



State of Health Estimation of Electric Vehicle Batteries Using Transformer-Based Neural Network

Yixin Zhao

Environmental Engineering Sciences,
University of Florida,
Gainesville, FL 32611
e-mail: yixin.zhao@ufl.edu

Sara Behdad¹

Associate Professor
Environmental Engineering Sciences,
University of Florida,
Gainesville, FL 32611
e-mail: sarabehdad@ufl.edu

Electric vehicles (EVs) are considered an environmentally friendly option compared to conventional vehicles. As the most critical module in EVs, batteries are complex electrochemical components with nonlinear behavior. On-board battery system performance is also affected by complicated operating environments. Real-time EV battery in-service status prediction is tricky but vital to enable fault diagnosis and prevent dangerous occurrences. Data-driven models with advantages in time-series analysis can be used to capture the degradation pattern from data about certain performance indicators and predict the battery states. The transformer model can capture long-range dependencies efficiently using a multi-head attention block mechanism. This paper presents the implementation of a standard transformer and an encoder-only transformer neural network to predict EV battery state of health (SOH). Based on the analysis of the lithium-ion battery from the NASA Prognostics Center of Excellence website's publicly accessible dataset, 28 features related to the charge and discharge measurement data are extracted. The features are screened using Pearson correlation coefficients. The results show that the filtered features can improve the model's accuracy and computational efficiency. The proposed standard transformer shows good performance in the SOH prediction. [DOI: 10.1115/1.4065762]

Keywords: electric vehicle batteries, state of health estimation, transformer networks, alternative energy sources, energy storage systems

1 Introduction

Due to the ongoing use of fossil fuels and reliance on internal combustion engine cars for more than a century, the transportation industry is one of the main contributors to global greenhouse gas (GHG) emissions [1]. Electric vehicles (EVs) have been identified as a promising solution to mitigate the GHG effect and the ever-growing energy demand [2]. Lithium-ion batteries are the most technologically advanced energy storage systems for EVs due to their high energy and power densities, strong environmental adaptability, and low self-discharge rate [3,4].

To make EVs more competitive with gasoline-powered vehicles, extremely fast charging is a key indicator that batteries need to achieve [5,6]. High rates of charging and discharging, combined with the wide range of operating temperatures to which EVs might be exposed (-20 to 70 °C), can accelerate battery degradation during cycling [7]. Battery degradation has been attributed to multiple mechanisms. The dominating factors are the growth of the solid electrolyte interphase layer, irreversible deposition of lithium metal on the anode, and loss of active material from the

cathode [8]. The batteries age over time leading to a reduction in their performance and safety [9]. The reliable operation of EV batteries requires real-time monitoring of their in-use states, such as state of health (SOH) [10,11]. However, battery in-use characteristics cannot be measured directly, but they need to be inferred by building degradation models [12–14]. A solid battery management system relies on a reliable battery degradation model for status monitoring and health state assessment to guarantee safe and high-performance operation.

There are two broad approaches to developing a battery degradation model for predicting and diagnosing: the physics-based approach and the data-driven approach. The physics-based model, also known as the electrochemical model, is a set of coupled partial differential equations that represent the microscopic chemical reactions occurring within the cell [15]. The pseudo-two-dimensional (P2D) model, for example, is one of the most widely used physics-based models [16]. This sophisticated model offers a thorough examination of the thermal energy balance, mass balances, charge balances, and kinetics of electrochemical reactions within the cell [17]. However, the physics-based model typically requires solving a set of tightly coupled differential equations [18]. The process is too complex and slow to be used for real-time management of in-use batteries.

The data-driven approaches estimate the effect of degradation on the battery's operational data. These approaches do not necessarily need to consider the underlying degradation behavior of the battery

¹Corresponding author.

Contributed by the Advanced Energy Systems Division of ASME for publication in the JOURNAL OF ENERGY RESOURCES TECHNOLOGY. Manuscript received April 12, 2024; final manuscript received June 13, 2024; published online July 18, 2024. Assoc. Editor: Tatiana Morozuk.

as well as its physical and chemical properties. However, the aging of EV batteries is nonlinear due to multiple factors such as charge–discharge rates, temperature variations, and usage patterns. Models that can handle high-dimensional data with complex dependencies are required to capture these patterns through measurable extrinsic signals. The machine-learning-based approaches can be trained on large datasets quickly, which makes them suitable for modeling complex systems like EV batteries. In addition, machine-learning models can be adapted to various battery chemistries and configurations [19,20]. In particular, deep neural networks can achieve higher prediction accuracy by training multiple layers of neural networks to match more closely to complex nonlinear battery systems [12,21–24].

Despite the accuracy, the deployment of deep learning models on-site still has significant challenges. These models typically require substantial computational resources and are challenging to be applied in resource-constrained environments (e.g., battery management systems). Many studies have proven that recurrent neural network (RNN) based neural networks, including gate recurrent unit (GRU) and long short-term memory (LSTM), are acceptable solutions for battery modeling because they can use internal states (memory) to represent battery aging information [25–27]. The LSTM network, for example, solves RNN's gradient disappearance and gradient explosion problems and has produced a relatively good performance for battery SOH prediction [28]. However, using LSTM to process sequential input in a recurrent manner is computationally intensive, which requires storing and updating memory cells at each time-step. Transformer is a newer type of neural network architecture that uses self-attention mechanisms to process input sequences without recurrence [29]. Transformers can process input sequences in parallel, which is much faster than the sequential processing used in LSTMs [30].

The main contribution of this study is to develop a highly efficient battery SOH prediction framework, thus making the model more applicable to real-time prediction in the field. To achieve this target, the transformer-based neural network that can process a sequence of data at once by using an attention mechanism is chosen as the prediction model. In addition to implementing a standard encoder–decoder transformer, the performance of the encoder-only transformer with a simpler architecture is investigated and compared. Moreover, the raw measurement data from the battery charge and discharge are extracted and screened for features before being input into the model. The computational efficiency of the model is thus further improved.

The rest of the structure of this paper can be summarized as follows. Section 2 is the related works on deep learning and feature extraction for battery state prediction. Section 3 describes the proposed SOH prediction framework, including the rules for feature extraction and selection, as well as the architecture of the transformer-based prediction models. Section 4 gives details of the experimental implementation and the SOH prediction results of the models. Finally, Sec. 5 concludes the work in this study and the perspectives for future research.

2 Related Work

2.1 Deep Learning for the State of Health Estimation. SOH and remaining useful life (RUL) are state indicators related to the aging behavior of the battery. To achieve a more comprehensive literature review, prediction models relating to both SOH and RUL are discussed in this section. Numerous studies applied deep learning-based models for battery degradation estimation owing to their advantages in modeling complex nonlinear problems. Among them, RNNs that can utilize sequential information are well suited for battery state prediction problems that require processing time-series data. As a variant of RNN, the LSTM network was designed to solve the gradient vanishing problem, which is one of the widely used models in the field of battery state prediction [31,32]. Kaur et al. [33] compared the performance of feed-forward neural networks (FNN), convolutional neural networks (CNN), and

LSTM for battery capacity estimation. Their results show that LSTM, which can recursively process time-series information, had the best accuracy. However, it requires greater computational cost than FNN and CNN. As a variant, the bidirectional long short-term memory (Bi-LSTM) integrates two LSTMs with positive and negative time series, allowing the model to detect information that the one-way network may overlook [34]. Sun et al. implemented the Bi-LSTM for SOH prediction, and the model can explore Li-ion batteries' degradation behavior from two sequence directions [26]. The results demonstrate that Bi-LSTM can achieve more accurate SOH estimation than a single LSTM. Meanwhile, the bidirectional behavior makes the Bi-LSTM model significantly slower to compute than LSTM [34]. For RUL and SOH prediction, Qu et al. combined the LSTM network with particle swarm optimization and an attention mechanism [35]. The paper illustrated that the attention mechanism assigns weights to each feature according to its impact on the output, which leads to higher accuracy of the model. The usefulness of utilizing attention mechanisms to advance the time-series prediction model has also been demonstrated in other areas. For example, Li et al. [36] applied a self-attention-based CNN and LSTM combined method for photovoltaic power prediction. Their model has improved adaptability to input features by re-assigning attention weights throughout the training phase, which reduces the prediction error.

The other deep learning model that constructs key modules with attention mechanisms is the transformer [29]. There is no recurrence or convolutional mechanism in the transformer. It can process a sequence of data all at once, using an attention mechanism that allows it to access any part of the historical data without being limited by distance [37]. In addition, it enables faster training than LSTM since the majority of operations can be computed in parallel [38]. Due to its ability to handle sequence data, the transformer has the potential to be employed in battery degradation modeling. Few studies apply transformers to battery status estimation. Chen et al. combined a denoising auto-encoder with the transformer encoder to complete the RUL estimation [30]. Their model achieved approximate or even better precision than a single LSTM and required significantly less training time than an LSTM. Mo et al. developed a transformer encoder-based neural network enhanced by a gated convolutional unit, which achieved relatively good performance and was computationally efficient [37]. These two studies prove the benefit of transformers in reducing the computational burden of battery models. However, both of them use encoder-only transformer-based structures, and the effect of the decoder on standard transformers for battery state estimation was not discussed. Recent advancements have further showcased the accuracy of transformers in this domain. Gu et al. proposed the hybrid CNN-transformer model to achieve an accurate estimation of the SOH [39]. The LSTM, single transformer, and CNN-LSTM are used as a comparative model. Their CNN transformer achieved high accuracy in different scenarios. Gomez et al. improved the temporal fusion transformer based on the Bi-LSTM encoder–decoder layer and performed high prediction accuracy [40]. Based on these findings, exploring different transformer architectures and incorporating feature engineering helps understand transformers' computational efficiency and prediction accuracy.

