be found in Appendix A.6. Due to the complex and shorter input data of the urban driving setting, the baseline models exhibit high dispersion in their estimations, highlighting their limitations under such demanding conditions. In contrast, our SOH-TEC model demonstrates remarkable resilience and maintains high performance, consistently delivering accurate SOH estimations across varied trip

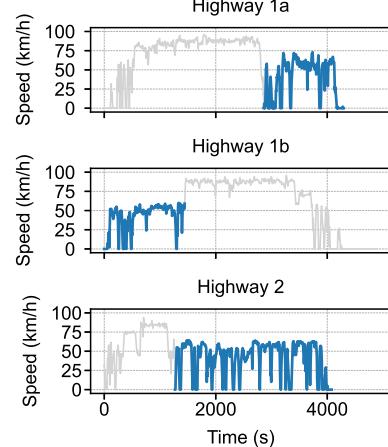


Fig. 8. Example of the urban trip segments extracted from the highway route trips. Blue lines are the urban trip parts that are used for the input data of Experiment 2.

routes. This robustness highlights the advanced capability of SOH-TEC to effectively handle the complexity of real-world urban driving conditions.

The comparative analysis of Experiments 1 and 2 highlights the superior performance of the SOH-TEC model. In Experiment 1, SOH-TEC outperformed other models, showing a clear advantage in the base setting with mixed driving segments. In Experiment 2, focusing on complex urban driving, SOH-TEC maintained high performance with minimal accuracy degradation. Specifically, while the second-best performing LSTM model's RMSE and MAE increased by 0.33 and 0.14 respectively, SOH-TEC's increases were only 0.13 and 0.05. This underscores SOH-TEC's robustness and its ability to effectively handle complex urban driving data for SOH estimation.

Table 2

Comparison of SOH estimation performance for different models in Experiment 2: urban driving setting. The best-performing model is highlighted in bold.

Model	RMSE↓	MAE↓	$R^2 \uparrow$
Dummy Regressor	6.822	5.459	-0.004
Route Aware Dummy Regressor	4.454	3.492	0.572
Linear Regression	2.826	2.187	0.828
Ridge Regression	2.895	2.229	0.819
Random Forest	2.488 (± 0.03)	1.765 (± 0.02)	0.866 (± 0.00)
LightGBM	2.200 (± 0.02)	1.533 (± 0.01)	0.896 (± 0.00)
MLP	3.592 (± 0.10)	2.577 (± 0.10)	0.722 (± 0.02)
CNN	2.412 (± 0.06)	1.702 (± 0.04)	0.874 (± 0.01)
LSTM	2.131 (± 0.09)	1.268 (± 0.07)	0.902 (± 0.01)
SOH-TEC (Ours)	1.295 (± 0.14)	0.767 (± 0.06)	0.963 (± 0.01)

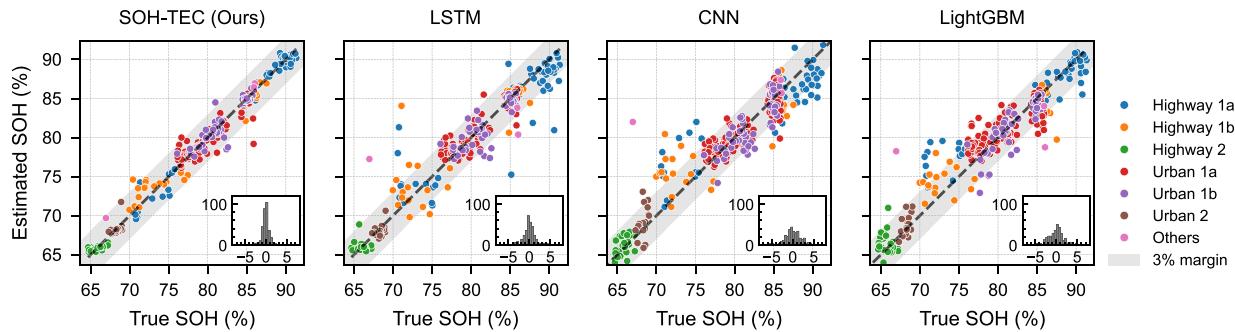


Fig. 9. The estimation results of Experiment 2: urban driving setting. The scatter plot illustrates true SOH vs. estimated SOH by the best-performing models. Marker colors represent different trip routes, and the gray region illustrates the MAE within a 3% error margin. The insets show the error distributions for each model.

6.3. Experiment 3: pre-training and fine-tuning using fewer labeled data

In this experiment, we investigate the effectiveness of our two proposed pre-training methods leveraging unlabeled data: masked token prediction and SOH ordinal comparison, detailed in Section 3.4. These pre-training approaches are particularly crucial given the considerable cost and time of acquiring SOH labels for real-world EVs.

To simulate realistic scenarios where only a limited amount of SOH-labeled data is available, we remove SOH labels from some of the training data used in Experiments 1 and 2. An illustration of the dataset configuration for Experiment 3 is shown in Fig. 10. Here, only α of the training split of the original dataset is still used as the labeled training dataset, and $1 - \alpha$ is the SOH unlabeled training dataset used for pre-training. The unlabeled data are randomly selected. The validation and test dataset remains the same as in Experiments 1 and 2. While this experimental setup provides valuable insights into the effectiveness of models and pre-training strategies in a controlled scenario with mixed labeled and unlabeled data, it should be noted that the setup may not fully capture the complexities of real-world environments. We explore these limitations and discuss potential avenues for future research in further detail in Section 7.4.

We compare the performance of models trained with different combinations of the pre-training strategies against models trained without any pre-training. We evaluate the performance of the models using the RMSE, MAE, and R^2 over $\alpha \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. The model is trained over eight different random seeds for each combination of the pre-training strategies and α values. The training and validation loss curves for the fine-tuning stage of the experiment are presented in Appendix A.5.

Fig. 11 shows the average RMSE and MAE versus α for the three different pre-training strategies and the no pre-training scenario. (The comprehensive results of the experiment are summarized in Appendix A.7.) The results clearly indicate that pre-training significantly enhances model performance, particularly when the amount of

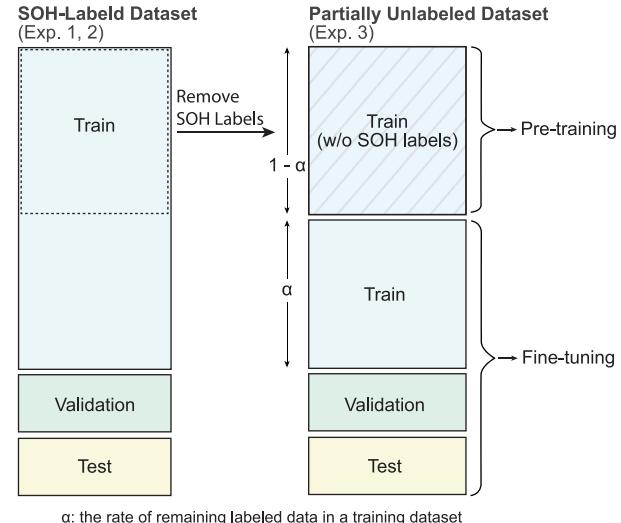


Fig. 10. Illustration of the datasets used in Experiment 3. A randomly selected portion α (where $\alpha < 1$) of the original dataset's training split is used as the labeled training dataset for supervised learning of SOH estimation. The remaining $1 - \alpha$ portion is used as the unlabeled SOH training dataset for pre-training purposes.

labeled data (or α) is low. This highlights the effectiveness of the pre-training approach leveraging unlabeled data to improve the model's capability in the downstream task of SOH estimation. For example, when $\alpha = 0.1$, SOH ordinal comparison pre-training enhances the model's performance by 36.4% for RMSE and 47.8% for MAE compared to the model without pre-training and Masked Token Prediction pre-training improves it by 14.2% for RMSE and 17.0% for MAE.

