A Flexible State-of-Health Prediction Scheme for Lithium-Ion Battery Packs With Long Short-Term Memory Network and Transfer Learning

Xing Shu, Graduate Student Member, IEEE, Jiangwei Shen, Guang Li, Senior Member, IEEE, Yuanjian Zhang, Member, IEEE, Zheng Chen, Senior Member, IEEE, and Yonggang Liu, Senior Member, IEEE

Abstract—The application of machine learning-based state-of-health (SOH) prediction is hindered by the large demand for training data. To conquer this defect, a flexible and easily transferred SOH prediction scheme for lithium-ion battery packs is developed. First, the charging duration for a predefined voltage range is hired as the health feature to quantify capacity degradation. Then, the long short-term memory (LSTM) network and transfer learning (TL) with fine-tuning strategy are incorporated to constitute the cell mean model (CMM) for SOH prediction with partial training data. Next, to evaluate the SOH inconsistencies among cells, the LSTM model is employed as the cell difference model (CDM), and the minimum estimation value of CDM is identified to determine a specific SOH. The experimental results reveal that, even when the first 360 cycle data, occupying only 40% in the whole 904 cycle data, are chosen and constituted to the data set for model training, the obtained estimation algorithm can still predict SOH precisely with the error of less than 3%, thus remarkably reducing the training data amount and mitigating the computation burden during model training. In addition, the preferable validation results on different types of lithium-ion batteries further manifest the extendibility of the proposed strategy.

Index Terms—Lithium-ion battery pack, long short-term memory (LSTM), state of health (SOH), transfer learning (TL).

Manuscript received November 28, 2020; revised January 30, 2021 and March 18, 2021; accepted April 13, 2021. Date of current version October 1, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB0104000, in part by the National Natural Science Foundation of China under Grant 61763021, and in part by the EU-funded Marie Skłodowska-Curie Individual Fellowships under Grant 845102.

1) Department of Mechanical Engineering, University of California, Berkeley, CA 94720, USA (e-mail: xshing92@kust.edu.cn; shenjiangwei6@163.com).
2) School of Mechanical and Aerospace Engineering, Queen’s University Belfast, Belfast BT9 5AG, U.K. (e-mail: g.li@qmul.ac.uk).
3) School of Engineering and Materials Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: y.zhang@qub.ac.uk).
4) School of Transportation Engineering, Kunming University of Science and Technology, Kunming 650000, China (e-mail: zhengchen2019@163.com).
5) State Key Laboratory of Mechanical Transmissions, Chongqing University, Chongqing 400044, China, and also with the College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China (e-mail: andylgy@umich.edu).

Digital Object Identifier 10.1109/TTE.2021.3074638

I. INTRODUCTION

Due to the massive consumption of fossil fuel and pressing concerns over carbon emissions, high-efficiency and environment-friendly electric vehicles (EVs) have become the development mainstream in transportation electrification [1], [2]. In EVs, lithium-ion batteries have been widely deployed as main energy sources, with a form of a pack consisting of tens to thousands of cells connected in series, parallel, or mixed topologies [3], [4]. However, with the increase in operation, battery performance degrades gradually, including capacity decline, resistance increase, and inconsistency aggravation among cells [5]. It is, therefore, critical to reliably estimate the healthy status of lithium-ion batteries, referred to as state-of-health (SOH) prediction [6]. Accurate knowledge of SOH can not only contribute to the estimation promotion of state of charge (SOC) but also facilitate the safe operation of batteries [7].

Battery SOH represents the ratio of maximum discharging capacity over the rated value [8]. Generally, the prediction methods of SOH can be sorted into three groups: direct calibration methods, filter-based methods, and machine learning-based methods [9]. Direct calibration methods determine battery SOH via specific experimental operations, such as full discharge of the battery after a complete charge [10]. This kind of method is simple and easy to implement. Nevertheless, it is an open-loop approach highly depending on the acquisition accuracy of current and voltage; furthermore, it is impractical to fully charge and discharge the batteries equipped in EVs. To cope with the restrictions of direct calibration methods, filter-based methods are developed, and they treat SOH as one parameter waiting to be identified in the model or one state variable that needs to be estimated in the built observer [11]. Ordinary filter algorithms, such as the extended Kalman filter (EKF) [12], the particle filter [13], and the H-infinity filter [14], have been intensively applied for online SOH estimation. However, the performance of these methods highly relies on the accuracy and robustness of the established battery model or observer.

Machine-learning-based methods have been progressively exploited to predict SOH due to their strong data processing
and nonlinear fitting capabilities [15]. These methods usually employ different machine learning algorithms to capture the degradation trend from measurement and generate healthy features representing SOH variation [16]. Tian et al. [17] reveal that the partial differential temperature variation is related to SOH degradation, and then, the support vector regression (SVR) is exploited to predict SOH based on temperature variation. In [18], the Gaussian process regression (GPR) is introduced to map the relationship between partial capacity increase and SOH. In addition to SVR and GPR methods, the random forest regression (RFR) [19], the radial basis function neural network (RBFNN) [20], the relevance vector machine (RVM) [21], and the long short-term memory (LSTM) network [22] are also successfully leveraged to estimate SOH. Among these machine learning algorithms, the LSTM network, as a typical representative of recurrent neural networks (RNNs), can capture the long-term dependence information through its specially designed gate structure and overcome the drawbacks of gradient explosion and gradient vanishing existing in RNN. Consequently, the LSTM network is widely adopted in state prediction due to its superior performance in capturing the long-term relationship of historical information [23].

Although the abovementioned approaches can achieve satisfactory SOH prediction, a large amount of training data is indispensable, undoubtedly increasing the duration of the off-line test. Moreover, when the battery type or the health feature is changed, the predetermined model needs to be reconstructed, or the model parameters need to be retrained, obviously complicating the algorithm design. Richardson et al. [24] exert the GPR to estimate SOH based on two different types of batteries. However, when the GPR model trained on the basis of data set 1 is transplanted to estimate the battery SOH described in data set 2, the repetitive operations, including reselecting the model parameters and retraining the model, need to be conducted once again. Similar difficulties are also encountered in [19], where the relative incremental capacity within three different voltage ranges is considered a healthy feature to characterize the SOH variation. However, the built model needs to be retrained with massive cycle data after the voltage range varies. In real-time applications, it is time-consuming, material-consuming, and labor-intensive to do so. How to mitigate the burdensome training job or reduce the required data amount when constructing the SOH model based on machine learning algorithms needs to be further investigated.