2.2 Battery Feature Extraction. An additional way to increase the computational efficiency and accuracy of the model is to extract, select, and optimize features from the raw data. The features from battery charge and discharge data can be divided into two categories: measured features and calculated features [41,42]. The measured features are extracted from the current, voltage, and temperature signals during the battery cycle that are available from the battery management system [43]. For example, Guo et al. extracted 14 features from the charging process, which are related to capacity, charge time, temperature, and current/voltage drop [44]. Their results show that after selection using gray relational analysis and dimensionality reduction by principal component analysis (PCA), the remaining features can well reflect

the internal aging process of the battery. Beganovic and Söffker [45] analyzed acoustic emission (AE) measurements from batteries and obtained AE energy as a feature that characterizes the degradation.

The calculated features are designed to mine more sensitive information from the measurement data. For this process, the signal that can be directly provided by the battery measurer is first transformed. The features extracted from the transformed signal profile are defined as calculated features [42]. For example, Li et al. applied the incremental capacity analysis technique to process battery voltage data [46]. This procedure could convert a flat voltage profile into an incremental capacity profile with a series of peaks and valleys. As the battery ages, the shape, amplitude, and position of the IC peaks change, which can be considered features reflecting degradation.

However, the computed feature extraction imposes more computational burden on the model than measured features. The main objective of this study is to develop a more efficient battery SOH prediction system. Therefore, the measured features from the previous studies are referred to as being used for feature extraction [41,43,44,46].

3 Methodology

3.1 Framework of the State of Health Estimation. A crucial factor to characterize the performance the battery is capable of in its present state is the SOH. It could also be a sign of battery deterioration. Many factors, including capacity and battery resistance, can be used to calculate the SOH. The SOH is determined in this study as the ratio of the nominal capacity to the releasable capacity, and it is displayed as

$$\text{SOH}(\%) = \frac{Q_{\max}}{Q_{\text{norm}}} \times 100\% \quad (1)$$

where $Q_{\max}$ is the maximal available capacity at the current cycle, and Q_{norm} is the nominal capacity given by the manufacturer. Since the nominal capacity of the battery is constant, the calculation of SOH is directly related to the available capacity of the battery.

Figure 1 depicts the flowchart for the SOH estimation. For feature extraction and optimization, the framework uses raw testing and monitoring data from Li-ion batteries. First, the initial extracted features are determined by investigating the relationship between the trends of the charge–discharge profiles and the degradation of the battery system. After extracting the features from the measured parameters, a correlation analysis is performed to evaluate the features' potential to capture battery degradation. Based on the selection of valuable features, a standard transformer neural network and an encoder-only transformer are employed to estimate the SOH.

3.2 Data Description. The data are obtained from the NASA lithium-ion battery aging dataset [47]. This public dataset tested commercially available lithium-ion rechargeable batteries with the model number 18650 at experiment control conditions. This study selected data from Battery 0005 for the experiment. The charging and discharging of Battery 0005 were repeated at room temperature. The charging process was carried out first at a constant current (CC) of 1.5 A. After the battery voltage reaches 4.2 V, it shifts to charging at a constant voltage (CV) mode until the current drops to 20 mA [48]. The discharging process was performed at a constant current of 2 A until the voltage dropped to 2.7 V. The signals of current (I), voltage (V), and temperature (T) during charging and discharging were recorded every 10 Hz. The battery aged significantly due to repeated charge and discharge cycles. The test came to an end when the battery's rated capacity had faded by 30%, signaling the end of its useful life. The variation in the capacity of the target battery is shown in Fig. 2.

3.3 Feature Extraction and Selection

3.3.1 Feature Extraction. The raw data consist of a continuous battery charging and discharging process. When reading the data, one full charging process with one full discharging process is considered a cycle. The available capacity recorded from the discharge process is used to calculate the SOH for the current cycle by Eq. (1). The data for feature extraction are selected as the full range of V , I , and T signals measured in the state of charging and discharging [46]. The changes in the charging and discharging profiles as the battery ages are illustrated in Fig. 3.

From Eq. (1), it is known that the extrapolation of the available battery capacity is the key to estimating the SOH. The battery capacity is the amount of electric charge that can be accumulated during charging and released reversibly during discharging [49]. The battery charging process is divided into two steps: CC mode and CV mode. According to the definition of capacity, the time required for the CC charging process is directly proportional to the charging capacity in the CC mode. From Figs. 3(a) and 3(b), it can be found that as the number of cycles increases, the time required to complete CC charging of the battery decreases. It thus implies that the time spent charging at a constant current can indicate the degree of battery aging.

The CV charging step is to keep the battery at a maximum specified potential while allowing the current to decrease through a current taper [50]. The more time spent in CV mode, the more difficult it proves to be for the lithium ions to migrate inside the battery [46]. Therefore, it shows the aggravation of the aging mechanism, such as elevated impedance and the formation of solid electrolyte interphase. In the charging temperature profile, the time point when the battery reaches its maximum temperature and the time point when the CC mode ends are largely coincident (Fig. 3(c)).

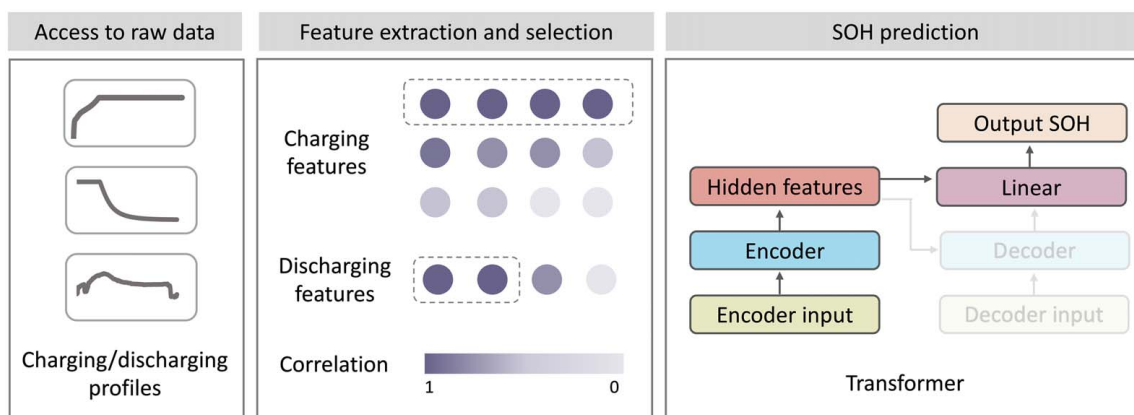


Fig. 1 Framework of the SOH estimation

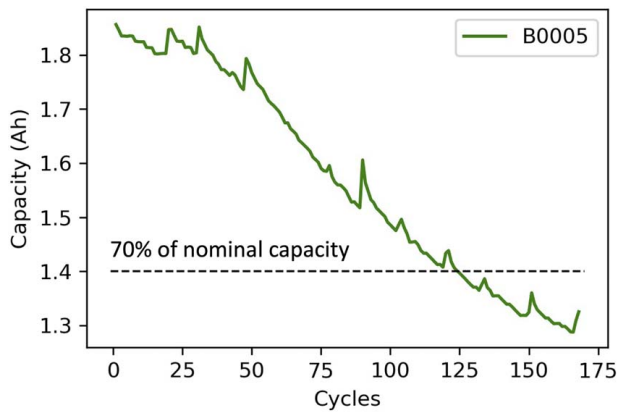


Fig. 2 Capacity degradation of Battery 0005

Based on the considerations above, it is worthwhile to extract and analyze the $I/V/T$ features associated with the beginning as well as the end of the CC and CV modes during the charging phase. As the current for CC charging, cutoff current, and the maximum voltage are fixed, these three values are not considered in the extraction of features. However, temperature is considered valuable for extraction and evaluation since temperature is one of the significant factors affecting the aging rate of the battery.

Figures 3(d)–3(f) present the profiles of battery discharge at different degradation levels. It can be seen that as the number of battery cycles increases, the time for the battery to complete constant current discharge, reach the cutoff voltage, and the time to reach the maximum temperature all become shorter and shorter. The available battery capacity is highly sensitive to the discharge process since it is gained by integrating the current curve over a full discharge process [49]. Therefore, several of the discharge features are also extracted pending subsequent evaluation. Based on the above analysis and on references from other similar battery feature extraction methods [41,43,46], a total of 28 charging and discharging features are initially extracted and summarized in Table 1. Examples of the profiles of the extracted features are displayed in Fig. 4.