Among the pre-training strategies, SOH ordinal comparison consistently outperformed masked token prediction across different values of α . The combination of masked token prediction and SOH ordinal

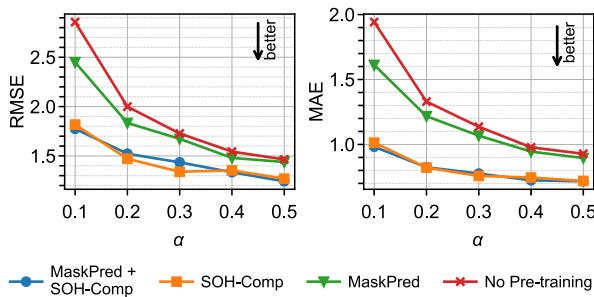


Fig. 11. RMSE and MAE of SOH estimation models for different α values. α is the ratio of the labeled training data to the total training data, meaning that the smaller α , the less SOH-labeled data is used for learning the model on the task of SOH estimation while the model can use more data for pre-training. MaskPred and SOH-Comp stand for masked token prediction and SOH ordinal comparison, respectively.

comparison performed similarly to SOH ordinal comparison alone, suggesting that masked token prediction does not provide additional benefits when used in conjunction with SOH ordinal comparison.

A noteworthy finding is that the model pre-trained with both masked token prediction and SOH ordinal comparison, utilizing only 50% of the labeled data ($\alpha = 0.5$), achieved performance nearly identical to the model trained on the full labeled dataset. Specifically, the RMSE and MAE for this pre-trained model are $1.243 (\pm 0.14)$ and $0.716 (\pm 0.05)$, while the RMSE and MAE for the model trained on the full labeled dataset are $1.167 (\pm 0.17)$ and $0.719 (\pm 0.04)$, respectively. This quantitative result underscores the significant impact of pre-training in drastically reducing the dependency on labeled data while maintaining high performance, highlighting its potential to make SOH estimation more efficient and cost-effective.

7. Discussion and limitations

7.1. Analyzing the self-attention map

Understanding the interpretative mechanisms of our model is crucial for reliable SOH estimation from EV operation data. The self-attention mechanism of the model differentially weights input features, highlighting the most relevant information for estimating SOH. In our context, the CLS token aggregates the sequence representation upon which the final SOH estimation is based. We analyze the self-attention weights assigned to this CLS token in the last transformer layer across various inputs to understand the model's focus under different driving conditions.

Fig. 12 illustrates the self-attention maps for different trip routes. The heatmaps visually represent the model's attention intensity, demonstrating distinct patterns for different driving environments. Notably, the heatmap from the highway trip in Fig. 12a shows a high concentration of attention by heads such as Head 1 and Head 5 on segments where the EV maintains consistent speed. Fig. 12b reveals that during mixed highway and city driving, attention shifts towards periods of stationary behavior, particularly in Head 2, Head 4, and Head 9, while Head 14 is focused on the consistent driving segments. In urban driving scenarios, as illustrated in Fig. 12c, the self-attention mechanism notably adjusts its focus, with Head 3 and Head 15 predominantly attending to the stationary segments where the EV's speed is zero.

These self-attention map analysis results reveal the model's concentration on data points from consistent driving or stationary phases, indicating that these patterns are particularly informative for SOH estimation. This insight can guide the refinement of data collection strategies to focus on the most impactful driving conditions and help hypothesize the crucial features for future algorithm enhancements. Additionally, the observed inter-head variability suggests that each

head captures different driving data aspects, a design feature that facilitates a thorough analysis of input features. Further research into these differences could lead to more specialized model architectures or input data selection strategies, enhancing the model's predictive accuracy.

7.2. SOH-TEC ablation study

To systematically understand the influence of various hyperparameters (and feature selection) on the performance of our transformer model, we conducted an extensive ablation study. We experimented with different configurations, varying one at a time while keeping others constant. This methodology allowed us to isolate and assess its specific impact on the model's accuracy in estimating SOH, as measured by RMSE. For consistency, this ablation study replicated the training and dataset setup from Experiment 1 (Section 6.1).

First, we examined five critical hyperparameters: the number of transformer blocks (N), the hidden dimension size (d), the kernel size of the first embedding layer (k_1), the number of self-attention heads (H), and the drop-path rate ($p_{\text{drop-path}}$). Fig. 13 illustrates the average RMSE across five random seeds with 95% confidence intervals, revealing key insights into hyperparameter performance. The model achieves the lowest RMSE with six transformer blocks, while the performance remains robust from two to eight blocks. For the hidden dimension size, performance peaks at 512, indicating that larger dimensions do not enhance accuracy. The kernel size of the first embedding layer optimizes at 8, suggesting that overly detailed or overly general representations are less effective. The number of attention heads shows a nearly monotonic decrease in RMSE with increasing heads. A moderate drop-path rate slightly enhances results, indicating the benefits of balanced regularization. This analysis reveals that while each hyperparameter exerts an influence on model performance, even the least optimal hyperparameter configuration in our study outperforms the baseline models, underlining the robustness of the SOH-TEC model against variations in hyperparameter settings.

The ablation study in Fig. 14 examines data augmentation and input feature effects on model performance. For data augmentation (Fig. 14a), we varied the random masking ratio p_{mask} and minimum length ratio λ . Optimal RMSE results occur at moderate levels of both parameters, with the lowest at $p_{\text{mask}} = 0.1$ and $\lambda = 0.8$, followed by $p_{\text{mask}} = 0.2$ and $\lambda = 0.9$. This suggests moderate data augmentation slightly improves performance. For input features (Fig. 14b), models including speed features marginally outperform those without. This result indicates that speed data, unique to the EV context, contributes valuable information for SOH estimation in EVs.

7.3. Dataset limitation and insights for future dataset creation

While the proposed method attains satisfactory accuracy, a noteworthy limitation of our study hinges on the limited variety and repetitiveness of driving routes in the dataset. This may cause the model to mistakenly correlate specific driving patterns to SOH, potentially preventing accurate learning of the relationship between EV operation data and SOH. For example, Highway 2 and Urban 2 routes were only driven during a short period toward the end of the data collection, which could lead to unintended associations. To improve and further confirm the model's performance and robustness, it is essential to incorporate a broader variety of driving scenarios that encompass a wide spectrum of environmental and operational conditions. Future iterations should implement a well-planned data collection strategy focused specifically on SOH estimation. We propose the following strategies to enhance dataset quality:

- **Avoid spurious correlations:** The dataset should be designed to minimize misleading correlations between SOH and incidental factors such as specific routes or SOC patterns.

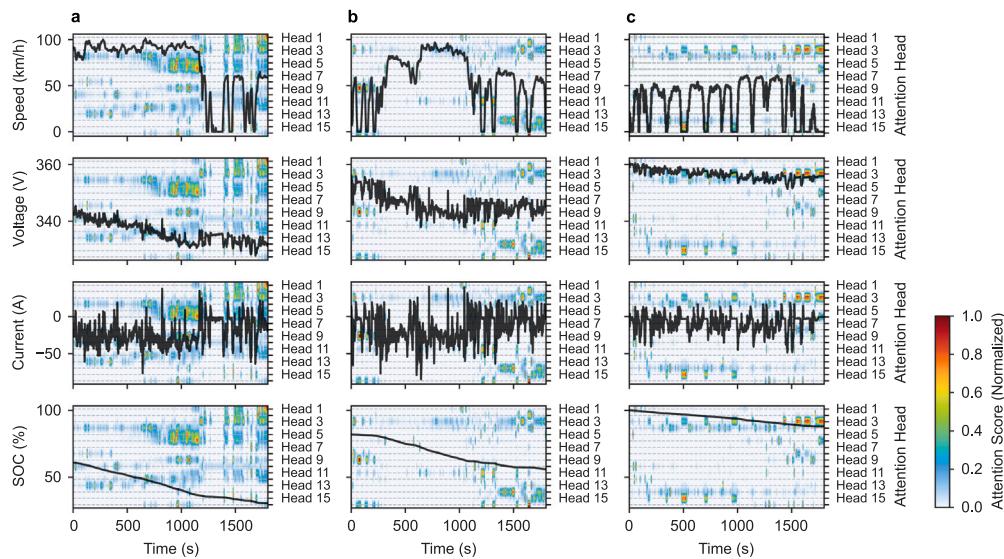


Fig. 12. Attention score corresponding to CLS token demonstrated by heatmap on the input sequence. (a) Highway driving input, (b) Mixed highway and city driving input, (c) City driving input.