Another critical concern for SOH estimation is the inconsistency among battery cells, which is ineluctable in production and usage, especially for EV applications. To the best of our knowledge, there is still a lack of efficient manners to characterize the SOH of battery packs on the basis of the cell’s SOH. To diagnose the health state of lithium-ion battery packs, Zheng et al. [25] apply a multistage constant current (CC) charging strategy to accelerate battery pack degradation, and a dual GPR framework is advanced to simultaneously forecast SOH and remaining useful life (RUL). However, the whole pack is deemed as an integrated unit without considering the inconsistency among cells. In [26], to identify the inconsistency of cell SOC, a cell mean model (CMM) and a cell difference model (CDM) are, respectively, constructed to delineate the overall SOC of the pack and the difference between the cell SOC and mean value. Nevertheless, few methods have been devoted to the estimation of cell SOH inconsistency in battery packs. It is, therefore, imperative to design a straightforward and efficient pack SOH estimation method to fill in the gap of existing research.

Aiming to promote the potential of machine learning-based methods and investigate an efficient pack SOH prediction algorithm, a flexible and transferable SOH prediction scheme is advanced for lithium-ion battery packs. Concretely, an LSTM network integrating a fully connected layer is constructed to learn the long-term dependence of battery cell SOH according to the features generated from partial charging voltage data. On this basis, a transfer learning (TL) method, which is applied to improve the learning ability by transferring information from a related domain, is proposed for easy adaption to the variation of feature voltages existing in different batteries, thereby significantly mitigating the burdensome training task. In addition, to fully consider the inconsistency among battery cells, the CMM and CDM are, respectively, addressed to simulate mean SOH level of pack and difference of cell SOH, thereby efficiently identifying the pack SOH. Eight battery cells and three types of batteries are experimented with to verify the effectiveness of the proposed method. In addition, more in-depth validations on two different types of lithium-ion batteries further justify the extendibility and flexibility of the proposed algorithm.

The remainder of this article is structured as follows. The main methodologies, including LSTM and TL, are introduced in Section II. The experimental tests and generation process of health features are described in Section III. Next, the detailed prediction procedure of pack SOH is detailed in Section IV. After that, Section V analyzes and discusses the experimental validation results. Finally, Section VI draws the main conclusions.

II. METHODOLOGIES

The objective of this study is to construct a uniform scheme to forecast SOH with partial training data based on LSTM and TL. To attain it, the theoretical basis of both algorithms is detailed first.

A. Long Short-Term Memory Network

The LSTM network, as an extended form of RNN, exploits the memory units in place of conventional hidden nodes to
the output gate, the state cell, and the output data, respectively; the predicted value of SOH, and N means the data size of SOH.

B. Transfer Learning

Existing machine learning methods usually assume that the data obey the same distribution, and also a large amount of labeled data is required to train the model. However, in real applications, the domains of the source data set and the target data set are usually different, and the limited label data in the target data hindered the promotion of machine learning methods. To recognize the knowledge and skills learned from previous tasks and transfer them to novel domains/tasks, the TL based on deep neural networks is proposed and has attracted wide attention to now [31]. As an efficient data mining framework, TL can transfer the data or features from previous tasks (named the source domain hereinafter) to facilitate the model construction of the target task (called the target domain), thus remediing the difficulty of information loss caused by the data insufficiency. For compensating the data loss on degradation prediction and improve the SOH estimation precision, a flexibility scheme for SOH prediction of different types of batteries is presented by integrating LSTM and TL. Among all TL solutions, fine-tuning a pretrained model based on a new data set is the most popular manner for knowledge transfer in the machine learning field [32]. On this account, it is engaged in SOH estimation in this study, and the schematic of TL with fine-tuning strategy is exhibited in Fig. 2.

The network consists of five layers, where the first layer is the input layer, and the last layer is the output layer. The input on the left and right networks includes, respectively, the source data and the target data. First, the network is pretrained through the source data to obtain the base model. Then, the target data are set as the input variable, and the TL with fine-tuning strategy is activated to adjust the parameters of one middle layer, while the parameters of other layers remain unchanged. In this manner, a new transfer model is constructed. Since only partial parameters of the base model are adjusted, the training speed is much faster than that of retraining the whole model. Simultaneously, the proposed method can cooperatively transfer the cycle data and the formerly trained model of source battery to the newly built model of target batteries for easier SOH prediction. Its strong learning capability can bridge the serviceable information implied in the data and model between two domains to facilitate the prediction model construction of the target domain, thus mitigating the requirement of aging data and obviously improving the training efficiency.

After detailing the principles of LSTM and TL, the related experiments and SOH prediction based on the proposed methods will be investigated.

III. EXPERIMENTS AND FEATURE GENERATION

This section details the aging experiment procedures and specifies the feature generation from the test data for SOH estimation.

A. Aging Experiment

In this study, three different sources of battery data are considered: 1) the nickel cobalt manganese (NCM) data set from our laboratory; 2) the lithium cobalt oxide (LCO) data set from the Center for Advanced Life Cycle Engineering (CALCE), which is an open-access depository supplied by the University of Maryland [33]; and 3) the commercial lithium iron phosphate (LFP) battery data set from the Massachusetts Institute of Technology (MIT) and Stanford University [34]. These three data sets are named data set 1, data set 2, and data
set 3, respectively. The specifications of these three batteries are tabulated in Table I. The laboratory batteries are repeatedly experimented with according to a CC charge/discharge strategy at room temperature. During the charging process, 2-A current is imposed in the CC charging step, and 4.15 V is sustained in the constant voltage (CV) charging stage until the current drops to the predefined cutoff value. Then, the test batteries are discharged with the 1-C current, where C denotes the current drops to the predefined cutoff value. After 5-min rest. The batteries in data set 2 belong to commercial prismatic cells with 1100-milli-Ah (mAh) nominal capacity. Three cells, referred to as 2-1, 2-2, and 2-3, are aged with a similar cycle experiment setting in data set 1, and the charging current is set to 0.55 A. More detailed test steps can be found in [35]. For data set 3, the charging protocol and the chemistry property of batteries are different from the first two batteries. The cells are charged with a two-step fast-charging strategy: C1(Q1)-C2 scheme, in which C1 and C2, respectively, indicate the first and second step currents, and Q1 denotes the SOC at which the current switches from C1 to C2. The second step charge is terminated at 80% SOC, followed by the 1-C current-based CC-CV charge mode. Finally, a 4-C current is imposed to discharge the battery until its voltage reaches the minimum cutoff value [36]. The experimental data from the aging test are employed to extract battery health features, train the model, and test the developed battery SOH estimation algorithm. First, the extraction process of health features will be introduced.