3.3.2 Feature Selection. Pearson correlation coefficient (PCC) is employed to identify the degree of correlation between each feature and the battery's health status [51]. The most relevant features for prediction and modeling can be identified by calculating the correlation coefficient. A Pearson correlation coefficient of 1 or -1 implies a perfect positive or negative linear relationship. A coefficient of 0 indicates no linear relationship.

The purpose of selecting the extracted features is to remove noise and redundant features in order to reduce computational cost while ensuring model performance. Figure 5 illustrates the feature selection using the Pearson correlation coefficient in this study, which consists of the following two main steps:

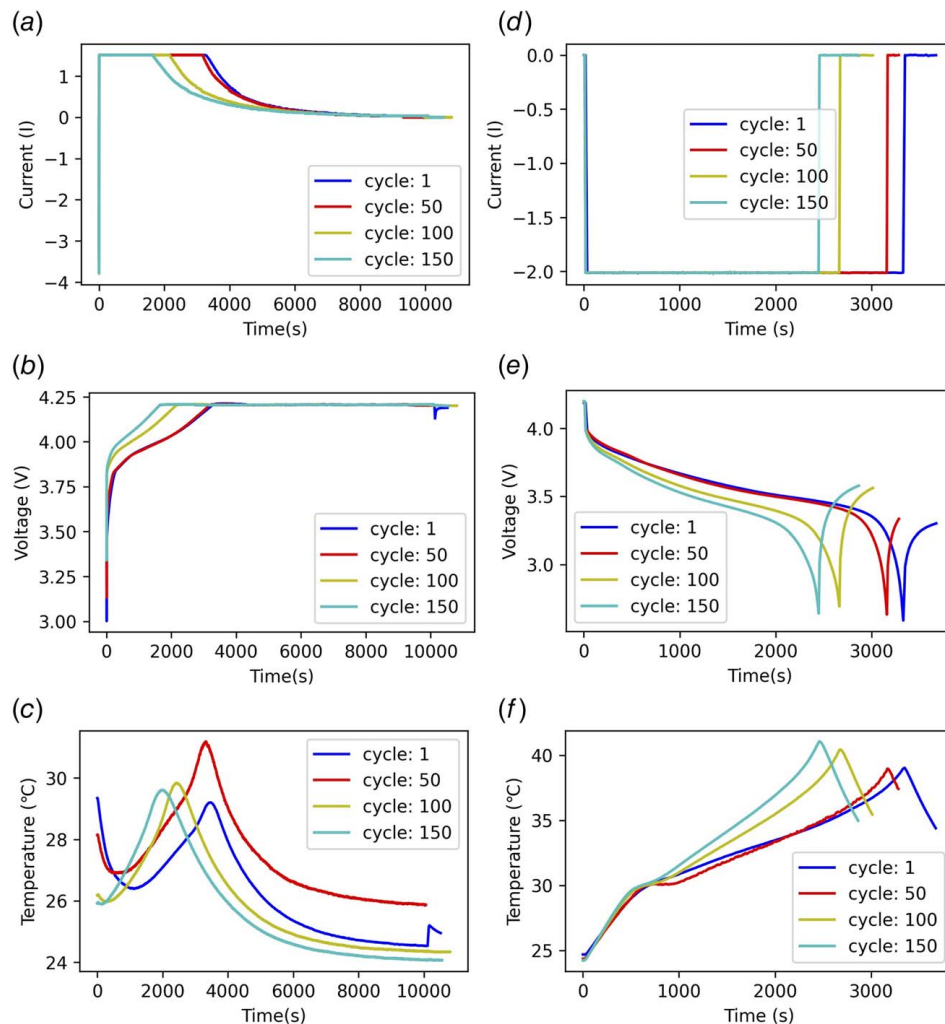


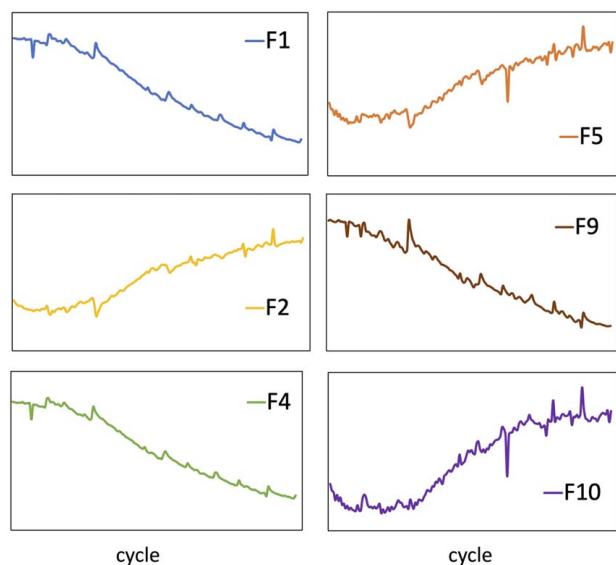
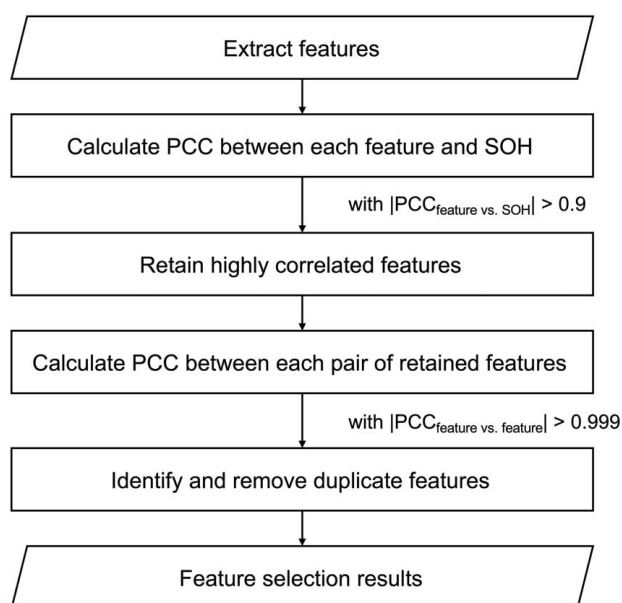
Fig. 3 Current, voltage, and temperature curves during charging and discharging of Battery 0005 after different number of cycles

Table 1 Feature extraction [41,43]

Group	Feature	Description	Type
Charging features	F1	Area covered by current curves of the CC charging	Current-related
	F2	Area covered by current curves of the CV charging	
	F3	Minimum slope of current curves in the CV charging	
	F4	Area covered by voltage curves of the CC charging	Voltage-related
	F5	Area covered by voltage curves of the CV charging	
	F6	Maximum slope of voltage curves in the CC charging	
	F7	Maximum temperature of charging	Temperature-related
	F8	Minimum temperature of charging	
	F9	Area covered by temperature curves of the CC charging	
	F10	Area covered by temperature curves of the CV charging	Time-related
	F11	Maximum temperature minus minimum temperature of charging	
	F12	The ratio of the CC charging area under the temperature curve to the corresponding area under the current curve	
	F13	The ratio of the CV charging area under the temperature curve to the corresponding area under the current curve	
	F14	CC charging time	
	F15	CV charging time	
	F16	CC charging time/(CC + CV charging time)	
	F17	Time to the minimum current in charging	
	F18	Time to reach the maximum voltage in charging	
	F19	Time to the maximum temperature in charging	
Discharging features	F20	Area covered by current curves of discharging	Current-related
	F21	Area covered by voltage curves of discharging	Voltage-related
	F22	Area covered by temperature curves of discharging	Temperature-related
	F23	Maximum temperature of discharging	
	F24	Minimum temperature of discharging	
	F25	Maximum temperature minus minimum temperature of discharging	Time-related
	F26	Time discharged under a constant current	
	F27	Time to the minimum voltage in discharging	
	F28	Time to the max temperature in discharging	

- Calculate the PCC between each feature and battery SOH. The features with a correlation coefficient absolute value greater than 0.9 are chosen.
- Calculate the PCC between each pair of features retained during step 1. A pair of features with an absolute PCC value greater than 0.999 is considered a “duplicate” feature. Only the one that is more correlated with SOH should be kept as model input. The remaining features that do not have “duplicates” are also kept.

The first step is retaining only closely related input information relevant to the prediction target. Specific screening criteria need to be determined for different datasets. In this paper, the threshold of 0.9 is chosen. After identifying strongly correlated features, the next step is to check for multicollinearity between these features. Multicollinearity occurs when two or more features are highly correlated with each other, leading to redundancy. This step confirms that each feature selected provides unique information about the health of the battery.

**Fig. 4 Profiles of extracted features F1, F2, F4, F5, F9, and F10****Fig. 5 Process of selecting features**

3.4 Model Structure. An encoder and a decoder compensate for the sequence-to-sequence architecture of the standard transformer. The encoder maps the input sequence into a high-dimensional vector, which is then provided to the decoder to generate the output sequence [30].