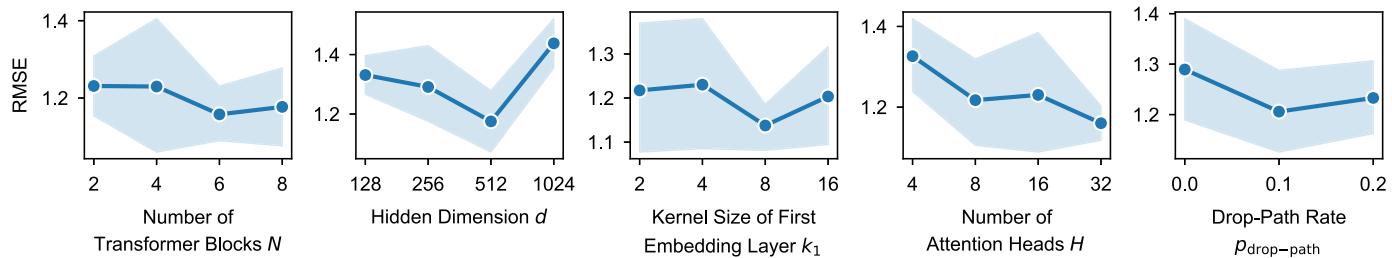


Fig. 13. Hyperparameter ablation study results of the transformer model. The plot line shows the average RMSE across the five random seeds for each hyperparameter setting, with the shaded region indicating the 95% confidence interval of the RMSE.

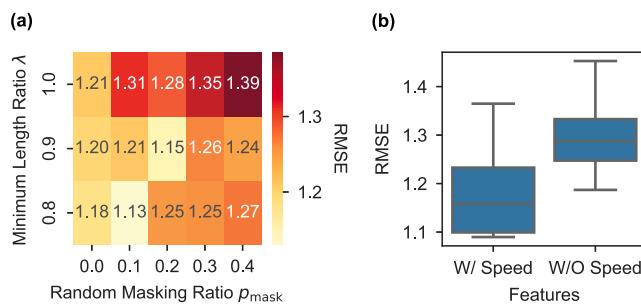


Fig. 14. Ablation study results. (a) The impact of data augmentation parameters on RMSE. The heatmap shows mean RMSE values over five random seeds for different combinations of random masking ratio p_{mask} and minimum length ratio λ . (b) Comparison of RMSE for models trained with and without speed features. The box plots represent the distribution of RMSE values over the eight random seeds.

- **Diverse driving behaviors:** The dataset should include driving patterns that are comprehensive enough to cover the entire spectrum of typical driving behavior.
- **Comprehensive SOH range:** The dataset should encompass a broad range of SOH values to ensure the model can accurately predict SOH across different stages of battery life.
- **Consistent driving:** The vehicle should be driven consistently throughout the data collection period so that battery degradation

is consistent and the linearly interpolated SOH based on the accumulated mileage can be justified.

Considering these elements, one potential strategy is to diversify the driving routes. The driven routes should be randomly chosen or systematically repeated to eliminate spurious correlations between driving patterns and SOH. For instance, a dataset could include 10 distinct urban routes that collectively represent typical urban driving patterns. Each route would be driven 50 times spread evenly over the course of a year to cover a wide range of SOH values. Ideally, data collection should span multiple years to mitigate seasonal effects. Since our model performs well with 30-minute operational data inputs, each route in the dataset can be relatively short. This approach ensures that the dataset includes a sufficient number of diverse instances to support robust model training and evaluation.

Another potential data collection strategy involves gathering driving data from a large number of corporate or privately owned EVs and conducting periodic RPTs. This approach, while promising, presents several challenges. First, there are high costs associated with collecting data and conducting RPTs. Second, data collection raises privacy concerns. Third, uncontrolled driving intervals necessitate devising a reliable method for SOH interpolation. Despite these challenges, this strategy could yield a dataset with diverse real-world driving scenarios.

7.4. Experiment setup limitation and further potential of pre-training

Considering real-world conditions where obtaining comprehensive SOH labels is often impractical, pre-training with unlabeled data is an important direction for real-world EV SOH estimation. Experiment

3 (in Section 6.3) was designed to confirm the effectiveness of this pre-training approach by employing a controlled mix of labeled and unlabeled data, aiming to replicate situations where labeled data is scarce. The results revealed that the proposed pre-training strategies leveraging unlabeled data enhanced the model's performance in the downstream SOH estimation task.

However, it is important to recognize the limitations of our experimental setup. Specifically, the random selection of unlabeled data from a relatively small dataset, while useful for initial validation, may not accurately represent real-world scenarios where diverse and vast amounts of unlabeled data are available. To build on the current work, future experiments could explore larger and more diverse SOH unlabeled datasets. This could enhance the model's performance and robustness across different driving conditions. Additionally, the flexibility of our model and its pre-training methods hints at the potential for adaptation across different domains and vehicle types. Incorporating data from multiple domains and vehicle types would be another important direction for future work to broaden the scope of its applicability.

8. Conclusion

This research introduces SOH-TEC, a transformer-based model for estimating the State of Health (SOH) of electric vehicle (EV) batteries using raw single-trip operation data. We have created a real-world EV-driving dataset with reliable SOH labels, collected over nearly three years from three EVs in Japan, including periodic reference performance tests for accurate labeling. Our experiments demonstrate SOH-TEC's superior performance compared to baseline models, with a root mean square error of 1.17 and a mean absolute error of 0.72 in the base setting. The model maintains robust performance even in challenging urban driving scenarios. Self-attention map analysis reveals the model's focus on stationary or consistent driving periods as key indicators for SOH estimation. Pre-training strategies, particularly SOH ordinal comparison, significantly improve performance when labeled data is limited. With only 50% labeled data, the pre-trained model performs comparably to training on the full labeled dataset.

However, the model's performance is constrained by the repetitiveness and limited variety of driving routes in the dataset. Future work should expand the dataset to include more diverse driving scenarios. Additionally, our model currently focuses on the vehicle driving session (discharging process), but integrating charging processes could provide a more holistic view of battery health and enhance model robustness and accuracy. This study confirms the effectiveness of transformer-based models for EV SOH estimation using real-world data. We hope this work paves the way for further research in EV battery management and supports the transition to cleaner energy solutions.

CRediT authorship contribution statement

Kosaku Nakano: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Sophia Vögler:** Writing – review & editing, Writing – original draft, Investigation. **Kenji Tanaka:** Writing – review & editing, Supervision, Data curation, Conceptualization.

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.

Data availability

The data that has been used is confidential. The code will be made available on request.