B. Feature Generation

As discussed above, it is critical to find a representative health feature to properly assess the aging state. The health feature needs to be easy to obtain, simple to calculate, and efficient for anti-interference. In EV applications, since driving conditions and driving demands are time-varying and stochastic, the discharging process of lithium-ion batteries is usually stochastic and unforeseeable [36], and it is tricky to apply the discharge data for SOH estimation. Nonetheless, the charging process is generally uniform as the battery is usually charged from the grid according to specific instructions. Hence, the charging voltage and capacity profiles are designated for feature generation and SOH prediction. Taking data set 1 as an example, the charging voltage profiles under different cycles are plotted in Fig. 3(a), and we can find that the charging time decreases with the increase of cycle numbers. Intuitively, the charging time for a fixed voltage range will be shortened when the battery capacity degrades. On this account, the duration for a fixed voltage range can be deemed as a health feature to measure the battery degradation. To be specific, for each charging curve, the duration $T_i$, in the designated voltage range, i.e., $V_1$ to $V_2$, is considered as a health feature, which can not only capture the degradation of capacity and increment of resistance but also evaluate the inconsistency of battery cells across different cycles plotted in Fig. 3(a), and we can find that the charging time decreases with the increase of cycle numbers. Intuitively, the charging time for a fixed voltage range will be shortened when the battery capacity degrades. On this account, the duration for a fixed voltage range can be deemed as a health feature to measure the battery degradation. To be specific, for each charging curve, the duration $T_i$, in the designated voltage range, i.e., $V_1$ to $V_2$, is considered as a health feature, which can not only capture the degradation of capacity and increment of resistance but also evaluate the inconsistency of battery cells across different cycles.

In the next step, the SOH prediction scheme will be designed according to the proposed strategy and the identified healthy features.

IV. BATTERY PACK SOH PREDICTION SCHEME

To avoid over operation of the individual battery cell in packs, the minimum SOH value among that of cells is defined as the SOH of the battery pack. Due to imperfect fabrication techniques in the manufacturing process and inconsistent heat distribution inside the battery pack, the aging speed of cells is different. In this section, a two-part SOH estimation model incorporating the CMM and CDM is developed, in which CMM describes the mean SOH performance among cells, and CDM mainly accounts for the difference between cell SOH and the mean value.

A. Construction of Cell Mean Model

The constructed block diagram of the CMM is depicted in Fig. 4. First, the health features are extracted by the method defined in Section III, and then, the basic model is built by the LSTM method to map the relationship between health features and the low and high boundaries are assumed as $V_1$ and $V_2$, respectively.

2) With the age of batteries, the polarization becomes severer with the form of internal resistance increase, and concretely, the CC charging time decreases. On this account, the charging time for a certain voltage range can be hired to measure the battery degradation. To be specific, for each charging curve, the duration $T_i$, in the designated voltage range, i.e., $V_1$ to $V_2$, is considered as a health feature, which can not only capture the degradation of capacity and increment of resistance but also evaluate the inconsistency of battery cells across different cycles.

In the next step, the SOH prediction scheme will be designed according to the proposed strategy and the identified healthy features.

IV. BATTERY PACK SOH PREDICTION SCHEME

To avoid over operation of the individual battery cell in packs, the minimum SOH value among that of cells is defined as the SOH of the battery pack. Due to imperfect fabrication techniques in the manufacturing process and inconsistent heat distribution inside the battery pack, the aging speed of cells is different. In this section, a two-part SOH estimation model incorporating the CMM and CDM is developed, in which CMM describes the mean SOH performance among cells, and CDM mainly accounts for the difference between cell SOH and the mean value.

A. Construction of Cell Mean Model

The constructed block diagram of the CMM is depicted in Fig. 4. First, the health features are extracted by the method defined in Section III, and then, the basic model is built by the LSTM method to map the relationship between health features and the low and high boundaries are assumed as $V_1$ and $V_2$, respectively.

2) With the age of batteries, the polarization becomes severer with the form of internal resistance increase, and concretely, the CC charging time decreases. On this account, the charging time for a certain voltage range can be hired to measure the battery degradation. To be specific, for each charging curve, the duration $T_i$, in the designated voltage range, i.e., $V_1$ to $V_2$, is considered as a health feature, which can not only capture the degradation of capacity and increment of resistance but also evaluate the inconsistency of battery cells across different cycles.
and SOH. Finally, the TL method with fine-tuning strategy is exerted to establish the autonomous learning model and attain SOH prediction based on the reduced training data and mutative feature range. The whole prediction scheme mainly includes three steps.

1) Feature Extraction: Partial voltage data are acquired from the voltage curves during the CC charging process to generate the first health feature coefficient, and the specified charging interval for the preset voltage range is calculated as the second health feature factor. Both factors are selected as the input of the base model.

2) Base Model Construction: According to the off-line test results, partial measurement data are selected as the training dataset, and the remaining data are assigned as the validation object. The ratio of partitioning the data between training and testing will be discussed in Section V. Then, the LSTM network with five layers is implemented to construct the base model. As well known, increasing layers will complicate the calculation process and aggravate the computation intensity. However, when the number of layers is too small, the model accuracy will be deteriorated. Through optimization, the number of layers is set to five after trading off the prediction accuracy and calculation complexity. It is known that environment temperature and charging/discharging current can generate significant influences on the aging of lithium-ion batteries; however, it is quite difficult to fully account for the massive influences incurred by different operation conditions in the long-term prediction of SOH. Hence, in the proposed scheme, a one-step-ahead prediction method is exploited to predict battery SOH. That is, no matter how much the battery degrades, it is not necessary to know the previous capacity; only if the specific characteristic parameter, i.e., health feature is extracted, the trained model can be employed to predict battery SOH efficiently. In addition, although the current health feature is extracted to estimate the SOH, the long-term dependence during SOH estimation is merged due to the long-term memory characteristics of the LSTM algorithm. In this manner, the up-to-date aging state of lithium-ion batteries can be estimated comprehensively, and the impact of the working environment on battery degradation can be considered to some extent.