An input layer, a layer for positional encoding, and stack-identical encoder layers make up the encoder module [52]. The input layer converts the raw input time-series data to a high-dimensional vector space (d_{model}), through a fully connected network. This procedure standardizes the dimensionality for all the subsequent computations and helps the model capture complex patterns and relationships in the data. The following positional encoding layer is used to address the lack of knowledge of the order of data points in a sequence. The positions of the data points in the sequence are characterized by performing sine and cosine functions and giving each position a unique representation between 0 and 1. This helps the model understand the time-series order and handle temporal dependencies in the data. The encoder layer generates encodings that contain information about which parts of the input are relevant to each other [39]. Two sub-layers comprise each encoder layer: multi-head self-attention and fully connected feed-forward. Each sub-layer is followed by a normalization layer. The encoder generates a vector of dimension d_{model} , which can then be used by the decoder in the standard transformer or directly for prediction in the encoder-only transformer.

The decoder is also composed of the input layer, decoder layers, and an output layer. The decoder in the transformer model takes two main inputs: the encoded source sequence and the previous decoder output. The encoded source sequence is generated by the encoder and contains representations of the input sequence. The previous decoder output, in the general transformer, consists of previously generated tokens from the decoder. It serves as the reference for the model to learn from during training and to generate predictions during inference. However, the transformer presented in this paper is simplified in this step due to the specific nature of this prediction task. The objective is to predict a single value (SOH at time t); therefore there are no previously generated tokens to provide context. The previous decoder output is replaced with a sequence consisting of the last data point of the encoder input. In other words, the input features at the last battery cycle are used as the second input to the decoder in our case. This modification uses the most recent information available, which is the feature of the system at the latest time-step. The decoder layer is similar to the encoder layer in that it includes the self-attention mechanism and feed-forward neural network. However, the self-attention mechanism in the decoder also attends to the output sequence of the

Encoder. The decoder can be used to apply self-attention to both the latest battery features and the encoder's outputs. This dual attention mechanism enhances the model's ability to capture complex dependencies within the battery data. Finally, a fully connected layer is used to map the representations learned by the last transformer unit, producing the SOH estimation.

In this study, the standard transformer and the encoder of the transformer are applied to learn the long-term dependencies of the SOH degradation from charging and discharging features (Fig. 6). With an encoder-only transformer, the hidden features output from the encoder enter directly into the final fully connected layer for prediction. This simplifies the pathway from input to prediction and improves the speed of computation.

4 Results and Discussion

4.1 Feature Selection Results. According to the features described in Table 1, 28 features are extracted from Battery 0005 in the NASA dataset. The correlation coefficient between features and battery SOH is summarized in Table 2 in descending order [41,44,46]. The PCC helps identify the most relevant features by measuring the linear relationship between each feature and the SOH. A higher absolute value of the PCC indicates a stronger relationship with the SOH and suggests that the feature is more informative for predicting battery health. Linear regression (LR) is also performed for each feature, and the resulting R^2 values are listed in Table 2. The R^2 value ranges from 0 to 1, with higher values indicating a stronger predictive relationship. The analysis shows that features with high PCC values generally had high R^2 values, which further indicates that these features can explain a significant portion of the variance in the SOH.

Selecting features with a high PCC directs the model to focus on the most relevant data. The threshold value is 0.9 to select the valuable features. Of the 28 features, 18 showed a PCC value higher than 0.9, which indicates that this threshold is useful in capturing the strongest signals while retaining sufficiency features. This helps the model gather information from different battery dimensions and reduces the risk of overfitting.

The results show that the features with a PCC greater than 0.9 with the SOH include: F1, F2, F4, F5, F9, F10, F14, F15, F16, and F18 from the charging group, and F20, F21, F22, F23, F25, F26, F27, and F28 from discharging group. Features such as the area covered by current and voltage curves during CC and CV charging (F1, F2, F4, F5) measure the battery's ability to accept and store charge. As the battery ages, a decrease in these areas

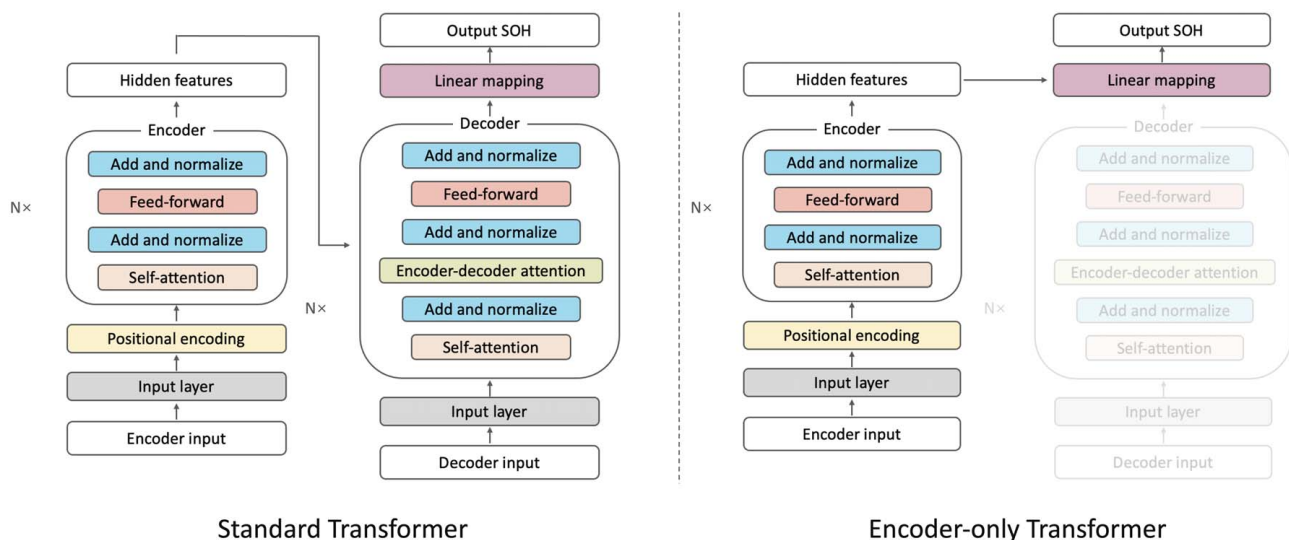


Fig. 6 Structure of standard transformer and encoder-only transformer

Table 2 Pearson correlation coefficient (absolute value) and R^2 between features and SOH in descending order

Group	Feature	PCC with the SOH (absolute value)	R^2 with SOH
Charging features	F4	0.996	0.992
	F14	0.996	0.992
	F1	0.996	0.992
	F16	0.994	0.987
	F9	0.990	0.980
	F2	0.988	0.976
	F15	0.976	0.952
	F5	0.976	0.952
	F10	0.969	0.939
	F18	0.958	0.917
	F17	0.774	0.560
	F6	0.637	0.405
	F11	0.634	0.402
	F3	0.610	0.372
	F13	0.600	0.360
	F12	0.401	0.161
	F7	0.383	0.147
	F8	0.016	0.000
	F19	0.006	0.000
Discharging features	F26	0.999	0.999
	F20	0.999	0.999
	F27	0.999	0.999
	F28	0.999	0.999
	F21	0.999	0.999
	F22	0.993	0.986
	F25	0.979	0.958
	F23	0.937	0.878
	F24	0.052	0.003

Note: The bold are the features whose PCC is higher than 0.9.

indicates reduced capacity and increased internal resistance. Temperature-related features (F9, F10) capture the thermal profile during charging. The higher temperatures represent the heat generation caused by increased internal resistance and potential degradation processes such as solid electrolyte interphase layer growth. The duration of the CC and CV phases (F14, F15) and their ratio (F16) are key indicators of charging efficiency, where longer CV times and shorter CC times show aging. In addition, the time to reach maximum voltage (F18) shows the charging speed of the battery can be charged, where longer times suggest higher internal resistance.

On the other hand, the features from the discharge data, except for the minimum temperature, are all highly correlated with SOH. Similarly, with the discharge area (F20, F21), time-related (F26, F27, F28), as well as the heat generation (F22, F23, F25) features are of high importance. In particular, F20, F21, F26, F27, and F28 have a high correlation of 0.9999 with the predicted target, which can be almost linear. This is because the experimental condition of Battery 0005 is a continuous release at a constant current during the discharge process until the cutoff voltage is reached. According to the calculation method of battery capacity, when the battery is discharged at a constant current, its capacity is given by the discharge current multiplied by the discharge duration [49]. Therefore, for the dataset adopted in this study, the features related to the discharge duration would be nearly linearly correlated with the battery SOH. However, in practical applications, the discharge pattern of EV batteries is completely dependent on driving behavior and is a random discharge behavior. The discharge features are still essential for predicting randomly discharged battery states, except that they will not have this strong linear relationship. In addition, it is not common in practice to allow an EV battery to complete a full discharge process. The operating conditions of EV discharging are not as stable as parking and charging at a charging station. Therefore, after feature selection, the model's performance

with only charging features is also tested and compared to the model with full feature input.

