Proceedings of the ASME 2023
International Design Engineering Technical Conferences and
Computers and Information in Engineering Conference
IDETC/CIE2023
August 20-23, 2023, Boston, Massachusetts

DETC2023-116426

STATE OF HEALTH ESTIMATION OF ELECTRIC VEHICLE BATTERIES USING TRANSFORMER-BASED NEURAL NETWORK

Yixin Zhao
Graduate Research Assistant
Environmental Engineering Sciences
University of Florida, Gainesville, FL, 32611
yixin.zhao@ufl.edu

Sara Behdad*
Associate Professor
Environmental Engineering Sciences
University of Florida, Gainesville, FL, 32611
sarabehdad@ufl.edu

ABSTRACT

Electric vehicles (EVs) are considered an environmentally friendly option 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 aid in the prevention of 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 is capable of capturing 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 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 effectively improve the accuracy of the model as well as the computational efficiency. The proposed standard Transformer shows good performance in SOH prediction.

Keywords: Lithium-ion battery; Electric vehicle; State of health; Transformer

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 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 need to be inferred by building degradation models [12–14]. An effective battery management system relies on a reliable battery degradation model for status monitoring and health state assessment to ensure their safe and high-performance operation.

There are two broad approaches to developing a battery degradation model for predicting and diagnosing battery in-use SOH: 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]. But 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 as well as its physical and chemical properties. In particular, the machine-learning based approaches can be trained on large datasets quickly, which makes them ideal 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].

Many studies have proven that recurrent neural network (RNN) based neural networks, including gate recurrent unit (GRU) and long short-term memory (LSTM), are effective 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 Transformers-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, Section 5 concludes the work in this study and the perspectives for future research.

2. RELATED WORK

2.1 Deep learning for SOH 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 are focused on the application of 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 feedforward 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 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. successfully implemented the Bi-LSTM for SOH prediction, and the model can explore the degradation behavior of Li-ion batteries deeply from two sequence directions [26]. The results demonstrate that Bi-LSTM can achieve more accurate SOH estimation than a single LSTM. Meanwhile, the bi-directional behavior makes the Bi-LSTM model significantly slower to compute than LSTM [34]. For RUL and SOH prediction, Qu et al. combined an 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 other deep learning model that constructs key modules with attention mechanisms is the Transformer [29]. There is no recurrence or convolutional mechanism in Transformer. It has the ability to 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 [36]. In addition, it enables faster training than LSTM since the majority of operations can be computed in parallel [37]. Due to its ability to handle sequence data, Transformer has the potential to be employed in battery degradation modeling. To the best of our knowledge, there are few works that 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 single LSTM and required significantly less training time than LSTM. Mo et al. developed a Transformer encoder based neural network enhanced by gated convolutional

unit, which achieved relatively good performance and was computationally efficient [36]. These two studies prove the effectiveness 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 in standard Transformer for battery state estimation was not discussed.

2.2 Battery feature extraction

An additional way to enhance 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 [38,39]. The measured features are extracted from the current, voltage, and temperature signals during the battery cycle that are available from the battery management system [40]. For example, Guo et al. extracted 14 features from the charging process, which are related to capacity, charge time, temperature, and current/voltage drop [41]. Their results show that after selection using gray relational analysis and dimensionality reduction by PCA, the features that remain can well reflect the internal aging process of the battery.

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 [39]. For example, Li et al. applied the incremental capacity analysis (ICA) technique to process battery voltage data [42]. This procedure could convert a flat voltage profile into an incremental capacity (IC) profile, which would have 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 more efficient battery SOH prediction system. Therefore, the measured features from the previous studies are referred to be used for feature extraction [38,40–42].

3. METHODOLOGY

3.1 Framework of the SOH estimation

A crucial factor to characterize the performance the battery is capable of in its present state is the state of health (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

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

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 makes use of the 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. Following the extraction of the features from the measured parameters, a correlation analysis is performed to evaluate the features' potential to capture battery degradation. Based on selecting valuable features, a standard transformer neural network and an encoder-only transformer are employed to estimate the SOH.