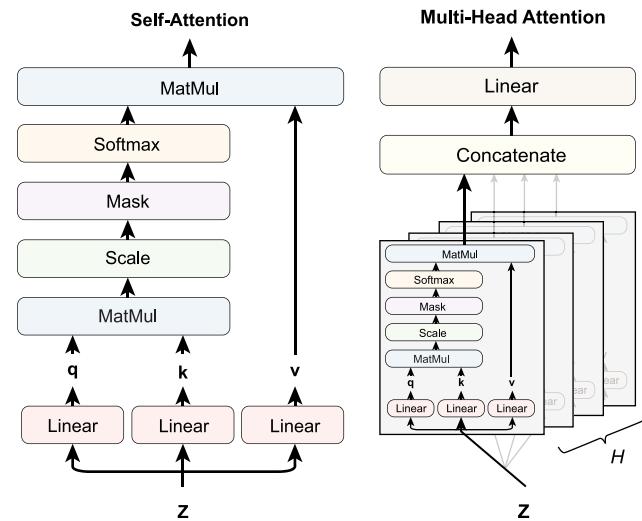


Fig. A.15. A graphical representation of the self-attention mechanism and multi-head attention.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT and Claude in order to improve the clarity of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgment

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We thank the Japan Automobile Research Institute for providing the data.

Appendix A. Supplementary information

A.1. Attention mechanism

Fig. A.15 illustrates the SA and MHA architecture. Each self-attention head computes a different attention score from query, key, and value. Query q , key k , and value v are mapped from the input z by linear projection respectively (Eq. (A.2) to Eq. (A.4)). The self-attention score is computed by the dot product of query and key, followed by a softmax function. Before the softmax is applied, the dot product is divided by the square root of the dimension of the key d_h . The self-attention score is then used to compute the weighted sum of the value vectors (Eq. (A.1)) [31].

$$SA(z) = \text{softmax} \left(\frac{qk^T}{\sqrt{d_h}} \right) v, \quad (A.1)$$

$$\text{where } q = zW_q \quad W_q \in \mathbb{R}^{d \times d_h}, \quad (A.2)$$

$$k = zW_k \quad W_k \in \mathbb{R}^{d \times d_h}, \quad (A.3)$$

$$v = zW_v \quad W_v \in \mathbb{R}^{d \times d_h}. \quad (A.4)$$

MHA is calculated by concatenating the output of H self-attention heads and passing them through a linear layer:

$$MHA(z) = \text{Concat}(SA_1(z), \dots, SA_H(z))W_o, \quad (A.5)$$

where $W_o \in \mathbb{R}^{Hd_h \times d}$ represents the learnable weight matrix for the output linear layer. Here, H denotes the number of self-attention heads, and d_h indicates the dimension of each self-attention head.

Table A.3

Number of trips across different routes included in the dataset.

Route	Number of trips
Highway 1a	484
Highway 1b	320
Highway 2	187
Urban 1a	1701
Urban 1b	406
Urban 2	208
Others	39

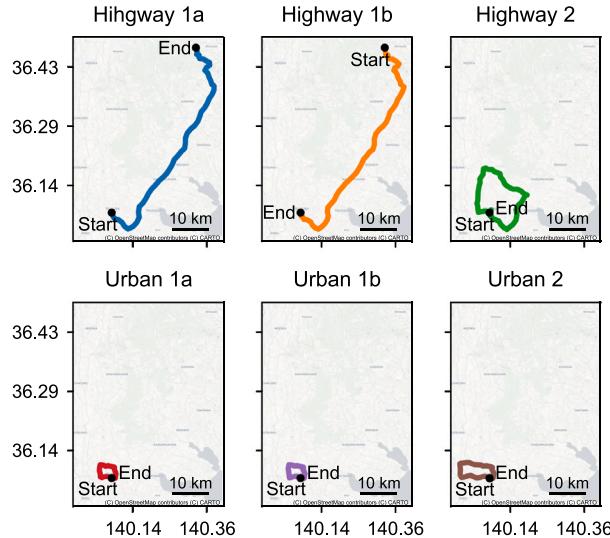


Fig. A.16. Six trip routes visualization.³

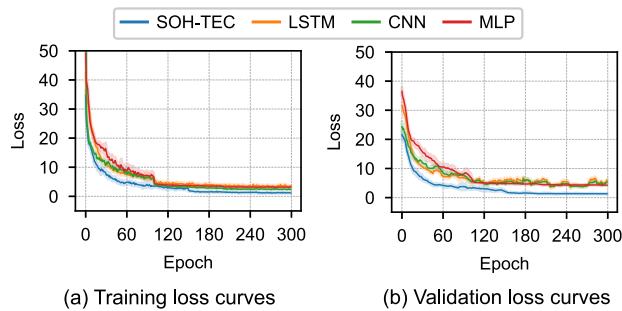


Fig. A.17. Training and validation loss curves for Experiment 1. Both curves have their vertical axes truncated at 50.0. The validation loss curves are smoothed using a moving average over 11 epochs.

A.2. Trip routes

Table A.3 shows the number of trips for each route pattern included in the dataset. Fig. A.16 illustrates the geographical layout of the trip routes.

A.3. Comparative baseline models

In this section, we provide a detailed description of the comparative baseline models used in our study. All the deep learning-based models

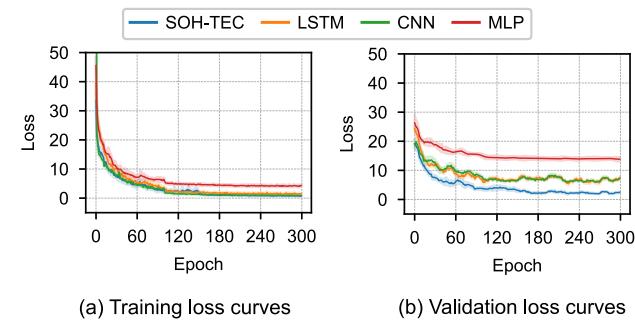


Fig. A.18. Training and validation loss curves for Experiment 2. Both curves have their vertical axes truncated at 50.0. The validation loss curves are smoothed using a moving average over 11 epochs.

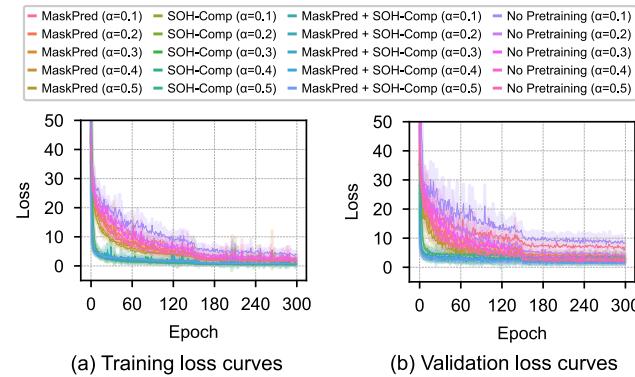


Fig. A.19. Training and validation loss curves for SOH estimation task in Experiment 3. Both curves have their vertical axes truncated at 50.0.

are implemented in PyTorch [71], and automatic mixed precision with fp16 was used for training and inference.

A.3.1. Dummy regressor

A Dummy Regressor outputs the average SOH of the training dataset as the predicted SOH. A modified version, the Route-aware Dummy Regressor, outputs the average SOH for each trip route. Given the dataset's repetitive and limited trip routes, the Route-aware Dummy Regressor serves as a sufficiently strong baseline.

A.3.2. MLP

The model referred to as MLP for simplicity, includes a convolutional embedding layer similar to SOH-TEC, followed by fully connected layers. The input sequence first passes through a convolutional layer (kernel size 8, stride 2) mapping it to 256 channels. This is followed by three fully connected layers with output dimensions of 1024, 256, and 1. ReLU activations are applied after the first two layers. The final output predicts the SOH.