While slope-related features (F3, F6) can indicate charge efficiency, they capture only a momentary aspect of the dynamic process, and provide less information than area features. The slopes are sensitive to noise and fluctuations, which make them less reliable for long-term SOH prediction. Then, the feature least associated with the SOH is the minimum temperature during charging and discharging. This is because the lowest temperature during battery operation usually occurs at the beginning or end of the charging and discharging phases when the battery is relatively inactive. This value is highly correlated with the ambient temperature and, therefore has little correlation with the status of the battery.

To eliminate redundant features and avoid overfitting, the similarity between each pair of high-correlation features is also calculated and presented in Fig. 7. According to the feature selection rules defined in Sec. 3.3.2, a pair of features with an absolute correlation value greater than 0.999 is considered a "duplicate." The one with a lower correlation with SOH should be removed. Finally, the valuable features that are selected include: F1, F2, F4, F5, F9, F10, F16, and F18 from charging group, and F20, F21, F22, F23, and F25 from discharging group.

In this particular dataset, multiple battery features from charging and discharging are present with high-correlation coefficients to the prediction target. The reason is that the batteries in the dataset are charged and discharged with a regular and stable behavior. It is possible that using the LR to fit the SOH curve would also achieve good performance. However, LR models are difficult to adapt if there is randomness in the battery data. Due to the consideration of the stochasticity of the battery usage in real situations, the feasibility of the transformer is still chosen to be discussed in this paper.

4.2 State of Health Estimation Results

4.2.1 Experiment Implementation Details. The two steps of feature selection described in Sec. 3.3.2 include selecting features with a higher correlation with the predicted target and deleting duplicate features. According to the results, a total of 13 valuable features from the raw charging and discharging data are obtained. However, due to the specificity of the dataset used in this study, a portion of the features from the discharge process have a very high linear correlation with SOH. Using these discharge features to examine the effectiveness of the developed time-series prediction model is not reasonable. Therefore, during the evaluation of the model, the model inputs are divided into four cases: all 28 available features, 19 available features from the charging data only, 13 selected features from the charging and discharging process, and 8 selected features from the charging data only (Table 3). In addition, stable charging data are more accessible than stable discharge data in real-world applications. The experimental discharge data used in this paper are different from practice in which the discharge pattern of EV batteries varies significantly with driving conditions, which makes it difficult to obtain consistent and reliable signals. On the other hand, charging typically occurs in more controlled environments, such as at charging stations, where conditions are stable and predictable. It would be helpful for practical applications to demonstrate the ability of the model to rely only on charging data.

In addition, groups 1 and 2 provide a more complete dataset but potentially increase computational cost and risk of overfitting. Groups 3 and 4 include the most predictive charging features, simplify the model, and reduce computational load. The effect of feature selection and discharge features are discussed in the following sections by comparing models' performance with different input groups.

The two transformer architectures employed for SOH estimation are implemented in PYTHON utilizing the PyTorch framework. The models are run on a single CPU and utilize the mean squared error loss function during the training process. The dataset is split into 70% training and 30% test sets. The developed transformer-based models have seven key parameters: the number of time-steps

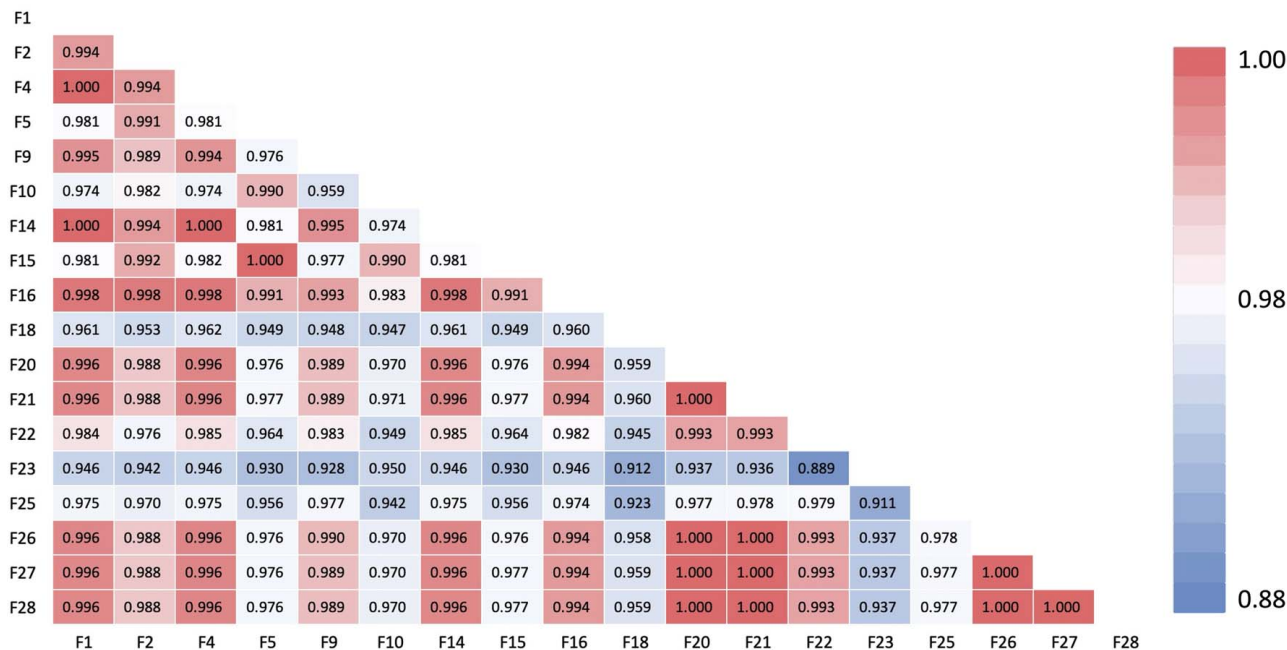


Fig. 7 Similarity between features with correlation coefficients greater than 0.9 with the SOH

Table 3 Four input groups

Input group	Description	Feature
1	All 28 available features	F1–F28
2	19 available features from the charging data only	F1–F19
3	13 selected features from the charging and discharging data	F1, F2, F4, F5, F9, F10, F16, F18, F20, F21, F22, F23, F25
4	8 selected features from the charging data only	F1, F2, F4, F5, F9, F10, F16, F18

to be input (sampler size), the number of expected features in the transformer encoder–decoder inputs (d_{model}), the number of encoder layers, the number of decoder layers, the number of heads in the multi-head attention mechanism, the dimension of the feed-forward network, and the fraction of neurons affected by dropout. The parameters are determined by grid search on the validation error. Table 4 summarizes the range of grid search and optimized hyperparameter settings for the transformer-based models. The parameter optimization process is based on input group 1 on a standard encoder–decoder transformer. The encoder-only transformer used in this study adopts the same parameter settings in Table 4, except that the decoder module is masked.

Table 4 Grid search and optimized hyperparameters for transformer-based models

Hyperparameters	Grid search	Setting
Sampler size	5, 10, 15, 20	15
d_{model}	64, 128, 256	64
Number of encoder layers	1, 3, 5	1
Number of decoder layers	1, 3, 5	1
Attention head	4, 8	8
Dimension of the feed-forward network	256, 512	512
Dropout	0, 0.1	0