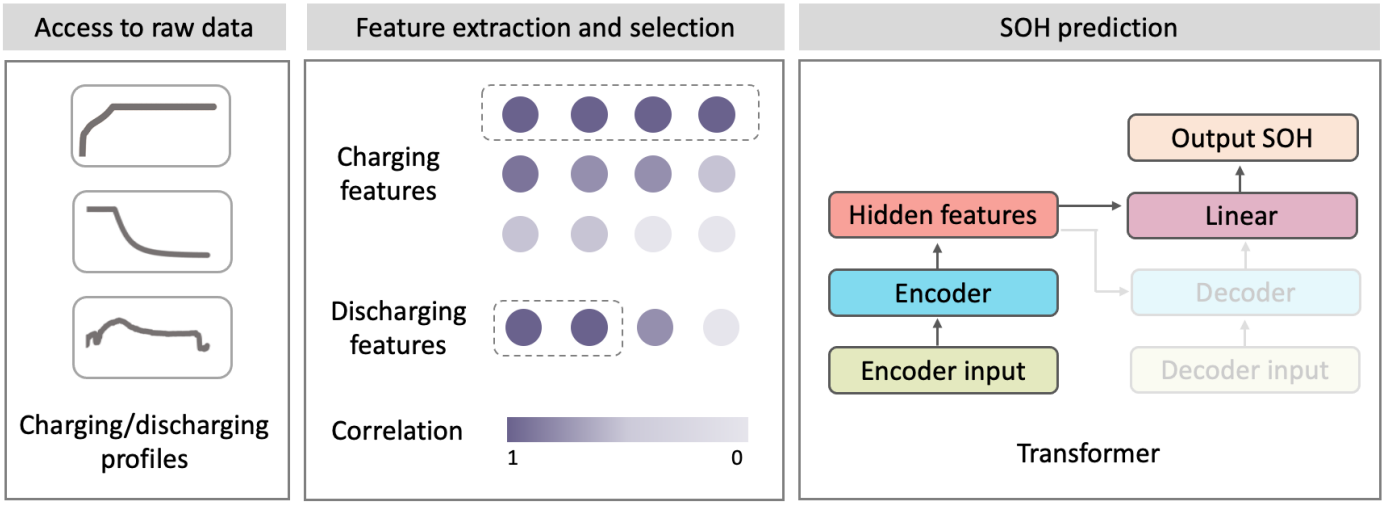

FIGURE 1: FRAMEWORK OF THE SOH ESTIMATION

3.2 Data description

The data are obtained from the NASA lithium-ion battery aging dataset [31]. This public dataset tested the 18,650 commercially available lithium-ion rechargeable batteries 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 [44]. 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 as it was subjected 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 Figure 2.

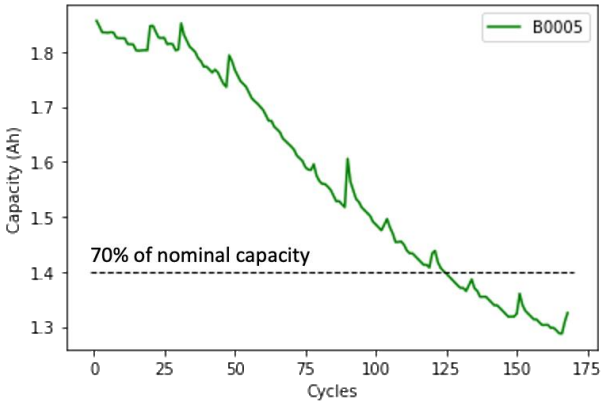

FIGURE 2: CAPACITY DEGRADATION OF BATTERY 0005

3.3 Feature extraction and selection

3.3.1 Feature extraction

The raw data for feature extraction is selected as the full range of V, I, and T signals measured in the state of charging and discharging [45]. The changes in the charging and discharging profiles as the battery ages are illustrated in Figure 3.

From Equation (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 [46]. 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 CC mode. From Figure 3(a) and (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 [47]. While the more time spent in CV mode, the more difficult it proves to be for the lithium ions to migrate inside the battery [45]. Consequently, it indicates the aggravation of the aging mechanism, such as elevated impedance and the formation of solid electrolyte interphase (SEI). 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 (Figure 3(c)). 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, cut-off current, and the maximum voltage are fixed, these three values are not considered in the extraction of features. However, temperature is considered valuable to be extracted and evaluated since temperature is one of the significant factors affecting the aging rate of the battery.