3) Transfer Model Construction: To ensure the flexibility of the proposed SOH prediction method, the TL with fine-tuning strategy is exploited to adjust the parameters of the base model. In this scheme, only the parameters of the bottom two layers, i.e., the fully connected layer and regression output layer, will be updated, and the parameters of other layers will remain unchanged. Note that, as the model parameters need to be adjusted according to the new features or battery types, the labeled data need to be prepared for retraining. In this manner, the transfer model is constructed.

B. Construction of Cell Difference Model

Here, we assume the SOH of the first cell in the pack as the CMM, the feature difference $\Delta T^i$, and the SOH difference $\Delta SOH^i$ between the mean value $SOH_{mean}$ and SOH$^i$ of the $i$th cell can be formulated as

\[
\Delta T^i = T^i - T_{mean}
\]

\[
\Delta SOH^i = SOH^i - SOH_{mean}
\]

where $T^i$ denotes the duration in the designated voltage range of cell $i$; $T_{mean}$ indicates the duration in the designated voltage range of mean cell SOH. $\Delta SOH$ is negative for lower SOH$^i$ and positive for higher SOH$^i$ compared to the mean value. To map the hidden relationship between $\Delta T^i$ and $\Delta SOH^i$, LSTM is employed to establish the CDM, in which $\Delta T^i$ and $\Delta SOH^i$ are assigned as the input and output. After attaining the mean value and the difference, the minimum difference under the same cycle is selected. Then, the minimum cell SOH, i.e., the pack SOH, can be deduced by the inverse transformation of (3). During the estimation process, only the duration for a preset voltage range needs to be determined online, and it is easy to acquire. Meanwhile, the real capacity or SOH of the training set needs to be known in advance for model training, and the real capacity of the test set is treated as the reference to evaluate the model performance.

In this manner, the SOH inconsistency within the battery pack can be well evaluated by the proposed CDM, and the proposed method can effectively distinguish the cell inconsistency and identify the battery pack SOH state. Next, the experimental validation will be conducted and discussed.

V. VALIDATION AND ANALYSIS

The performance of the proposed scheme for SOH prediction is evaluated by conducting a large number of comparisons raised by different predictors under various operation conditions. To demonstrate the feasibility and advancement of the proposed method, four representative tests are conducted, including the performance comparison of different predictors in the base model with training and testing data sets of different lengths, the portability over other voltage ranges and battery types, and validations on battery packs. The evaluation criteria include average absolute error (AAE), maximum absolute error (MAE), and RMSE; the AAE and MAE are formulated as

\[
\begin{align*}
AAE &= \frac{1}{N} \sum_{i=1}^{N} |SOH_{Ref} - SOH_{Pre}| \\
MAE &= \max \left( |SOH_{Ref} - SOH_{Pre}| \right)
\end{align*}
\]

A. Comparison of Different Methods on Base Model

Based on the experimental data, the feasibility of the proposed base model is first validated in terms of prediction of data set 1. Another three commonly used SOH
predictors, including backpropagation NN (BP-NN), SVM, and least-squares SVM (LS-SVM), are employed to conduct the estimation for performance comparison. The influence of voltage range selection is also addressed in this part. In the three commonly used predictors, the voltage range of 3.6–3.8 V is applied to extract the health feature. In the proposed method, three different ranges are considered for healthy feature generation. The first two ranges express individual voltage intervals, and the third one includes two ranges, as presented in Table II. For each prediction scenario, the first 70% of aging profiles are applied for model training, and all four methods are applied to forecast the same degradation data. Moreover, to fully promote the advantages of each network, the grid search method is leveraged to find the optimal parameters of each model.

The prediction performance of these four methods is compared with the reference curve, and the results are shown in Fig. 5 and Table II, respectively. As can be found, the capacity degradation trend is generally consistent, and no obvious rapid aging process emerges. Note that, as the phenomenon of capacity recovery exists due to the intermittent experiment, the capacity degradation curve is not monotonously decreasing. From Fig. 5 and Table II, it can be observed that all four approaches can effectively track the reference curve during the entire lifetime. The AAEs by the BP-NN, SVM, and LS-SVM are 0.43%, 0.50%, and 0.43%, respectively; in contrast, the largest AAE for the proposed scheme is less than 0.43%. After removing the estimation results in the initial stage, the MAE of the proposed method is much lower than that of the other three algorithms, except when two voltage ranges are selected to extract the health feature. In addition, for the BP-NN, SVM, and LS-SVM algorithms, their RMSEs are, respectively, 0.63%, 0.69%, and 0.60%, obviously higher than that of the proposed method. The running time of the proposed method is evaluated on a desktop computer, which is equipped with Intel Xeon E3-1230 (3.30 GHz) processor and 32-GB memory. The calculation time of the proposed method is 3.02 s when using data set 1 to predict SOH. Since SOH can be merely estimated once per cycle, the average running time of each cycle for data set 1 is 3.34 ms. Compared with the duration of one cycle experiment that lasts for a number of hours, the short estimation duration can fully meet the requirements of online implementation. By comparing the prediction results, it can be concluded that different voltage ranges as health features lead to different estimation performances. It is also distinct that the developed method with the input data under range from 3.6 to 3.8 V can raise better prediction performance compared to that with other voltage ranges. In addition, the model inputs with two voltage ranges will possibly increase the model instability and consequent computational complexity. Therefore, we can infer that the proposed algorithm with five layers and the health feature extracted from a single voltage range are feasible and reasonable.