A.3.3. CNN

The CNN model is composed of multiple convolutional layers, each followed by BN and ReLU activation functions. The detailed configuration of the model is shown in Table A.4. The final output (estimated SOH) is produced by a linear layer that takes the flattened output from the last convolutional block as an input. No activation function is applied to the final layer.

A.3.4. LSTM

The LSTM baseline uses a BN layer, followed by a 1D convolutional layer (512 output features, kernel size 32, stride 8) to reduce sequence length. After LN, features are processed by a bidirectional LSTM (hidden size 768), then mapped to a single SOH prediction via a linear

³ Map data from OpenStreetMap (openstreetmap.org/copyright) and CARTO (<https://carto.com/>).

Table A.4
Convolutional layers of the CNN model.

No.	Input channels	Output channels	Kernel size	Stride
0	4	256	4	2
1	256	256	4	2
2	256	256	4	2
3	256	128	16	1
4	128	128	4	1

Table A.5
Summary of hyperparameters for models training.

Model	Batch size	Grad clip max norm	Sequence crop range	Learning rate		
				Base	Decay step size	γ
MLP	16	5	(1800, 1800)	1e-4	100	0.1
LSTM	32	1	(2400, 2400)	5e-5	100	0.1
CNN	16	—	(1920, 2400)	1e-4	100	0.1
SOH-TEC (Ours)	32	10	(1620, 1800)	5e-5	150	0.2

Table A.6
Parameter search configuration of SOH-TEC.

Parameter	Values
Attention Drop Rate $p_{\text{attn-dropout}}$	0, 0.1, 0.2
Batch Size	16, 32
Number of Transformer Blocks N	4, 6, 8, 10
Drop Path Rate $p_{\text{drop-path}}$	0, 0.1
FFN Drop Rate $p_{\text{ffn-dropout}}$	0, 0.1, 0.2
Embedding Dimension d	256, 512
Gradient Clipping Max Norm	null, 3, 5, 10
Learning Rate	1e-4, 5e-5, 1e-5, 5e-6
Random Masking Ratio Range	(0, 0), (0, 0.1), (0, 0.2), (0.1, 0.2), (0.1, 0.3)
Random Masking Type	arbitrary, continuous
Minimum Length Ratio λ	1, 0.9, 0.8, 0.6
Number of Embed Layers L	1, 2, 3, 5
Number of Heads H	16
Kernel Size of First Embedding Layer k_1	4, 8, 16
Position Drop Rate $p_{\text{pos-dropout}}$	0, 0.1
Pre-embed Norm	LayerNorm, BatchNorm, InstanceNorm, null
Sequence Length S	1200, 1500, 1800, 2100, 2400
Learning Rate Decay Gamma	0.1, 0.2
Learning Rate Decay Step Size	100, 150
Stride of First Embedding Layer s_1	2, 4

Table A.7
Parameter search configuration of MLP.

Parameter	Values
Batch Size	16, 32, 64
Dropout Rate	0, 0.1, 0.2
Embedding Dimension d	128, 256, 512, 768
Gradient Clipping Max Norm	null, 1, 3, 5, 10
Learning Rate	1e-4, 5e-5, 1e-5, 5e-6
Minimum Length Ratio λ	1, 0.9, 0.8, 0.6
First Hidden Layer Dimension	256, 512, 1024, 2048
Second Hidden Layer Dimension	256, 512, 1024, 2048
Patch Size	4, 8, 16, 32
Pre-embed Norm	LN, LN1D, BatchNorm, InstanceNorm, null
Sequence Length S	1200, 1500, 1800, 2100, 2400
Learning Rate Decay Gamma	0.1, 0.2
Learning Rate Decay Step Size	100, 150
Stride of First Embedding Layer s_1	1, 2, 4, 8

layer. This architecture is often referred to as CNN-LSTM, which is widely adopted for battery state estimation tasks.

A.3.5. Feature-based models

In addition to the deep learning models, we also evaluate the performance of four feature-based models: Linear Regression, Ridge Regression, Random Forest, and LightGBM, which were found to be

Table A.8
Parameter search configuration of CNN.

Parameter	Values
Batch Size	16, 32
Convolutional Block 1 Channels	128, 256, 512
Convolutional Block 1 Kernel Size	2, 4, 8, 16
Convolutional Block 1 Stride	1, 2, 4
Convolutional Block 2 Channels	64, 128, 256, 512
Convolutional Block 2 Kernel Size	2, 4, 8, 16
Convolutional Block 2 Stride	1, 2, 4
Convolutional Block 3 Channels	32, 64, 128, 256
Convolutional Block 3 Kernel Size	2, 4, 8, 16
Convolutional Block 3 Stride	1, 2, 4
Batch Norm First	True, false
Number of Convolutional Blocks 1	1, 2, 3
Number of Convolutional Blocks 2	1, 2, 3
Number of Convolutional Blocks 3	0, 1, 2, 3
Dropout rate	0, 0.1, 0.2
Gradient Clipping Max Norm	Null, 1, 3, 5, 10
Learning Rate	1e-4, 5e-5, 1e-5, 5e-6
Minimum Length Ratio λ	1, 0.9, 0.8, 0.6
Sequence Length S	1200, 1500, 1800, 2100, 2400
Learning Rate Decay Gamma	0.1, 0.2
Learning Rate Decay Step Size	100, 150

Table A.9
Parameter search configuration of LSTM.

Parameter	Values
Batch Size	16, 32, 64
Dropout Rate	0, 0.1, 0.2
Embedding Dimension d	128, 256, 512, 768
Gradient Clipping Max Norm	null, 1, 3, 5, 10
Learning Rate	1e-4, 5e-5, 1e-5, 5e-6
LSTM Bidirectional	true, false
LSTM Hidden Size	128, 256, 512, 768
LSTM Number of Layers	1, 2, 3, 4
Minimum Length Ratio λ	1, 0.9, 0.8, 0.6
Number of Embed Layers L	1
Patch Size	4, 8, 16, 32
Pre-embed Norm	LN, LN1D, BatchNorm, InstanceNorm
Sequence Length S	1200, 1500, 1800, 2100, 2400
Learning Rate Decay Gamma	0.1, 0.2
Learning Rate Decay Step Size	100, 150
Stride of First Embedding Layer s_1	1, 2, 4, 8

the best-performing models for SOH estimation for lab-experimented cell data [72]. These models utilize a comprehensive set of features extracted from the dataset to predict the SOH. The features include statistical metrics such as mean, standard deviation, minimum, maximum, skewness, and kurtosis for speed, voltage, current, and SOC. Additionally, incremental capacity (IC) is calculated, along with voltage differences across SOC ranges, mean voltage to mean current ratios, and voltage range. Voltage and SOC values at maximum incremental capacity, trip duration, and various discharge metrics (amp-hours and kilowatt-hours) are also included. It should be noted that feature extraction is more challenging on real-world driving data than on lab-experimented data due to the fluctuations and variability inherent in real-world conditions. These fluctuations can introduce noise and make it more difficult to extract meaningful features, thereby impacting the performance of feature-based models.

A.4. Hyperparameters

The training hyperparameters for the deep learning-based baseline models (MLP, CNN, and LSTM) and our proposed model (SOH-TEC) are summarized in Table A.5. During training, each sequence is cropped at a randomly selected starting point to a length that is uniformly sampled from a sequence crop range (same as random cropping explained in Section 3.3). The learning rate is decayed by γ for every decay step size. For example, in MLP, the learning rate starts with 1e-4, and after every 100 epoch, it is decayed by 0.1.