Figure 3(d)/(e)/(f) presents 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 cut-off 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 [46]. 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 [38,40,45], 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 Figure 4.

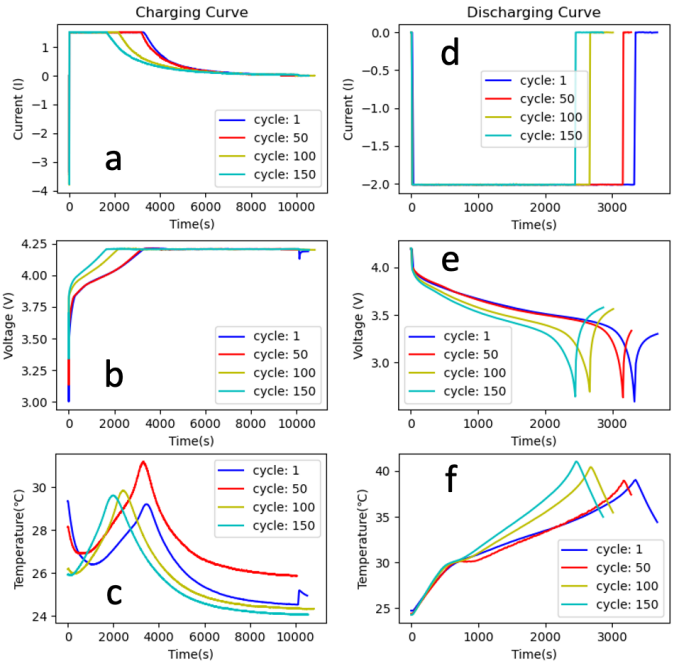

FIGURE 3: CURRENT, VOLTAGE, AND TEMPERATURE CURVES DURING CHARGING AND DISCHARGING OF BATTERY 0005 AFTER DIFFERENT NUMBER OF CYCLES

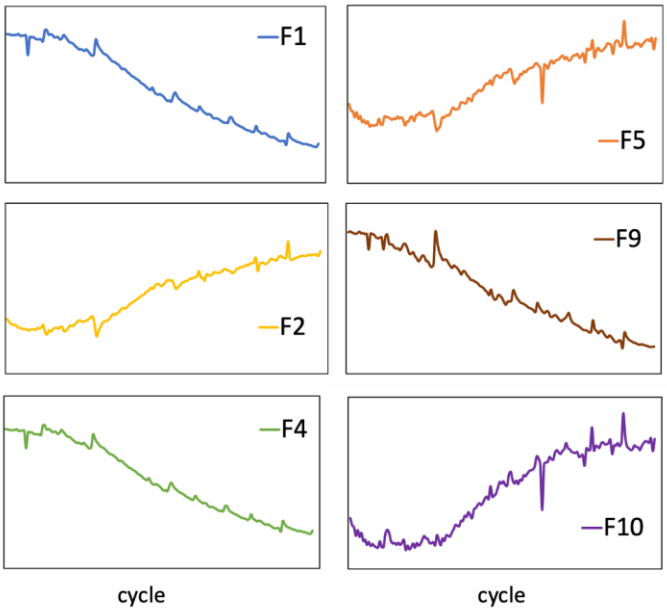

FIGURE 4: PROFILES OF EXTRACTED FEATURES F1, F2, F4, F5, F9, AND F10

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 [48]. 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. The feature selection using the Pearson correlation coefficient in this study has the following two steps.

- 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. The pair of features with an absolute PCC value greater than 0.999 is considered "duplicate" features. 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.