### B. Performance Evaluation on Different Voltage Ranges

In previous discussions, a random voltage range is selected to construct the base model. To validate the adaptability of the proposed scheme, the voltage range for health feature extraction is modified to 3.5–4 and 3.5–4.15 V, significantly deviating from the preset voltage range of the base model. Fig. 6 displays the results for different voltage ranges by the proposed method and the single-LSTM method. In this case, the model trained by the first 360 cycle data (40% of the experimental data) is utilized to predict the SOH of the remaining 544 cycles. From Fig. 6, it can be found that the predicted SOH is, in general, less accurate than before, and the MAE becomes larger, due to the improper selection of the voltage range and the amount decrease of the training data set. Even so, the estimation performance based on the proposed method is still acceptable. However, the SOH curves obtained by the proposed method and the single-LSTM method show different trends, and the prediction profile of the former method can track the degradation curve more accurately during the whole process; nonetheless, the latter only follows the reference curve during training and gradually moves away from the true value during the test.

Similar to before, the proposed approach achieves more robust results with less training data and varying feature ranges. Fig. 6(c) shows the boxplot of the prediction errors across all the voltage ranges based on the proposed and individual LSTM methods, and the red lines represent the median value of error. We can find that the boxplots of LSTM- and TL-based methods for different voltage ranges are flatter and show that the error distribution of this method is more concentrated. Moreover, the median error line based on the LSTM and TL methods is closer to zero. The maximum error of the proposed method occurs in the initial cycle, which is included in the training set. However, the maximum error of the individual LSTM method appears in the test set and higher than 9%. It can be concluded that the proposed scheme outperforms the single-LSTM approach. Moreover, it also can be observed from the above analysis that the proposed method can result in precise SOH estimation with an MAE of 3.2%. Moreover, this method can reduce the training data amount.

### Table II

| Algorithm   | Voltage range | AAE (%) | MAE (%) | RMSE (%) |
|-------------|---------------|---------|---------|----------|
| BP-NN       | [3.6 3.8]     | 0.43    | 2.56    | 0.63     |
| SVM         | [3.6 3.8]     | 0.50    | 2.21    | 0.69     |
| LS-SVM      | [3.6 3.8]     | 0.43    | 2.63    | 0.60     |
|             | [3.6 3.8]     | 0.31    | 1.88    | 0.42     |
| Proposed    | [3.5 3.7]     | 0.36    | 1.81    | 0.54     |
|             | [3.5 3.7] &   | 0.42    | 2.38    | 0.59     |
|             | [3.6 3.8]     |         |         |          |
by about 30% (about 270 cycles). From the perspective of balancing data preparation and accuracy, it is still acceptable to sacrifice slight accuracy while reducing a variety of test data and saving much testing time.

### C. Performance Evaluation on Other Batteries

To verify the adaptability and scalability of the proposed scheme to other types of batteries, data sets 2 and 3 specified in Table 1 are applied to estimate the SOH via the proposed method, and the performance of data set 2 is presented in Fig. 7. Compared with data set 1, the capacity recovery phenomenon of data set 2 is more obvious. With the help of TL with fine-tuning strategy, the model is trained by the data of the first 100 cycles (occupying 18%–20% of the aging data), which is determined through trial and error. The remaining cycles are applied for validation, and the selected voltage interval for feature generation ranges from 3.8 to 4.1 V. Although only 100-cycle data are leveraged to train the model, the developed method leads to satisfactory performance during the entire aging process. In the worst case, the maximum AAE, MAE, and RMSE are 0.82%, 3.9%, and 1.07%, respectively. Note that the MAEs of the cells 2-1 and 2-2 exceed 3%, and it can be seen from Fig. 7(b), (e), and (h) that most of the errors can be restricted within 2.8%. Concretely, the error distribution can be schematically summarized in Fig. 7(c), (f), and (i). As can be found, the zero (or near) error emerges frequently. With the increase in the absolute value of error, the error occurrence probability decreases gradually, being in line with the Gaussian distribution with the mean value of 0.04%, 0.3%, and 0.006% and the variance of 0.0002, 0.00008, and 0.00006. The detailed validation results reveal that the distribution of error is reasonable and manifest that the proposed method can accurately and stably estimate battery SOH, even in the case of obvious size reduction of the training data set. Moreover, the proposed method can effectively reduce the degradation experiment by 285 cycles, which will cost around 855 h in this case. This, in turn, highlights the strong timeliness and high efficiency of the proposed framework when it is leveraged to predict the SOH of different batteries, manifesting its adaptability and extendibility.

After the comparison based on data set 2, data set 3 is further experimented with to examine the performance of the proposed method. The cathode material of data set 3 is LFP, which is different from the first two types of batteries. Here, only one cell is chosen for validation, and the voltage range of 2.5–3.5 V is selected as the health feature. The prediction results are sketched in Fig. 8, from which we can find that the general degradation trend looks like a polynomial curve, and hence, an efficient polynomial fitting method may gain preferable estimation with the input of cycle number. However, from the zoomed-in figure in Fig. 8(a), local partial capacity recovery emerges, making it difficult to accurately estimate the SOH by the simple polynomial fitting method. On the other hand, since the battery in data set 3 is fully charged and discharged during the experiment, the SOH value with respect to cycle number is close to a polynomial curve. Nevertheless, it is difficult to encounter full charge and discharge operations all the time in practice. Once partial charge and discharge behaviors occur, the number of cycles will increase, while the SOH degradation rate will be slower, and in this case, the polynomial curve fitting method cannot work well. In addition, as the research target of this article is to investigate a general efficient SOH prediction algorithm that can adapt to different types of batteries with the reduction of training data size, it is obvious that the polynomial fitting method and the related settings based on data set 3 are not applicable for data sets 1 and 2. Instead, the LSTM network and TL with a fine-tuning strategy based on data set 1 can be transplanted to predict the SOH of data set 3 with preferable and robust estimation. In addition, the single-LSTM algorithm is continuously exploited to compare the size of training data.