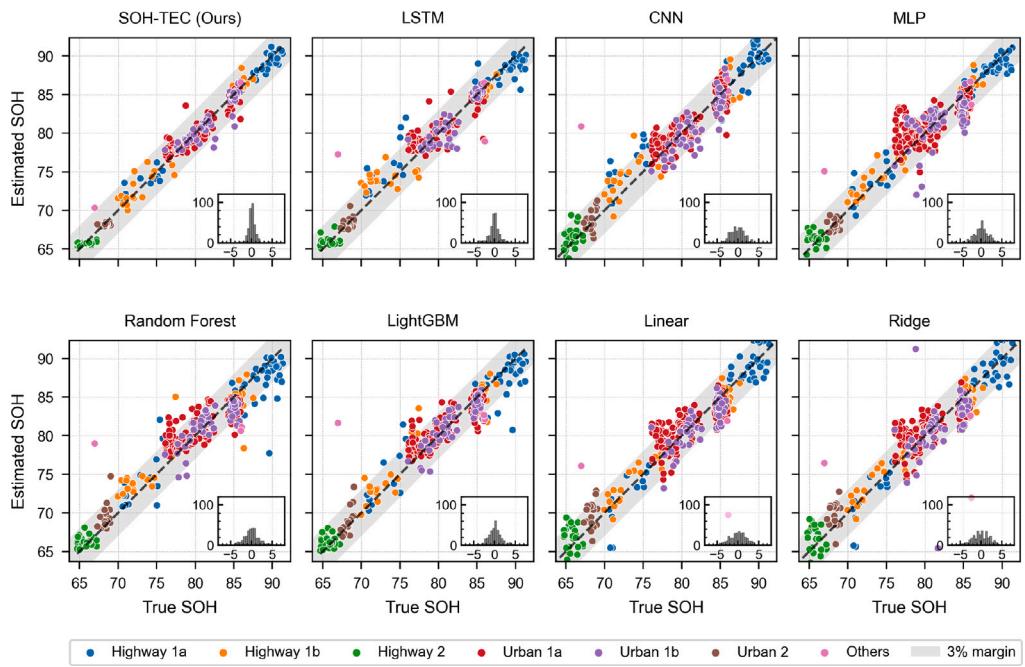


Fig. A.20. The estimation results of Experiment 1: base setting. The scatter plot illustrates true SOH vs. estimated SOH by the best-performing models for the test dataset. Marker colors represent different trip routes, and the gray region illustrates the 3% error margin. The insets show the error distributions for each model.

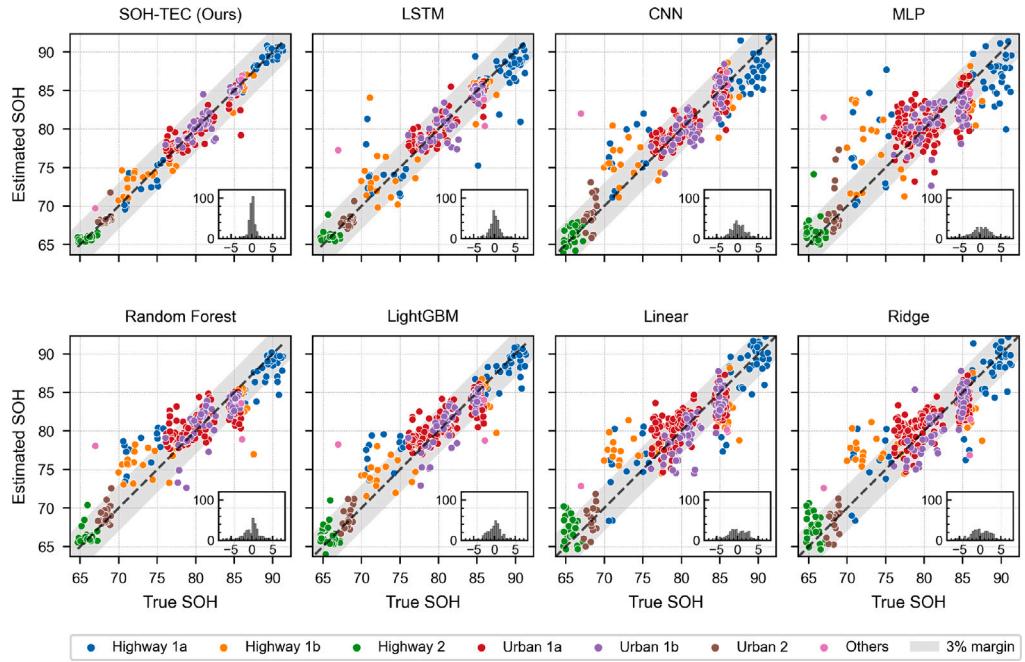


Fig. A.21. The estimation results of Experiment 2: urban driving setting. The scatter plot illustrates true SOH vs. estimated SOH by the best-performing models for the test dataset. Marker colors represent different trip routes, and the gray region illustrates the 3% error margin. The insets show the error distributions for each model.

These hyperparameters and others regarding the model architecture were determined by a random search for each model using training and validation datasets. Tables A.6–A.9 show the hyperparameter search configuration of SOH-TEC, MLP, CNN, and LSTM, respectively. We employed random search with Hyperband [73] early stopping. The maximum epoch was set to 200. The random masking type of SOH-TEC can be either ‘arbitrary’ or ‘continuous.’ When set to ‘arbitrary,’ the masking locations within the sequence are chosen randomly. When set to ‘continuous,’ the masking is applied over a continuous subsequence of the input sequence. The sequence length S and the minimum length

ratio λ are used to determine the length range to which the original input sequence is cropped, where the length is sampled from $U(\lambda S, S)$. The pre-embed norm refers to the normalization method before the embedding module. The position drop rate $p_{\text{pos-dropout}}$ in SOH-TEC is a dropout in the positional encoding layer, which is set to 0 in this study.

A.5. Training and validation loss curves

Fig. A.17, Fig. A.18, and Fig. A.19 show the training and validation loss curves for Experiments 1, 2, and 3, respectively.

Table A.10

SOH estimation results with different pre-training strategies over different α values. α is the ratio of the labeled training data to the total training data.

Model	Pre-training		$\alpha = 0.1$			$\alpha = 0.2$			$\alpha = 0.3$			$\alpha = 0.4$			$\alpha = 0.5$		
	MaskPred ^a	SOH-Comp ^b	RMSE \downarrow	MAE \downarrow	$R^2 \uparrow$	RMSE \downarrow	MAE \downarrow	$R^2 \uparrow$	RMSE \downarrow	MAE \downarrow	$R^2 \uparrow$	RMSE \downarrow	MAE \downarrow	$R^2 \uparrow$	RMSE \downarrow	MAE \downarrow	$R^2 \uparrow$
SOH-TEC	—	—	2.857 (± 0.24)	1.942 (± 0.17)	0.823 (± 0.03)	2.000 (± 0.15)	1.330 (± 0.07)	0.913 (± 0.01)	1.727 (± 0.13)	1.136 (± 0.02)	0.935 (± 0.01)	1.544 (± 0.13)	0.978 (± 0.04)	0.948 (± 0.01)	1.464 (± 0.13)	0.928 (± 0.05)	0.953 (± 0.01)
	✓	—	2.450 (± 0.21)	1.612 (± 0.12)	0.870 (± 0.02)	1.834 (± 0.12)	1.217 (± 0.09)	0.927 (± 0.01)	1.673 (± 0.12)	1.068 (± 0.05)	0.939 (± 0.01)	1.480 (± 0.20)	0.945 (± 0.07)	0.952 (± 0.01)	1.439 (± 0.14)	0.896 (± 0.06)	0.955 (± 0.01)
	—	✓	1.818 (± 0.27)	1.014 (± 0.12)	0.927 (± 0.02)	1.471 (± 0.11)	0.822 (± 0.04)	0.953 (± 0.01)	1.339 (± 0.08)	0.758 (± 0.05)	0.961 (± 0.00)	1.354 (± 0.20)	0.745 (± 0.06)	0.960 (± 0.01)	1.270 (± 0.20)	0.719 (± 0.06)	0.964 (± 0.01)
	✓	✓	1.775 (± 0.18)	0.982 (± 0.10)	0.931 (± 0.01)	1.522 (± 0.12)	0.825 (± 0.04)	0.950 (± 0.01)	1.435 (± 0.11)	0.776 (± 0.02)	0.955 (± 0.01)	1.335 (± 0.16)	0.724 (± 0.06)	0.961 (± 0.01)	1.243 (± 0.14)	0.716 (± 0.05)	0.966 (± 0.01)

^a Masked token prediction.