TABLE 1: FEATURE EXTRACTION [38,40]

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 CV charging	
	F3		Minimum slope of current curves in CV charging	
	F4		Area covered by voltage curves of CC charging	Voltage-related
	F5		Area covered by voltage curves of CV charging	
	F6		Maximum slope of voltage curves in 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	
	F11		Maximum temperature minus minimum temperature of charging	
	F12		The ratio of CC charging area under temperature curve to the corresponding area under the current curve	
	F13		The ratio of CV charging area under temperature curve to the corresponding area under the current curve	
	F14		CC charging time	Time-related
	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	
	F26		Time discharged under a constant current	Time-related
	F27		Time to the minimum voltage in discharging	
	F28		Time to the max temperature in discharging	

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, stack-identical encoder layers make up the encoder module [49]. Through a fully-connected network, the input layer converts the input time series data to a vector of self-defined dimensions (d_model). The positions of the data points in the sequence are characterized by the position encoding layer performing sine and cosine functions, giving each position a unique representation between 0 and 1. The encoder layer is tasked with generating encodings that contain information about which parts of the input are relevant to each other [50]. There are two sub-layers that comprise each encoder layer: multi-Head self-attention and a fully-connected feed-forward. Each sub-layer is followed by a normalization layer. The encoder generates a vector of dimension d_model for utilization by subsequent decoders.

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. 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. The decoder layer is similar to the encoder layer in that it includes 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. 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 (Figure 5). With an encoder-only Transformer, the hidden features output from the encoder enters directly into the final fully-connected layer for prediction.

4. RESULTS AND DISCUSSION

4.1 Feature selection results

According to the features described in Table 1, twenty-eight features are extracted from the Battery 0005 in the NASA dataset. The correlation coefficient between features and battery SOH is summarized in Table 2 in descending order [38,41,45]. The results show that the features with a PCC greater than 0.9 with 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, F28 from discharging group. The feature least associated with SOH is the minimum temperature during charging and discharging. This is due to the fact that 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.

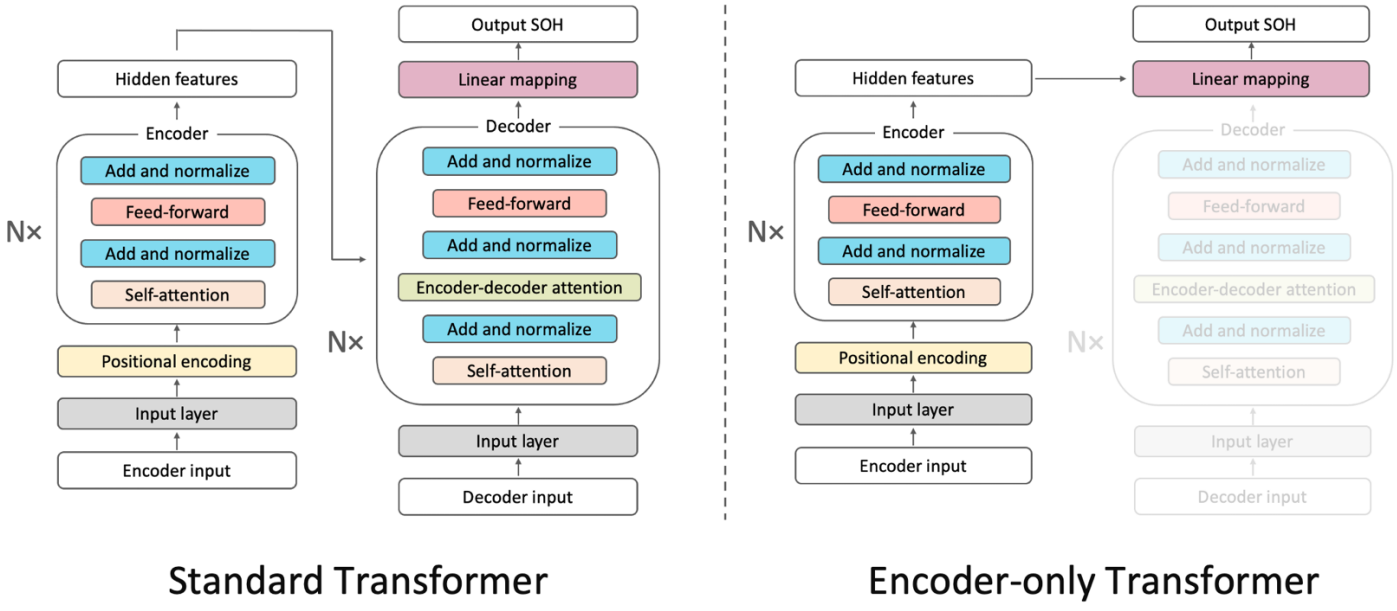

FIGURE 5: STRUCTURE OF STANDARD TRANSFORMER AND ENCODER-ONLY TRANSFORMER

In this particular dataset, there are multiple battery features from charging and discharging that 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 linear regression (LR) to fit the SOH curve would also achieve good performance. However, LR model is 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.