In the two networks, except for the training data, the other parameters remain the same. Through repetitive optimization, the first 80% (1076 cycles) and 60% (611 cycles) of aging data are, respectively, applied for training the single-LSTM and the present algorithm, and all the data are exerted to assess the prediction performance. As can be found, the prediction results do not show the distinct difference, and both methods can provide precise predictions. Concretely, the RMSE and AAE by the LSTM are 0.27% and 0.16%, and those by the proposed method remain close and are 0.36% and 0.24%. Nonetheless, the MAE by the proposed method is 0.09%, much lower than that by the LSTM; moreover, the data set size for training the proposed method is 20% less than that for training the single-LSTM algorithm. In other words, with the same estimation results, 465 cycle tests can be saved by the proposed method in this study, and most of the estimation error can still be controlled within 2.8%, which is only 0.92% higher than that of the base model. Thus, it can be summarized that the proposed algorithm can be transplanted from NCM batteries to LCO and LFP batteries, manifesting its superior extendable capability. The only task that needs to be conducted is to adjust the parameters of the fully connected layer and the output layer. In this manner, the feasibility of the proposed algorithm when applied to different types of batteries is verified.

### D. Performance Evaluation on Battery Pack

Four cells grouped into a small pack by the serial connection are experimented with the same environmental conditions.
These four cells are with the same property in data set 1 and are referred to as cells 1-1, 1-2, 1-3, and 1-4, respectively. The acquired data are utilized to identify the SOH by the proposed pack estimation algorithm. In this study, the estimation results of cell 1-1 are hired to construct the CMM. Note that the validation in Part A is based on the experiment of cell 1-1, and hence, the estimation results of cell 1-1 are not repeated here.

The estimation curves of CDM for cells 1-2, 1-3, and 1-4 are depicted in Fig. 9(a), (d), and (g). It can be clearly found the evolution trend of CDM for each cell shows different trends. In the first 200 cycles, the differences are not obvious, and the maximum SOH difference is less than 2%. However, as the battery ages, the difference of CDM becomes more and more obvious; when the cycle reaches 440 times, the maximum difference exceeds 3%.

The prediction results of cell SOH are sketched in Fig. 9(b), (e), and (h), and their corresponding errors are plotted in Fig. 9(c), (f), and (i) and summarized in Table III, where “LSTM” means the SOH obtained by the single-LSTM algorithm. As can be seen, the prediction curves of the proposed method are closer to the reference profiles than those by the single LSTM. The prediction error of the proposed method for all the cells is located closer to the horizontal axis. Also, the result of MAE is better than that of the single-LSTM method, as portrayed in Fig. 9(c), (f), and (i). Moreover, due to the long-term memory characteristics of single LSTM, the initial state of LSTM fluctuates obviously, raising larger SOH estimation error. By contrast, due to the LSTM-TL method and CDM, the SOH prediction results by the proposed method do not show a similar phenomenon.

The diagnosis results of battery pack SOH for single LSTM and the proposed method are shown in Fig. 10, and
the statistical error is listed in Table III. As can be seen, the prediction value of the single LSTM is mostly below the reference value, while the estimation value of the proposed method oscillates around the reference. It is obvious that the proposed method enables a more reasonable variation track of pack SOH compared to the single-LSTM method. According to Table III, the evaluation indexes reveal that the proposed method performs more accurately than single LSTM. To be specific, the AAE, MAE, and RMSE of the proposed method are less than two-thirds, one-third, and three-fifths of that by the single-LSTM method.

To sum up, the proposed LSTM integrating the TL algorithm and CDM can achieve satisfactory pack SOH estimation. As such, the abuse of battery packs can be effectively avoided by reliably diagnosing the pack SOH, and safe operation of batteries can be guaranteed. Moreover, the prediction procedure of the proposed scheme is executed during the charging phase, indicating its easy application potential in practice. In addition, in the proposed pack SOH prediction scheme, the LSTM model is employed as the CDM, and the minimum estimation value of CDM is identified to determine the pack SOH. When facing the SOH of a battery pack with multiple cells (for example, tens to thousands), first, the CDM value of each cell needs to be compared, and the minimum CDM will be selected. Then, the estimation method will be exploited to predict the entire battery pack SOH. From this point of view, applying the proposed scheme to the pack with multiple cells increases only the amount of CDM comparison, and thus, the estimation performance of pack SOH by the proposed algorithm will be still the same, as addressed in this study.

VI. Conclusion

The knowledge of battery pack SOH is of vital importance for safe operation. Machine learning-based methods require a large amount of training data to portray the hidden nonlinear relationship of lithium-ion batteries to predict SOH. In this study, a flexible LSTM-based battery pack SOH prediction scheme is advanced with the combination of the TL strategy. The relative charging time for a preset voltage range is extracted as the health feature to represent the degradation state of lithium-ion battery cells. Then, an improved LSTM network integrating TL is proposed to establish the CMM and fully substantiated by comparing it with the commonly used SOH prediction algorithms. The validation results based on the proposed model manifest that the cell SOH error is less than 3% with 70% training data. Moreover, the flexibility of the proposed TL algorithm is validated on different voltage ranges and different types of batteries, and the training data can be reduced by 20%–40% without discrediting the estimation performance, thereby justifying the remarkable contribution of TL in mitigating the training data amount. Based on the developed CDM, the SOH inconsistency existing in battery cells can be effectively taken into account, and the pack SOH is properly determined by the CMM and the minimum value of CDM. According to the proposed cell and pack SOH prediction methods, the pack SOH can reasonably follow the variation trend of the minimum SOH of cells and, therefore, enables the avoidance of improper abuse operations more efficiently.

Although a smaller power battery pack is employed as the validation object in this study, the proposed scheme is also feasible to predict the SOH of larger battery packs, which will be our validation focus in the next step. In the future, the aging mechanism and SOH diagnosis of lithium-ion batteries from the perspective of electrochemical reactions and the long-term prediction of battery SOH is imperative to be explored for the promotion of estimation precision. In addition, the impact of temperature on performance variation and SOH prediction of lithium-ion batteries also need to be further investigated.