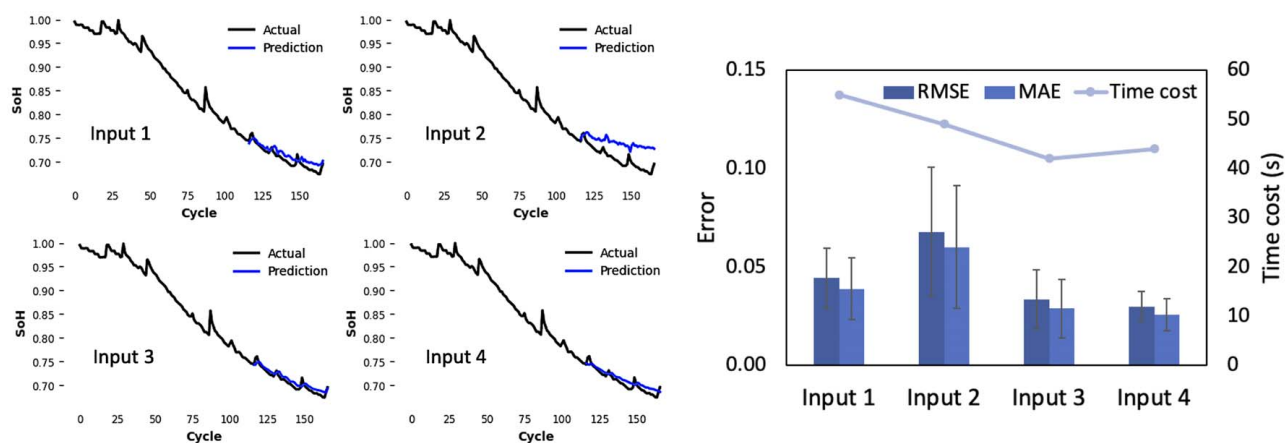
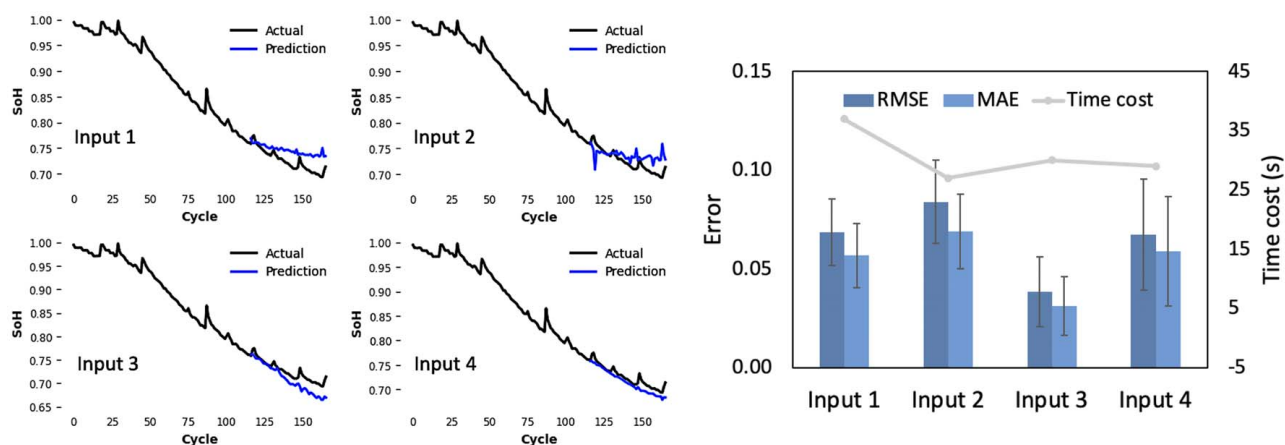
Root mean square error (RMSE) and mean absolute error (MAE) are utilized as evaluation criteria to assess the efficacy of the designed model for the SOH estimation. The model's stability is indicated by the RMSE, which measures the difference between the true and predicted values. MAE, a linear score of the prediction error, can be used to show how accurate a model is. In addition, to verify the results, the model performance metrics reported in this paper are based on the average and standard deviation of five separate runs.

4.2.2 Model Performance. The performance of the two transformer-based neural networks using four different sets of inputs is summarized in Table 5. Based on the comparison of RMSE and MAE, the standard transformers with input groups 3 and 4 perform best.

The performance of the standard transformers and encoder-only transformers for the SOH prediction task can be compared and discussed. The transformer encoder uses self-attention to capture the dependencies between the different time-steps in the sequence. The resulting hidden representations are then used as input to downstream prediction tasks. Therefore, the hidden representations obtained by the encoder module can be directly fed into a fully connected layer for the SOH prediction. The SOH prediction curves from both transformer-based models are shown in Figs. 8 and 9. The comparison shows that the encoder-only transformer does not perform as well as the standard transformer, especially when the number of features in the input is small. For example, the model with a decoder and without a decoder has the largest performance gap when the input is group 4. However, the advantage of encoder-only transformers is their computational speed. It runs over 30% faster than the standard transformer without a decoder module. In future studies, if the performance of the transformer encoder can be enhanced by further parameter optimization or feature enhancement, it will be a more suitable model for real applications. In real-time monitoring and decision-making, updating SOH predictions on time facilitates prompt management. Specifically, it can help to achieve timely dynamic optimization of the corresponding charging protocols to improve battery life and performance. In addition, the reduction in computational complexity helps integrate deep learning models into resource-constrained in-vehicle systems.

Table 5 Performance of two transformer-based models with different inputs

Input group	Standard transformer			Encoder-only transformer		
	RMSE	MAE	Time cost (s)	RMSE	MAE	Time cost (s)
1	0.0443 ± 0.0152	0.0389 ± 0.0158	55	0.0687 ± 0.0170	0.0569 ± 0.0162	37
2	0.0679 ± 0.0327	0.0600 ± 0.0311	49	0.0841 ± 0.0211	0.0692 ± 0.0188	27
3	0.0335 ± 0.0148	0.0287 ± 0.0151	42	0.0384 ± 0.0176	0.0312 ± 0.0150	30
4	0.0300 ± 0.0077	0.0258 ± 0.0081	44	0.0674 ± 0.0282	0.0592 ± 0.0277	29

**Fig. 8 Predicted results for standard transformer****Fig. 9 Predicted results for encoder-only transformer**

On the other hand, the effect of the features from the discharge data on the model is worth discussing. As explained in Sec. 4.1, since the experimental conditions of Battery 0005 are constant current continuous discharge, the duration of discharge and SOH show a nearly linear relationship. The impact of features with such a high correlation with the predicted target on model performance is apparent. As shown in Figs. 8 and 9, models that employed discharge features, with or without a decoder module, can predict more accurately than models that use only charging features.

However, constant current continuous discharge of batteries is not common in real-world applications. Although the extraction and selection of discharge features can substantially improve the model's performance in the current dataset, the usage of this type of feature in actual EV battery SOH prediction needs to be verified. Without the discharge feature, the standard transformer with the

selected charging feature (input group 4) can still achieve good model performance, with RMSE and MAE of 0.030 and 0.0258, respectively. Regarding the computational cost of feature extraction, it takes about 20 s to compute features for 168 charge-discharge data cycles.

The validity of feature selection can be verified on standard transformers which perform more stable. It is also necessary to exclude the interference of discharging features on the model performance due to the high correlation between discharge features and SOH. Feature selection is evaluated only by comparing the influence of all available and selected features from charging data on the model. The RMSE and MAE of standard transformers using input group 2 are 0.0679 and 0.0600, respectively. After feature selection, the RMSE and MAE of standard transformers with input group 4 are reduced to 0.0300 and 0.0258. This demonstrates the impact

Table 6 Performance of two transformer-based models with different types of features

Feature type	Standard transformer		Encoder-only transformer	
	RMSE	MAE	RMSE	MAE
Current-related	0.0727 ± 0.0405	0.0681 ± 0.0424	0.0509 ± 0.0415	0.0426 ± 0.0415
Voltage-related	0.0608 ± 0.0378	0.0562 ± 0.0396	0.0932 ± 0.0490	0.0813 ± 0.0461
Temperature-related	0.0493 ± 0.0187	0.0453 ± 0.0189	0.0344 ± 0.0131	0.0282 ± 0.0141
Time-related	0.0328 ± 0.0190	0.0276 ± 0.0198	0.0907 ± 0.0410	0.0784 ± 0.0379

Table 7 Comparison of model computational costs

Model	RMSE	MAE	Time cost (s)
Standard transformer	0.0300 ± 0.0077	0.0258 ± 0.0081	44
Encoder-only transformer	0.0674 ± 0.0282	0.0592 ± 0.0277	29
LSTM	0.0681 ± 0.0053	0.0636 ± 0.0049	60
GRU	0.0281 ± 0.0044	0.0181 ± 0.0055	54

Table 8 Comparison of model prediction results with other literature

Model	RMSE	MAE	Reference
Standard transformer	0.0300	0.0258	This study
LSTM	0.0387	0.0439	[39]
ARIMA	0.0375	—	[53]
LSTM with swarm optimization and attention mechanism	0.006	—	[35]
Bi-LSTM	0.005	—	[26]

of the feature selection rule proposed in the paper. In addition to improving the model's accuracy, the computation time of the model has been improved by around 10%. Speed and accuracy are the characteristics required for EV battery management systems. Our results demonstrate that feature extraction and selection, and, feature optimization, such as noise reduction and enhancement, are also interesting problems to explore in the future.

This study also analyzes the response of the model to different types of highly related features. The 13 screened features from input group 2 can be categorized into four types: current-related (F1, F2, F20), voltage-related (F4, F5, F21), temperature-related (F9, F10, F22, F23, F25), and time-related (F16, F18). The performance of the model respectively, using these four types of features as inputs is summarized in Table 6. Due to the reduced number of input features, the error and the deviation of predictions are increased. The five features correlated with temperature provide relatively accurate and consistent model performance. The other three types with small numbers are more at risk of overfitting the model. This demonstrates the importance of providing multidimensional information to improve the robustness of the model. However, due to the randomness of the deep learning model and the unbalanced set, this experiment does not provide a fair assessment of the model's sensitivity to different types of features.

The proposed model is also performed on other batteries from the NASA dataset. The results show that fine-tuning the hyperparameters is necessary when applying the model to other batteries since the charging and discharging conditions are not identical.

4.2.3 Comparison of Estimation Results and Computational Cost With Other Models. The computational cost of transformers is compared with LSTM and GRU in Table 7, which are widely used in the field of battery SOH estimation. All the models use group 4 features as input with a sliding window size of 15. The parameter settings for the two transformers are shown in Table 4 for optimum parameters. The LSTM and GRU are designed with two LSTM/GRU layers with 256 hidden units per layer.