TABLE 2: PEARSON CORRELATION COEFFICIENT (ABSOLUTE VALUE) BETWEEN FEATURES AND SOH IN DESCENDING ORDER

Group	Feature	PCC with SOH (absolute value)
Charging features	**F4**	**0.9962**
	F14	**0.9961**
	F1	**0.9961**
	F16	**0.9936**
	F9	**0.9898**
	F2	**0.9877**
	F15	**0.9757**
	F5	**0.9756**
	F10	**0.9692**
	F18	**0.9578**
	F17	0.7744
	F6	0.6366
	F11	0.6341
	F3	0.6098
	F13	0.6000
	F12	0.4008
	F7	0.3829
	F8	0.0159
	F19	0.0063
Discharging features	**F26**	**0.9999**
	F20	**0.9999**
	F27	**0.9999**
	F28	**0.9999**
	F21	**0.9999**
	F22	**0.9927**
	F25	**0.9788**
	F23	**0.9369**
	F24	0.0519

* The bold are the features whose PCC is higher than 0.9

The features with high correlation from the charging data are mainly related to the start and end time of the CC and CV step-charging. The available capacity of a battery is determined by the amount of electric charge accumulated during charging, which is why the duration of the two phases of charging is critical to SOH estimation. On the other hand, the features from the discharge data, except for the minimum temperature, are all highly correlated with SOH. In particular, F20, F21, F26, F27, and F28 have a high correlation of 0.9999 with the predicted target, which can be almost linear.

The reason for this is that the experimental condition of Battery 0005 is a continuous release at a constant current during the discharge process until the cut-off 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 [46]. 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 the prediction of 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 performance of the model with only charging features is also tested and compared to the model with full feature input.

The similarity between high-correlation features is also calculated and presented in Figure 6. According to the feature selection rules defined in Section 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, F18 from charging group, and F20, F21, F22, F23, F25 from discharging group.

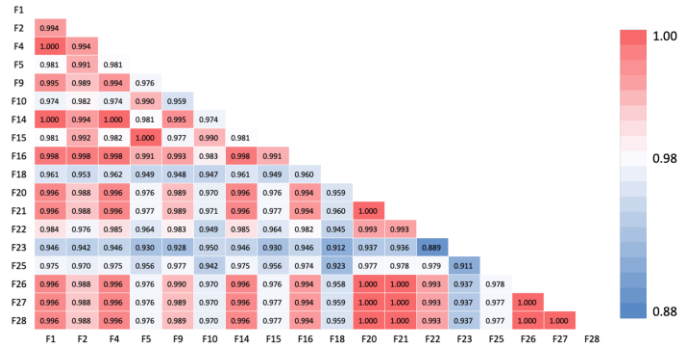

FIGURE 6: SIMILARITY BETWEEN FEATURES WITH CORRELATION COEFFICIENTS GREATER THAN 0.9 WITH SOH

4.2 SOH estimation results

4.2.1 Experiments implementation details

The two steps of feature selection as described in Section 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 has a very high linear correlation with SOH. It is not reasonable to use these discharge features to examine the effectiveness of the developed time series prediction model. 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). By comparing the performance of models with different input groups, the effect of feature selection and the effect of discharge features are discussed in the following sections.

TABLE 3: FOUR INPUT GROUPS

Input group	Description	Feature
1	All the 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

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 utilizes the Mean Squared Error (MSE) loss function during the training process. The dataset is split into a 70% training set and a 30% test set. The developed Transformer-based models have seven key parameters: number of time steps to be input (sampler size), number of expected features in the transformer encoder/decoder inputs (d_model), number of encoder layers, number of decoder layers, number of heads in the multi-head attention mechanism, 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 provides a summary of the range of grid search and optimized hyperparameter settings for the Transformer-based models. The process of parameter optimization 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.