^b SOH ordinal comparison.

A.6. Full scatter plots of test results

Fig. A.20 and Fig. A.21 show the full scatter plots of the estimated SOH vs. the true SOH for the test dataset of Experiment 1 and Experiment 2, respectively.

A.7. Comprehensive results of experiment 3

The comprehensive results of Experiment 3 are summarized in Table A.10. For each value of α , we report the average RMSE, MAE, and R^2 along with their standard deviations. These results further confirm the superior performance of the SOH ordinal comparison and the effectiveness of pre-training in enhancing model performance.

References

- [1] International Energy Agency. Tracking clean energy progress 2023. 2023, <https://www.iea.org/reports/tracking-clean-energy-progress-2023>. [Accessed: April 12, 2024].
- [2] International Energy Agency. Global EV outlook 2023. 2023, <https://www.iea.org/reports/global-ev-outlook-2023>. [Accessed: April 12, 2024].
- [3] Miao Y, Hynan P, Von Jouanne A, Yokochi A. Current Li-ion battery technologies in electric vehicles and opportunities for advancements. Energies 2019;12(6):1074.
- [4] Lu L, Han X, Li J, Hua J, Ouyang M. A review on the key issues for lithium-ion battery management in electric vehicles. J Power Sources 2013;226:272–88.
- [5] Saw LH, Ye Y, Tay AA. Integration issues of lithium-ion battery into electric vehicles battery pack. J Clean Prod 2016;113:1032–45.
- [6] Ding Y, Cano ZP, Yu A, Lu J, Chen Z. Automotive Li-ion batteries: current status and future perspectives. Electrochim Energy Rev 2019;2:1–28.
- [7] Han X, Lu L, Zheng Y, Feng X, Li Z, Li J, et al. A review on the key issues of the lithium ion battery degradation among the whole life cycle. ETransportation 2019;1:100005.
- [8] Barré A, Deguilhem B, Grolleau S, Gérard M, Suard F, Riu D. A review on lithium-ion battery ageing mechanisms and estimations for automotive applications. J Power Sources 2013;241:680–9.
- [9] Xiong R, Tian J, Mu H, Wang C. A systematic model-based degradation behavior recognition and health monitoring method for lithium-ion batteries. Appl Energy 2017;207:372–83.
- [10] Kaushubham, Koorata PK, Panchal S, Fraser R, Fowler M. Investigation of the thermal performance of biomimetic minichannel-based liquid-cooled large format pouch battery pack. J Energy Storage 2024;84:110928.
- [11] Huang S-C, Tseng K-H, Liang J-W, Chang C-L, Pecht MG. An online SOC and SOH estimation model for lithium-ion batteries. Energies 2017;10(4):512.
- [12] Bercicar M, Gandiaga I, Villarreal I, Omar N, Van Mierlo J, Van den Bossche P. Critical review of state of health estimation methods of Li-ion batteries for real applications. Renew Sustain Energy Rev 2016;56:572–87.
- [13] Ozkurt C, Camci F, Atamuradov V, Odorry C. Integration of sampling based battery state of health estimation method in electric vehicles. Appl Energy 2016;175:356–67.
- [14] Babin A, Rizoug N, Mesbahi T, Boscher D, Hamdoun Z, Larouci C. Total cost of ownership improvement of commercial electric vehicles using battery sizing and intelligent charge method. IEEE Trans Ind Appl 2018;54(2):1691–700.
- [15] Wang J, Yang H, Wang Z, Zhou Y, Liu P, Hong J. State of health analysis of batteries at different stages based on real-world vehicle data and machine learning. J Energy Storage 2024;88:111616.
- [16] Chen M, Ma X, Chen B, Arsenault R, Karlson P, Simon N, et al. Recycling end-of-life electric vehicle lithium-ion batteries. Joule 2019;3(11):2622–46.
- [17] Zhang C, Yan J, You F. Critical metal requirement for clean energy transition: a quantitative review on the case of transportation electrification. Adv Appl Energy 2023;9:100116.
- [18] Ku AY, Alonso E, Eggert R, Graedel T, Habib K, Hool A, et al. Grand challenges in anticipating and responding to critical materials supply risks. Joule 2024;8(5):1208–23.
- [19] Ungurean L, Cărstoiu G, Micea MV, Groza V. Battery state of health estimation: a structured review of models, methods and commercial devices. Int J Energy Res 2016;41(2):151–81.
- [20] Oji T, Zhou Y, Ci S, Kang F, Chen X, Liu X. Data-driven methods for battery soh estimation: Survey and a critical analysis. Ieee Access 2021;9:126903–16.
- [21] Zhao J, Han X, Ouyang M, Burke AF. Specialized deep neural networks for battery health prognostics: Opportunities and challenges. J Energy Chem 2023;87:416–38.
- [22] Gu X, See KW, Li P, Shan K, Wang Y, Zhao L, et al. A novel state-of-health estimation for the lithium-ion battery using a convolutional neural network and transformer model. Energy 2023;262:125501.
- [23] Bockrath S, Lorentz V, Pruckner M. State of health estimation of lithium-ion batteries with a temporal convolutional neural network using partial load profiles. Appl Energy 2023;329:120307.
- [24] Weng C, Feng X, Sun J, Peng H. State-of-health monitoring of lithium-ion battery modules and packs via incremental capacity peak tracking. Appl Energy 2016;180:360–8.
- [25] Pozzato G, Allam A, Pulvirenti L, Negoita GA, Paxton WA, Onori S. Analysis and key findings from real-world electric vehicle field data. Joule 2023;7(9):2035–53.
- [26] Zhao Y, Behdad S. State of health estimation of electric vehicle batteries using transformer-based neural network. J Energy Resour Technol 2024;146(10):101703.
- [27] Baure G, Dubarry M. Synthetic vs. real driving cycles: A comparison of electric vehicle battery degradation. Batteries 2019;5(2):42.
- [28] Wen S, Lin N, Huang S, Wang Z, Zhang Z. Lithium battery health state assessment based on vehicle-to-grid (V2G) real-world data and natural gradient boosting model. Energy 2023;284:129246.
- [29] Liu H, Deng Z, Yang Y, Lu C, Li B, Liu C, et al. Capacity evaluation and degradation analysis of lithium-ion battery packs for on-road electric vehicles. J Energy Storage 2023;65:107270.
- [30] Rezvanianian SM, Liu Z, Chen Y, Lee J. Review and recent advances in battery health monitoring and prognostics technologies for electric vehicle (EV) safety and mobility. J Power Sources 2014;256:110–24.
- [31] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Advances in neural information processing systems (neurIPS). 2017, p. 6000–10.
- [32] He Z, Ni X, Pan C, Hu S, Han S. Full-process electric vehicles battery state of health estimation based on informer novel model. J Energy Storage 2023;72:108626.
- [33] Hannan MA, How DN, Lipu MH, Mansor M, Ker PJ, Dong Z, et al. Deep learning approach towards accurate state of charge estimation for lithium-ion batteries using self-supervised transformer model. Sci Rep 2021;11(1):19541.
- [34] Singh S, Budarapu P. Deep machine learning approaches for battery health monitoring. Energy 2024;300:131540.
- [35] Zhang W, Jia J, Pang X, Wen J, Shi Y, Zeng J. An improved transformer model for remaining useful life prediction of lithium-ion batteries under random charging and discharging. Electronics 2024;13(8):1423.
- [36] Zhao W, Ding W, Zhang S, Zhang Z. Enhancing lithium-ion battery lifespan early prediction using a multi-branch vision transformer model. Energy 2024;302:131816.
- [37] Biggio L, Bendinelli T, Kulkarni C, Fink O. Ageing-aware battery discharge prediction with deep learning. Appl Energy 2023;346:121229.
- [38] Costa N, Ansén D, Dubarry M, Sánchez L. ICFormer: A deep learning model for informed lithium-ion battery diagnosis and early knee detection. J Power Sources 2024;592:233910.
- [39] Chen D, Zhou X. AttMoE: Attention with mixture of experts for remaining useful life prediction of lithium-ion batteries. J Energy Storage 2024;84:110780.