As can be seen from the results, the standard transformer works with high accuracy while using the competitive time cost. The self-attention mechanism helps the transformer models give different weights to different components and focus on the information that is most relevant to predicting SOH. This capability is particularly helpful for battery data that may show complex degradation patterns. In addition, transformers can process all elements in a sequence at the same time. On the other hand, RNN variants such as LSTM and GRU process data sequentially in a recurrent manner.

That results in a proper computational speed for the transformers, especially the encoder-only one. However, the accuracy of the current encoder-only transformer should still be improved. In this case, the LSTM and GRU are only roughly tuned for hyperparameters, so these do not represent the best results that can be achieved by RNN models. Deepening the model architecture may improve the prediction results of LSTM or GRU, but it may also cause a further increase in computational cost.

The performance of models from standard transformers using the selected charging features is compared with models from other literature. The prediction results from the other models were also implemented on Battery 0005 from the NASA public dataset. Moreover, they have a similar training and test set split (70%: 30%) as the one in this study. Although the raw data are the same, the data pre-processing or feature extraction employed by different models is varied. Therefore, the performance comparison here is rough and not entirely fair. The results of the comparison are presented in Table 8. The results show that our standard transformers have advantages in terms of accuracy compared to single models, such as LSTM and autoregressive integrated moving average model (ARIMA). However, advanced variants of the simple model, such as LSTM with enhanced optimization mechanisms, self-attention mechanisms, and bidirectional LSTM, perform better than the transformers in this paper. The standard transformer has the advantage of simplicity and the ability to make fast predictions for the battery timing prediction problem, but there is still potential for accuracy improvement.

5 Conclusion

This paper presents an efficient framework for estimating battery SOH that comprises feature extraction/selection and a predictive model based on transformers. Based on battery current, voltage, and temperature data measured during battery charging and discharging, 28 features for battery SOH estimation are extracted. The correlation between features and SOH and the similarity between features are subsequently evaluated using the Pearson correlation coefficient. This way features with a high correlation with the prediction target are retained, and duplicate features are removed to avoid adding noise and computational burden to the model.

The performances of the standard transformer and the encoder-only transformer model are tested on four different input sets. The results show that the standard transformer performs better than the

encoder-only model. The standard transformer with the selected charging-only features can achieve good model performance, with RMSE and MAE of 0.030 and 0.0258, respectively. In addition, the discharge characteristics of the cells with constant current continuous discharge behavior improve the accuracy of the models. However, the constant and stable discharge behavior results from controlled experimental conditions. The benefit of using the discharge characteristics in predicting the state of batteries with random discharge behavior remains to be verified.

Real-world EV usage is characterized by stochasticity compared to stable data from the laboratory. Batteries in practical use are subjected to various usage patterns and environmental conditions, which can significantly impact the degradation trajectory. To adjust the model to these uncertainties, it is necessary to use a comprehensive dataset that incorporates more aging conditions, such as varying temperatures, humidity, non-constant charging and discharging parameters, and diverse driving cycles. In addition, techniques to advance the robustness of the model should be tested, such as transfer learning and advanced data augmentation. Incorporating additional sensor data also has the potential to improve the model's reliability.

Although transformers have been compared with LSTM and GRU models, future research could include a more comprehensive analysis to identify areas for improvement. For example, quantifying the uncertainty in predictions caused by the model's randomness and identifying the sources of error is needed. Future directions can also focus on feature extraction for partial charge-discharge profiles. Most current prediction models for SOH use the full range of charge and discharge data. However, EV batteries do not always complete a whole charging or discharging cycle in real-driving situations. Therefore, using partial battery data for state prediction would be more practical. Furthermore, adaptive feature extraction techniques that dynamically select the most relevant features based on current battery conditions could improve the model's capability.

Acknowledgment

This material is based upon work supported by the National Science Foundation–USA under Grants Nos. 2324950 and 2026 276. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

References

- Reinhardt, R., Christodoulou, I., Gassó-Domingo, S., and Amante García, B., 2019, "Towards Sustainable Business Models for Electric Vehicle Battery Second Use: A Critical Review," *J. Environ. Manag.*, **245**, pp. 432–446.
- Fang, W., Kwon, O. J., and Wang, C.-Y., 2010, "Electrochemical–Thermal Modeling of Automotive Li-Ion Batteries and Experimental Validation Using a Three-Electrode Cell," *Int. J. Energy Res.*, **34**(2), pp. 107–115.
- Capitaine, J.-A., and Wang, Q., 2018, "Design of a Test Platform for the Determination of Lithium-Ion Batteries State of Health," *ASME J. Mech. Des.*, **141**(2), p. 021702.
- Quijano-Ortiz, F., and Seepersad, C., 2022, "Design Recommendations for Reducing the Environmental Impact of Battery Packs," *Proceedings of the IDETC/CIE*, St. Louis, MO, Aug. 14–17, American Society of Mechanical Engineers Digital Collection, p. V005T05A010.
- Paulus, G. K., Hornstra, L. M., Alygizakis, N., Slobodnik, J., Thomaidis, N., and Medema, G., 2019, "The Impact of On-Site Hospital Wastewater Treatment on the Downstream Communal Wastewater System in Terms of Antibiotics and Antibiotic Resistance Genes," *Int. J. Hyg. Environ. Health*, **222**(4), pp. 635–644.
- Patel, P., and Nelson, G. J., 2020, "The Influence of Structure on the Electrochemical and Thermal Response of Li-Ion Battery Electrodes," *ASME J. Energy Resour. Technol.*, **142**(5), p. 050906.
- Waldmann, T., Wilka, M., Kasper, M., Fleischhammer, M., and Wohlfahrt-Mehrens, M., 2014, "Temperature Dependent Ageing Mechanisms in Lithium-Ion Batteries—A Post-Mortem Study," *J. Power Sources*, **262**, pp. 129–135.
- Tanim, T. R., Dufek, E. J., Evans, M., Dickerson, C., Jansen, A. N., Polzin, B. J., Dunlop, A. R., et al., 2019, "Extreme Fast Charge Challenges for Lithium-Ion Battery: Variability and Positive Electrode Issues," *J. Electrochem. Soc.*, **166**(10), p. A1926–A1938.
- Kim, T. J., Youn, B. D., and Kim, H. J., 2014, "Online-Applicable Temperature Prediction Model for EV Battery Pack Thermal Management," *Proceedings of the IDETC/CIE*, Portland, OR, Aug. 4–7, 2013, American Society of Mechanical Engineers Digital Collection, p. V008T12A008.
- Noura, N., Boulon, L., and Jemef, S., 2020, "A Review of Battery State of Health Estimation Methods: Hybrid Electric Vehicle Challenges," *World Electr. Veh. J.*, **11**(4), p. 66.
- Lui, Y. H., Li, M., Sadoughi, M., Hu, C., and Hu, S., 2018, "Physics-Based State of Health Estimation of Lithium-Ion Battery Using Sequential Experimental Design," *Proceedings of the IDETC/CIE*, Quebec City, Quebec, Canada, Aug. 26–29, American Society of Mechanical Engineers Digital Collection, p. V02BT03A061.
- Li, W., Sengupta, N., Dechent, P., Howey, D., Annaswamy, A., and Sauer, D. U., 2021, "One-Shot Battery Degradation Trajectory Prediction With Deep Learning," *J. Power Sources*, **506**, p. 230024.
- Hu, C., Hong, M., Li, Y., and Jeong, H.-L., 2016, "On-Board Analysis of Degradation Mechanisms of Lithium-Ion Battery Using Differential Voltage Analysis," *Proceedings of the IDETC/CIE*, Charlotte, NC, Aug. 21–24, American Society of Mechanical Engineers Digital Collection, p. V02AT03A044.
- Xi, Z., Jing, R., and Lee, C., 2016, "Diagnostics and Prognostics of Lithium-Ion Batteries," *Proceedings of the IDETC/CIE*, Boston, MA, Aug. 2–5, American Society of Mechanical Engineers Digital Collection, p. V02AT03A038.
- Fotouhi, A., Auger, D. J., Propp, K., Longo, S., and Wild, M., 2016, "A Review on Electric Vehicle Battery Modelling: From Lithium-Ion Toward Lithium–Sulphur," *Renewable Sustainable Energy Rev.*, **56**, pp. 1008–1021.
- Doyle, M., Fuller, T. F., and Newman, J., 1993, "Modeling of Galvanostatic Charge and Discharge of the Lithium/Polymer/Insertion Cell," *J. Electrochem. Soc.*, **140**(6), p. 1526.
- Han, S., Tang, Y., and Khaleghi Rahmian, S., 2021, "A Numerically Efficient Method of Solving the Full-Order Pseudo-2-Dimensional (P2D) Li-Ion Cell Model," *J. Power Sources*, **490**, p. 229571.
- Edge, J. S., O'Kane, S., Prosser, R., Kirkaldy, N. D., Patel, A. N., Hales, A., Ghosh A., et al., 2021, "Lithium Ion Battery Degradation: What You Need to Know," *Phys. Chem. Chem. Phys.*, **23**(14), pp. 8200–8221.
- Patil, M. A., Tagade, P., Hariharan, K. S., Kolake, S. M., Song, T., Yeo, T., and Doo, S., 2015, "A Novel Multistage Support Vector Machine Based Approach for Li Ion Battery Remaining Useful Life Estimation," *Appl. Energy*, **159**, pp. 285–297.
- Zhang, Y., Xiong, R., He, H., and Pecht, M., 2019, "Validation and Verification of a Hybrid Method for Remaining Useful Life Prediction of Lithium-Ion Batteries," *J. Clean. Prod.*, **212**, pp. 240–249.
- Lee, S., Cui, H., Rezvanizadeh, M., and Ni, J., 2013, "Battery Prognostics: SOC and SOH Prediction," *Proceedings of the IDETC/CIE*, Notre Dame, IN, June 4–8, American Society of Mechanical Engineers Digital Collection, pp. 689–695.
- Li, W., Jiao, Z., Du, L., Fan, W., and Zhu, Y., 2019, "An Indirect RUL Prognosis for Lithium-Ion Battery Under Vibration Stress Using Elman Neural Network," *Int. J. Hydrogen Energy*, **44**(23), pp. 12270–12276.
- Dai, H., Zhao, G., Lin, M., Wu, J., and Zheng, G., 2019, "A Novel Estimation Method for the State of Health of Lithium-Ion Battery Using Prior Knowledge-Based Neural Network and Markov Chain," *IEEE Trans. Ind. Electron.*, **66**(10), pp. 7706–7716.
- Bai, G., and Wang, P., 2015, "A Self-Cognizant Dynamic System Approach for Health Management: Lithium-Ion Battery Case Study," *Proceedings of the IDETC/CIE*, Buffalo, NY, Aug. 17–20, American Society of Mechanical Engineers Digital Collection, p. V02AT03A041.
- Yayan, U., Arslan, A. T., and Yucel, H., 2021, "A Novel Method for SoH Prediction of Batteries Based on Stacked LSTM With Quick Charge Data," *Appl. Artif. Intell.*, **35**(6), pp. 421–439.
- Sun, H., Sun, J., Zhao, K., Wang, L., and Wang, K., 2022, "Data-Driven ICA-Bi-LSTM-Combined Lithium Battery SOH Estimation," *Math. Probl. Eng.*, **2022**(1), p. e9645892.
- Fan, Y., Xiao, F., Li, C., Yang, G., and Tang, X., 2020, "A Novel Deep Learning Framework for State of Health Estimation of Lithium-Ion Battery," *J. Energy Storage*, **32**, p. 101741.
- Li, X., Zhang, L., Wang, Z., and Dong, P., 2019, "Remaining Useful Life Prediction for Lithium-Ion Batteries Based on a Hybrid Model Combining the Long Short-Term Memory and Elman Neural Networks," *J. Energy Storage*, **21**, pp. 510–518.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I., 2017, "Attention Is All You Need."
- Chen, D., Hong, W., and Zhou, X., 2022, "Transformer Network for Remaining Useful Life Prediction of Lithium-Ion Batteries," *IEEE Access*, **10**, pp. 19621–19628.