Root mean square error (RMSE) and mean absolute error (MAE) are utilized as evaluation criteria to assess the efficacy of the designed model for 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.

TABLE 4: GRID SEARCH AND OPTIMIZED HYPERPARAMETERS SETTING FOR THE 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

4.2.2 Model performance

The performance of the two Transformers-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 are the ones with the best performance.

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 SOH prediction. The SOH prediction curves from both Transformer-based models are shown in Figure 7 and Figure 8. 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 its computational speed. Without decoder module, it runs over 30% faster than the standard Transformer. 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-time monitoring.

On the other hand, the effect of the features from the discharge data on the model is worth being discussed. As explained in Section 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 that have such a high correlation with the predicted target on model performance is apparent. As shown in Figures 7 and 8, models that employed discharge features, with or without a decoder module, can be more accurate in prediction than models that use only charging features.

TABLE 5: PERFORMANCE OF TWO TRANSFORMER-BASED MODES WITH DIFFERENT INPUT

Input group	Standard Transformer			Encoder-only Transformer		
	RMSE	MAE	Time cost (s)	RMSE	MAE	Time cost (s)
1	0.0359	0.0322	55	0.0553	0.0465	37
2	0.1109	0.1021	49	0.1073	0.0958	27
3	0.0238	0.0212	42	0.0672	0.0614	30
4	0.0290	0.0258	44	0.2180	0.2030	29

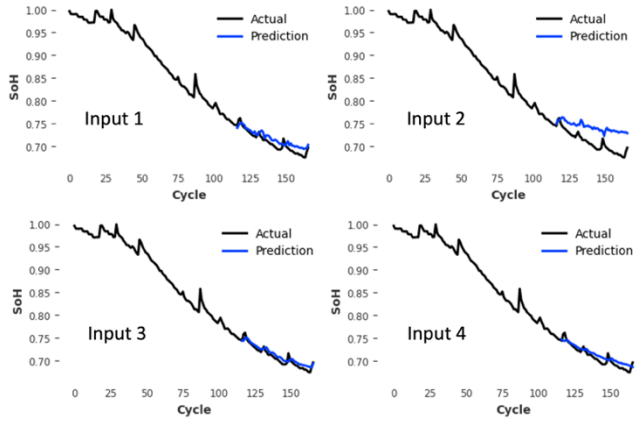

FIGURE 7: PREDICTED RESULTS FOR STANDARD TRANSFORMER

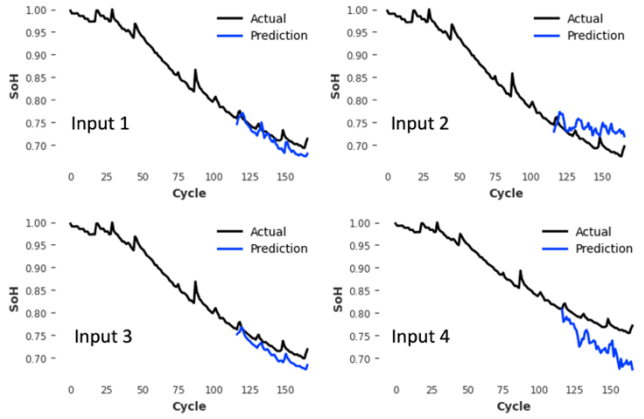

FIGURE 8: PREDICTED RESULTS FOR ENCODER-ONLY TRANSFORMER

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 performance of the model in the current dataset, the effectiveness of this type of feature in actual EV battery SOH prediction needs to be verified. With the absence of 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 0.029 and 0.0258, respectively. Regarding the computational cost of feature extraction, it takes about 20 seconds to compute features for 168 cycles of charge/discharge data.

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 in this respect 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 0.1109 and 0.1021, respectively. After feature selection, the RMSE and MAE of standard Transformers with input group 4 are reduced to 0.0290 and 0.0258. This demonstrates the effectiveness of the feature selection rule proposed in the paper. In addition to the improvement in the accuracy of the model, the computation time of the model has been improved by around 10%. Fast and accurate are the characteristics required for EV battery management systems. Our results demonstrate that feature extraction and selection is a promising direction to be explored for EV batteries. In addition, feature optimization, such as noise reduction and enhancement, is also an interesting problem to be explored in the future.