REFERENCES

[1] N. Guo, X. Zhang, Y. Zou, B. Lenzo, and T. Zhang, “A computationally efficient path-following control strategy of autonomous electric vehicles with yaw motion stabilization,” IEEE Trans. Transport. Electrific., vol. 6, no. 2, pp. 728–739, Jun. 2020, doi: 10.1109/TTE.2020.2993862.
[2] X. Hu, C. Zou, X. Tang, T. Liu, and L. Hu, “Cost-optimal energy management of hybrid electric vehicles using fuel cell/battery health-aware predictive control,” IEEE Trans. Power Electron., vol. 35, no. 1, pp. 382–392, Jan. 2020, doi: 10.1109/TPEL.2019.2915675.
[3] X. Shu, G. Li, J. Shen, W. Yan, Z. Chen, and Y. Liu, “An adaptive fusion estimation algorithm for state of charge of lithium-ion batteries considering wide operating temperature and degradation,” J. Power Sources, vol. 462, Jun. 2020, Art. no. 228132.
[4] X. Ding et al., “A novel active equalization topology for series-connected lithium-ion battery packs,” IEEE Trans. Ind. Appl., vol. 56, no. 6, pp. 6892–6903, Nov. 2020.
[5] Z. Deng, X. Hu, X. Lin, Y. Che, L. Xu, and W. Guo, “Data-driven state of charge estimation for lithium-ion battery packs based on Gaussian process regression,” Energy, vol. 205, Aug. 2020, Art. no. 118000, doi: 10.1016/j.energy.2020.118000.
[6] X. Hu, L. Xu, X. Lin, and M. Pecht, “Battery lifetime prognostics,” Joule, vol. 4, no. 2, pp. 310–346, Feb. 2020.
[7] X. Shu, G. Li, J. Shen, Z. Lei, Z. Chen, and Y. Liu, “A uniform estimation framework for state of health of lithium-ion batteries considering feature extraction and parameters optimization,” Energy, vol. 204, Aug. 2020, Art. no. 117957, doi: 10.1016/j.energy.2020.117957.
[8] Z. Deng, X. Hu, X. Lin, L. Xu, Y. Che, and L. Hu, “General discharge voltage information enabled health evaluation for lithium-ion batteries,” IEEE/ASME Trans. Mechatronics, early access, Nov. 24, 2020, doi: 10.1109/TMECH.2020.3040010.
[9] X. Shu, G. Li, Y. Zhang, J. Shen, Z. Chen, and Y. Liu, “Online diagnosis of state of health for lithium-ion batteries based on short-term charging profiles,” J. Power Sources, vol. 471, Sep. 2020, Art. no. 228478.

[10] Y. Tan and G. Zhao, “Transfer learning with long short-term memory network for state-of-health prediction of lithium-ion batteries,” IEEE Trans. Ind. Electron., vol. 67, no. 10, pp. 8723–8731, Oct. 2020, doi: 10.1109/TIE.2019.2946653.

[11] D. Xiao et al., “Reduced-coupling coestimation of SOC and SOH for lithium-ion batteries based on convex optimization,” IEEE Trans. Power Electron., vol. 35, no. 11, pp. 12332–12346, Nov. 2020, doi: 10.1109/TPEL.2020.2984248.

[12] B. Duan, Q. Zhang, F. Geng, and C. Zhang, “Remaining useful life prediction of lithium-ion battery based on extended Kalman particle filter,” Int. J. Energy Res., vol. 44, no. 3, pp. 1724–1734, 2020.

[13] X. Tang, Y. Wang, C. Zou, K. Yao, X. Xia, and F. Gao, “A novel framework for lithium-ion battery modeling considering uncertainties of temperature and aging,” Energy Convers. Manage., vol. 180, pp. 162–170, Jan. 2019.

[14] C. Chen, R. Xiong, and W. Shen, “A lithium-ion battery-in-the-loop approach to test and validate multiscale dual h infinity filters for state-of-charge and capacity estimation,” IEEE Trans. Power Electron., vol. 33, no. 1, pp. 332–342, Jan. 2018.

[15] J. He, Z. Wei, X. Bian, and F. Yan, “State-of-health estimation of lithium-ion batteries using incremental capacity analysis based on voltage–capacity model,” IEEE Trans. Transport. Electrific., vol. 6, no. 2, pp. 417–426, Jun. 2020, doi: 10.1109/TTE.2020.2994543.

[16] X. Tang, K. Liu, X. Wang, F. Gao, J. Macro, and W. D. Widanage, “Model migration neural network for predicting battery aging trajectories,” IEEE Trans. Transport. Electrific., vol. 6, no. 2, pp. 363–374, Jun. 2020, doi: 10.1109/TTE.2020.2979547.

[17] J. Tian, R. Xiong, and W. Shen, “State-of-health estimation based on differential temperature for lithium-ion batteries,” IEEE Trans. Power Electron., vol. 35, no. 10, pp. 10363–10373, Oct. 2020, doi: 10.1109/TPEL.2020.2978300.

[18] X. Li, Z. Wang, and J. Yan, “Prognostic health condition for lithium battery using the partial incremental capacity and Gaussian process regression,” J. Power Sources, vol. 421, pp. 56–57, Mar. 2019, doi: 10.1016/j.jpowsour.2019.03.008.

[19] Y. Li et al., “Random forest regression for online capacity estimation of lithium-ion batteries,” Appl. Energy, vol. 232, pp. 210–219, Dec. 2018.

[20] C. She, Z. Wang, F. Sun, P. Liu, and L. Zhang, “Battery aging assessment for real-world electric buses based on incremental capacity analysis and radial basis function neural network,” IEEE Trans. Ind. Informat., vol. 16, no. 5, pp. 3345–3354, May 2020.

[21] P. Guo, Z. Cheng, and L. Yang, “A data-driven remaining capacity estimation approach for lithium-ion batteries based on charging health feature extraction,” J. Power Sources, vol. 412, pp. 442–450, Feb. 2019.

[22] Y. Zhou et al., “Hybrid genetic algorithm method for efficient and robust evaluation of remaining useful life of supercapacitors,” Appl. Energy, vol. 262, Feb. 2020, Art. no. 114169.

[23] Y. Wu, Q. Xue, J. Shen, Z. Lei, Z. Chen, and Y. Liu, “State of health estimation for lithium-ion batteries based on healthy features and long short-term memory,” IEEE Access, vol. 8, pp. 28533–28547, 2020.

[24] R. R. Richardson, C. R. Birkl, M. A. Osborne, and D. A. Howey, “Gaussian process regression for in situ capacity estimation of lithium-ion batteries,” IEEE Trans. Ind. Informat., vol. 15, no. 1, pp. 127–137, Jan. 2019.