- [31] Zhou, Y., Huang, Y., Pang, J., and Wang, K., 2019, "Remaining Useful Life Prediction for Supercapacitor Based on Long Short-Term Memory Neural Network," *J. Power Sources*, **440**, p. 227149.
- [32] Zhang, Y., Xiong, R., He, H., and Pecht, M. G., 2018, "Long Short-Term Memory Recurrent Neural Network for Remaining Useful Life Prediction of Lithium-Ion Batteries," *IEEE Trans. Veh. Technol.*, **67**(7), pp. 5695–5705.
- [33] Kaur, K., Garg, A., Cui, X., Singh, S., and Panigrahi, B. K., 2021, "Deep Learning Networks for Capacity Estimation for Monitoring SOH of Li-Ion Batteries for Electric Vehicles," *Int. J. Energy Res.*, **45**(2), pp. 3113–3128.
- [34] Siami-Namini, S., Tavakoli, N., and Namin, A. S., 2019, "The Performance of LSTM and BiLSTM in Forecasting Time Series," *2019 IEEE International Conference on Big Data (Big Data)*, Los Angeles, CA, Dec. 9–12, pp. 3285–3292.
- [35] Qu, J., Liu, F., Ma, Y., and Fan, J., 2019, "A Neural-Network-Based Method for RUL Prediction and SOH Monitoring of Lithium-Ion Battery," *IEEE Access*, **7**, pp. 87178–87191.
- [36] Li, J., Niu, H., Meng, F., and Li, R., 2022, "Prediction of Short-Term Photovoltaic Power Via Self-Attention-Based Deep Learning Approach," *ASME J. Energy Resour. Technol.*, **144**(10), p. 101301.
- [37] Mo, Y., Wu, Q., Li, X., and Huang, B., 2021, "Remaining Useful Life Estimation Via Transformer Encoder Enhanced by a Gated Convolutional Unit," *J. Intell. Manuf.*, **32**(7), pp. 1997–2006.
- [38] Zeyer, A., Bahar, P., Irie, K., Schlüter, R., and Ney, H., 2019, "A Comparison of Transformer and LSTM Encoder Decoder Models for ASR," *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, Sentosa, Singapore, Dec. 14–18, pp. 8–15.
- [39] Gu, X., See, K. W., Li, P., Shan, K., Wang, Y., Zhao, L., Lim, K. C., and Zhang, N., 2023, "A Novel State-of-Health Estimation for the Lithium-Ion Battery Using a Convolutional Neural Network and Transformer Model," *Energy*, **262**, p. 125501.
- [40] Gomez, W., Wang, F.-K., and Chou, J.-H., 2024, "Li-Ion Battery Capacity Prediction Using Improved Temporal Fusion Transformer Model," *Energy*, **296**, p. 131114.
- [41] Ma, Y., Shan, C., Gao, J., and Chen, H., 2022, "A Novel Method for State of Health Estimation of Lithium-Ion Batteries Based on Improved LSTM and Health Indicators Extraction," *Energy*, **251**, p. 123973.
- [42] Hu, X., Li, S., and Peng, H., 2012, "A Comparative Study of Equivalent Circuit Models for Li-Ion Batteries," *J. Power Sources*, **198**, pp. 359–367.
- [43] Yao, X.-Y., Chen, G., Pecht, M., and Chen, B., 2023, "A Novel Graph-Based Framework for State of Health Prediction of Lithium-Ion Battery," *J. Energy Storage*, **58**, p. 106437.
- [44] Guo, P., Cheng, Z., and Yang, L., 2019, "A Data-Driven Remaining Capacity Estimation Approach for Lithium-Ion Batteries Based on Charging Health Feature Extraction," *J. Power Sources*, **412**, pp. 442–450.
- [45] Beganovic, N., and Söffker, D., 2019, "Estimation of Remaining Useful Lifetime of Lithium-Ion Battery Based on Acoustic Emission Measurements," *ASME J. Energy Resour. Technol.*, **141**(4), p. 041901.
- [46] Li, Y., Stroe, D.-I., Cheng, Y., Sheng, H., Sui, X., and Teodorescu, R., 2021, "On the Feature Selection for Battery State of Health Estimation Based on Charging–Discharging Profiles," *J. Energy Storage*, **33**, p. 102122.
- [47] NASA Open Data Portal. "Li-Ion Battery Aging Datasets." <https://data.nasa.gov/dataset/Li-ion-Battery-Aging-Datasets/uj5r-zjdb>, Accessed March 11, 2023.
- [48] Khumprom, P., and Yodo, N., 2019, "A Data-Driven Predictive Prognostic Model for Lithium-Ion Batteries Based on a Deep Learning Algorithm," *Energies*, **12**(4), p. 660.
- [49] Kirchev, A., 2015, "Chapter 20—Battery Management and Battery Diagnostics," *Electrochemical Energy Storage for Renewable Sources and Grid Balancing*, P. T. Moseley, and J. Garche, eds., Elsevier, Amsterdam, The Netherlands, pp. 411–435.
- [50] Wang, H., Frisco, S., Gottlieb, E., Yuan, R., and Whitacre, J. F., 2019, "Capacity Degradation in Commercial Li-Ion Cells: The Effects of Charge Protocol and Temperature," *J. Power Sources*, **426**, pp. 67–73.
- [51] Xiong, R., Sun, Y., Wang, C., Tian, J., Chen, X., Li, H., and Zhang, Q., 2023, "A Data-Driven Method for Extracting Aging Features to Accurately Predict the Battery Health," *Energy Storage Mater.*, **57**, pp. 460–470.
- [52] Wu, N., Green, B., Ben, X., and O'Banion, S., 2020, "Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case".
- [53] Kunlong, C., Jiuchun, J., Fangdan, Z., Bingxiang, S., and Yanru, Z., 2016, "SOH Estimation for Lithium-Ion Batteries: A Cointegration and Error Correction Approach," *2016 IEEE International Conference on Prognostics and Health Management (ICPHM)*, Ottawa, ON, Canada, June 20–22 pp. 1–6.