The proposed model is also performed on other batteries from the NASA dataset. The results show that fine-tuning of 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 standard Transformers is compared with LSTM and GRU in Table 6, that are widely used in the field of battery SOH estimation. All three models use the input group 4 and the number of time steps to be input is 15. The parameter settings for the standard transformer 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 the best prediction results while using the lowest time cost. 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.

TABLE 6: COMPARISON OF MODEL COMPUTATIONAL COSTS

Model	RMSE	MAE	Time cost (s)
Standard Transformer	0.0290	0.0258	44
LSTM	0.0637	0.0594	60
GRU	0.0345	0.0267	54

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 are varied. Therefore, the performance comparison here is rough and not entirely fair. The results of the comparison are presented in Table 7. The results show that our standard Transformers have advantages in terms of accuracy compared to single models, such as LSTM and ARIMA. However, advanced variants of the simple model, such as LSTM with the addition of enhanced optimization mechanisms as well as self-attention mechanisms and bi-directional 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.

TABLE 7: COMPARISON OF MODEL PREDICTION RESULTS WITH OTHER LITERATURE

Model	RMSE	MAE	Reference
Standard Transformer	0.0290	0.0258	This study
LSTM	0.0387	0.0439	[50]
Autoregressive Integrated Moving Average Model (ARIMA)	0.0375	-	[51]
LSTM with swarm optimization and attention mechanism	0.006	-	[35]
Bi-LSTM	0.005	-	[26]

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, as well as the similarity between features, is subsequently evaluated using the Pearson correlation coefficient. In 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 performance of the standard Transformer, as well as 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 0.029 and 0.0258, respectively. In addition, the discharge characteristics of the cells with constant-current continuous discharge behavior effectively improve the accuracy of the models. However, the constant and stable discharge behavior is the result of controlled experimental conditions. The effectiveness of the discharge characteristics in predicting the state of batteries with random discharge behavior remains to be verified.

Future research directions can focus on feature extraction for partial charge-discharge profiles. Most of the current prediction models for SOH use the full range of charge and discharge data. However, in real driving situations, EV batteries do not always complete a whole charging or discharging cycle. Therefore, it would be more practical to use partial battery data for state prediction.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation–USA under grant # 2026276. 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.

REFERENCES

[1] 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," Journal of Environmental Management, **245**, pp. 432–446.

[2] 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," International Journal of Energy Research, **34**(2), pp. 107–115.

[3] Capitaine, J.-A., and Wang, Q., 2018, "Design of a Test Platform for the Determination of Lithium-Ion Batteries State of Health," Journal of Mechanical Design, **141**(021702).

[4] Quijano-Ortiz, F., and Seepersad, C., 2022, "Design Recommendations for Reducing the Environmental Impact of Battery Packs," American Society of Mechanical Engineers Digital Collection.

[5] 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," International Journal of Hygiene and Environmental Health, **222**(4), pp. 635–644.

[6] Rodgers, L., Karplus, P., Gogoana, R., and Nawrot, M., 2011, "Rapidly Charging Battery Systems," American Society of Mechanical Engineers Digital Collection, pp. 265–273.

[7] 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," Journal of Power Sources, **262**, pp. 129–135.

[8] Tanim, T. R., Dufek, E. J., Evans, M., Dickerson, C., Jansen, A. N., Polzin, B. J., Dunlop, A. R., Trask, S. E., Jackman, R., Bloom, I., Yang, Z., and Lee, E., 2019, "Extreme Fast Charge Challenges for Lithium-Ion Battery: Variability and Positive Electrode Issues," J. Electrochem. Soc., **166**(10), p. A1926.

[9] Kim, T. J., Youn, B. D., and Kim, H. J., 2014, "Online-Applicable Temperature Prediction Model for EV Battery Pack Thermal Management," American Society of Mechanical Engineers Digital Collection.

[10] Noura, N., Boulon, L., and Jemeï, S., 2020, "A Review of Battery State of Health Estimation Methods: Hybrid Electric Vehicle Challenges," World Electric Vehicle Journal, **11**(4), p. 66.