[25] X. Hu, Y. Che, X. Lin, and Z. Deng, “Health prognosis for electric vehicle battery packs: A data-driven approach,” IEEE/ASME Trans. Mechatronics, vol. 25, no. 6, pp. 2622–2632, Dec. 2020, doi: 10.1109/TMECH.2020.2986364.

[26] Y. Zheng et al., “Cell state-of-charge inconsistency estimation for LiFePO 4 battery pack in hybrid electric vehicles using mean-difference model,” Appl. Energy, vol. 111, pp. 571–580, Nov. 2013.

[27] Y. Tian, R. Lai, X. Li, X. Jiang, and J. Tian, “A combined method for state-of-charge estimation for lithium-ion batteries using a long short-term memory network and an adaptive cubature Kalman filter,” Appl. Energy, vol. 265, May 2020, Art. no. 114789.

[28] J. Hong, Z. Wang, and Y. Yao, “Fault prognosis of battery system based on accurate voltage abnormality prognosis using long short-term memory neural networks,” Appl. Energy, vol. 251, Oct. 2019, Art. no. 113381, doi: 10.1016/j.apenergy.2019.113381.

[29] X. Shu, G. Li, Y. Zhang, S. Shen, Z. Chen, and Y. Liu, “Stage of charge estimation of lithium-ion battery pack based on improved cubature Kalman filter with long short-term memory model,” IEEE Trans. Transport. Electrific., early access, Dec. 1, 2020, doi: 10.1109/TTE.2020.3041757.

[30] Y. Zhang, R. Xiong, H. He, and M. G. Pecht, “Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries,” IEEE Trans. Veh. Technol., vol. 67, no. 7, pp. 5695–5705, Jul. 2018.

[31] Y. Guo, H. Shi, A. Kumar, K. Grauman, T. Rosing, and R. Feris, “SpotTune: Transfer learning through adaptive fine-tuning,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 4805–4814.

[32] M. Long, Y. Cao, J. Wang, and M. Jordan, “Learning transferable features with deep adaptation networks,” in Proc. Int. Conf. Mach. Learn. (ICML), 2015, pp. 97–105.

[33] X. Xing, Y. W. M. Ma, K.–L. Tsui, and M. Pecht, “An ensemble model for predicting the remaining useful performance of lithium-ion batteries,” Microelectron. Rel., vol. 53, no. 6, pp. 811–820, Jun. 2013.

[34] K. A. Severson et al., “Data-driven prediction of battery cycle life before capacity degradation,” Nature Energy, vol. 4, no. 5, pp. 383–391, May 2019.

[35] W. He, N. Williard, M. Osterman, and M. Pecht, “Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method,” J. Power Sources, vol. 196, no. 23, pp. 10314–10321, Dec. 2011.

[36] X. Hu, Y. Che, X. Lin, and S. Onori, “Battery health prediction using fusion-based feature selection and machine learning,” IEEE Trans. Transp. Electrific., early access, Aug. 17, 2020, doi: 10.1109/TTE.2020.3017090.

Xing Shu (Graduate Student Member, IEEE) received the M.S. degree from the Kunming University of Science and Technology, Kunming, China, in 2018, where he is currently pursuing the Ph.D. degree in transportation engineering. His research interests include battery management systems and battery status estimation for electric vehicles.

Jiangwei Shen received the B.S. degree in traffic engineering and the M.S. degree in power machine and engineering from the Kunming University of Science and Technology, Kunming, China, in 2008 and 2011, respectively, where he is currently pursuing the Ph.D. degree in automobile engineering. He is currently an Experimentalist with the Kunming University of Science and Technology. His research interests include battery management systems and energy management of hybrid electric vehicles.

Guang Li (Senior Member, IEEE) received the Ph.D. degree in electrical and electronics engineering, specialized in control systems, from The University of Manchester, Manchester, U.K., in 2007. He is currently a Reader in dynamics modeling and control with the Queen Mary University of London, London, U.K. His research interests include constrained optimal control, model predictive control, adaptive robust control, and control applications, including renewable energies and energy storage.
Yuanjian Zhang (Member, IEEE) received the M.S. degree in automotive engineering from Coventry University, Coventry, U.K., in 2013, and the Ph.D. degree in automotive engineering from Jilin University, Changchun, China, in 2018.

In 2018, he joined the University of Surrey, Guildford, U.K., as a Research Fellow in advanced vehicle control. He currently works at the Sir William Wright Technology Centre, Queen’s University Belfast, Belfast, U.K. His current research interests include advanced control on electric vehicle powertrains, vehicle-environment-driver cooperative control, vehicle dynamic control, and intelligent control for driving-assist systems.

Zheng Chen (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering and the Ph.D. degree in control science engineering from Northwestern Polytechnical University, Xi’an, China, in 2004, 2007, and 2012, respectively.

He was a Post-Doctoral Fellow and a Research Scholar with the University of Michigan at Dearborn, Dearborn, MI, USA, from 2008 to 2014. He is currently a Professor with the Faculty of Transportation Engineering, Kunming University of Science and Technology, Kunming, Yunnan, China, and also a Marie-Curie Research Fellow with the School of Engineering and Materials Science, Queen Mary University of London, London, U.K. He has conducted over 30 projects. He has published over 130 peer-reviewed journal articles and conference proceedings. His research interests include battery management systems, battery status estimation, and energy management of hybrid electric vehicles.

Dr. Chen is also a fellow of the Institution of Engineering and Technology. He was a recipient of the Yunnan Oversea High Talent Project, China. He got the second place in the IEEE VTS Motor Vehicle Challenge in 2017 and 2018.

Yonggang Liu (Senior Member, IEEE) received the B.S. and Ph.D. degrees in automotive engineering from Chongqing University, Chongqing, China, in 2004 and 2010, respectively.

He was a Joint Ph.D. Student with the University of Michigan at Dearborn, Dearborn, MI, USA, from 2007 to 2009. He is currently a Professor and a Doctoral Supervisor with the College of Mechanical and Vehicle Engineering, Chongqing University. He has led more than 20 research projects. He has published more than 70 research articles. He holds ten patents. His research interests mainly include optimization and control of intelligent electric vehicle power systems and integrated control of vehicle automatic transmissions.

Dr. Liu is also a Committeeman of the Technical Committee on Vehicle Control and Intelligence of the Chinese Association of Automation.