- [40] Wang W, Yang K, Zhang L, Zhou S, Ren B, Lu Y, et al. An end-cloud collaboration approach for online state-of-health estimation of lithium-ion batteries based on multi-feature and transformer. *J Power Sources* 2024;608:234669.
- [41] Fauzi MR, Yudistira N, Mahmudy WF. State-of-health prediction of lithium-ion batteries using exponential smoothing transformer with seasonal and growth embedding. *IEEE Access* 2024;12:14659–70.
- [42] Gomez W, Wang F-K, Chou J-H. Li-ion battery capacity prediction using improved temporal fusion transformer model. *Energy* 2024;296:131114.
- [43] Luo K, Zheng H, Shi Z. A simple feature extraction method for estimating the whole life cycle state of health of lithium-ion batteries using transformer-based neural network. *J Power Sources* 2023;576:233139.
- [44] Jia C, Tian Y, Shi Y, Jia J, Wen J, Zeng J. State of health prediction of lithium-ion batteries based on bidirectional gated recurrent unit and transformer. *Energy* 2023;285:129401.
- [45] Zhao H, Chen Z, Shu X, Shen J, Lei Z, Zhang Y. State of health estimation for lithium-ion batteries based on hybrid attention and deep learning. *Reliab Eng Syst Saf* 2023;232:109066.
- [46] Fan Y, Li Y, Zhao J, Wang L, Yan C, Wu X, et al. Online state-of-health estimation for fast-charging lithium-ion batteries based on a transformer-long short-term memory neural network. *Batteries* 2023;9(11):539.
- [47] Zou C, Chen X, Zhang Y. State of health prediction of lithium-ion batteries based on temporal degeneration feature extraction with deep cycle attention network. *J Energy Storage* 2023;65:107367.
- [48] Gao M, Shen H, Bao Z, Deng Y, He Z. A correlation-augmented informer-based method for state-of-health estimation of Li-ion batteries. *IEEE Sens J* 2023;24(3):3342–53.
- [49] Chen L, Xie S, Lopes AM, Bao X. A vision transformer-based deep neural network for state of health estimation of lithium-ion batteries. *Int J Electr Power Energy Syst* 2023;152:109233.
- [50] Bai T, Wang H. Convolutional transformer-based multi-view information perception framework for lithium-ion battery state-of-health estimation. *IEEE Trans Instrum Meas* 2023;72:1–12.
- [51] Shi D, Zhao J, Wang Z, Zhao H, Eze C, Wang J, et al. Cloud-based deep learning for co-estimation of battery state of charge and state of health. *Energies* 2023;16(9):3855.
- [52] Cai N, Qin Y, Chen X, Wu K. Dual time-scale state-coupled co-estimation of state of charge, state of health and remaining useful life for lithium-ion batteries via deep inter and intra-cycle attention network. *J Energy Storage* 2024;77:109797.
- [53] Ke L, Fang L, Meng J, Peng J, Wu J, Lin M, et al. Identification of the aging state of lithium-ion batteries via temporal convolution network and self-attention mechanism. *J Energy Storage* 2024;84:110999.
- [54] Tang A, Jiang Y, Yu Q, Zhang Z. A hybrid neural network model with attention mechanism for state of health estimation of lithium-ion batteries. *J Energy Storage* 2023;68:107734.
- [55] Sun Z, Fan G, Liu Y, Zhou B, Wang Y, Chen S, et al. Interactive fusion of local and global degradation representations for rapid estimation of lithium-ion battery state-of-health. *J Energy Storage* 2024;90:111832.
- [56] Chen X, Qin Y, Zhao W, Yang Q, Cai N, Wu K. A self-attention knowledge domain adaptation network for commercial lithium-ion batteries state-of-health estimation under shallow cycles. *J Energy Storage* 2024;86:111197.
- [57] Giordano G, Klass V, Behm M, Lindbergh G, Sjöberg J. Model-based lithium-ion battery resistance estimation from electric vehicle operating data. *IEEE Trans Veh Technol* 2018;67(5):3720–8.
- [58] Zhang Y, Wilk T, Bergström J, Pecht M, Zou C. A machine learning-based framework for online prediction of battery ageing trajectory and lifetime using histogram data. *J Power Sources* 2022;526:231110.
- [59] She C, Wang Z, Sun F, Liu P, Zhang L. Battery aging assessment for real-world electric buses based on incremental capacity analysis and radial basis function neural network. *IEEE Trans Ind Inf* 2020;16(5):3345–54.
- [60] Zhao X, Hu J, Hu G, Qiu H. A state of health estimation framework based on real-world electric vehicles operating data. *J Energy Storage* 2023;63:107031.
- [61] Song L, Zhang K, Liang T, Han X, Zhang Y. Intelligent state of health estimation for lithium-ion battery pack based on big data analysis. *J Energy Storage* 2020;32:101836.
- [62] Zhao J, Ling H, Liu J, Wang J, Burke AF, Lian Y. Machine learning for predicting battery capacity for electric vehicles. *ETransportation* 2023;15:100214.
- [63] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies (NAACL-HLT), vol. 1, 2019, p. 4171–86.
- [64] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16×16 words: Transformers for image recognition at scale. In: International conference on learning representations. ICLR, 2021.
- [65] Baevski A, Zhou Y, Mohamed A, Auli M. wav2vec 2.0: A framework for self-supervised learning of speech representations. In: Advances in neural information processing systems (neurIPS). 2020, p. 12449–60.
- [66] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. ICML, 2015, p. 448–56.
- [67] Ba JL, Kiros JR, Hinton GE. Layer normalization. 2016, arXiv preprint arXiv: 1607.06450.
- [68] Loshchilov I, Hutter F. Decoupled weight decay regularization. 2017, arXiv preprint arXiv:1711.05101.
- [69] Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I, et al. Language models are unsupervised multitask learners. 2019, preprint.
- [70] Liu K, Peng Q, Che Y, Zheng Y, Li K, Teodorescu R, et al. Transfer learning for battery smarter state estimation and ageing prognostics: Recent progress, challenges, and prospects. *Adv Appl Energy* 2023;9:100117.
- [71] Ansel J, Yang E, He H, Gimelshain N, Jain A, Voznesensky M, et al. Pytorch 2: Faster machine learning through dynamic python bytecode transformation and graph compilation. In: International conference on architectural support for programming languages and operating systems. ASPLOS, 2024, p. 929–47.
- [72] Zhang H, Gui X, Zheng S, Lu Z, Li Y, Bian J. BatteryML: An open-source platform for machine learning on battery degradation. In: International conference on learning representations. ICLR, 2024.
- [73] Li L, Jamieson K, DeSalvo G, Rostamizadeh A, Talwalkar A. Hyperband: A novel bandit-based approach to hyperparameter optimization. *J Mach Learn Res* 2018;18(185):1–52.