[11] 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," American Society of Mechanical Engineers Digital Collection.

[12] Li, W., Sengupta, N., Dechent, P., Howey, D., Annaswamy, A., and Sauer, D. U., 2021, "One-Shot Battery Degradation Trajectory Prediction with Deep Learning," Journal of Power Sources, **506**, p. 230024.

[13] 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," American Society of Mechanical Engineers Digital Collection.

[14] Xi, Z., Jing, R., and Lee, C., 2016, "Diagnostics and Prognostics of Lithium-Ion Batteries," American Society of Mechanical Engineers Digital Collection.

[15] 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 and Sustainable Energy Reviews, **56**, pp. 1008–1021.

[16] 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.

[17] Han, S., Tang, Y., and Khaleghi Rahimian, S., 2021, "A Numerically Efficient Method of Solving the Full-Order Pseudo-2-Dimensional (P2D) Li-Ion Cell Model," Journal of Power Sources, **490**, p. 229571.

[18] S. Edge, J., O'Kane, S., Prosser, R., D. Kirkaldy, N., N. Patel, A., Hales, A., Ghosh, A., Ai, W., Chen, J., Yang, J., Li, S., Pang, M.-C., Diaz, L. B., Tomaszewska, A., Waseem Marzook, M., N. Radhakrishnan, K., Wang, H., Patel, Y., Wu, B., and J. Offer, G., 2021, "Lithium Ion Battery Degradation: What You Need to Know," Physical Chemistry Chemical Physics, **23**(14), pp. 8200–8221.

[19] 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," Applied Energy, **159**, pp. 285–297.

[20] 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," Journal of Cleaner Production, **212**, pp. 240–249.

[21] Lee, S., Cui, H., Rezvanizaniani, M., and Ni, J., 2013, "Battery Prognostics: SoC and SoH Prediction," American Society of Mechanical Engineers Digital Collection, pp. 689–695.

[22] 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," International Journal of Hydrogen Energy, **44**(23), pp. 12270–12276.

[23] 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 Transactions on Industrial Electronics, **66**(10), pp. 7706–7716.

[24] Bai, G., and Wang, P., 2015, "A Self-Cognizant Dynamic System Approach for Health Management: Lithium-Ion Battery Case Study," American Society of Mechanical Engineers Digital Collection.

[25] 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," Applied Artificial Intelligence, **35**(6), pp. 421–439.

[26] Sun, H., Sun, J., Zhao, K., Wang, L., and Wang, K., 2022, "Data-Driven ICA-Bi-LSTM-Combined Lithium Battery SOH Estimation," Mathematical Problems in Engineering, **2022**, p. e9645892.

[27] 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," Journal of Energy Storage, **32**, p. 101741.

[28] 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," Journal of Energy Storage, **21**, pp. 510–518.

[29] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I., 2017, "Attention Is All You Need."

[30] 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," Journal of 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 Transactions on Vehicular Technology, **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," International Journal of Energy Research, **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)*, 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] 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.

[37] 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)*, pp. 8–15.

[38] 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.

[39] Hu, X., Li, S., and Peng, H., 2012, "A Comparative Study of Equivalent Circuit Models for Li-Ion Batteries," Journal of Power Sources, **198**, pp. 359–367.

[40] 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," Journal of Energy Storage, **58**, p. 106437.

[41] 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," Journal of Power Sources, **412**, pp. 442–450.

[42] 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," Journal of Energy Storage, **33**, p. 102122.

[43] "Li-Ion Battery Aging Datasets | NASA Open Data Portal" [Online]. Available: https://data.nasa.gov/dataset/Li-ion-Battery-Aging-Datasets/uj5r-zjdb. [Accessed: 11-Mar-2023].

[44] 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.

[45] 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," Journal of Energy Storage, **33**, p. 102122.

[46] 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, pp. 411–435.

[47] 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," Journal of Power Sources, **426**, pp. 67–73.

[48] 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 Materials, **57**, pp. 460–470.

[49] Wu, N., Green, B., Ben, X., and O'Banion, S., 2020, "Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case."

[50] 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.

[51] 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)*, pp. 1–6.