Overview of Machine Learning-Enabled Battery State Estimation Methods

Yingjian Zhuge[1], Hengzhao Yang[1], and Haoyu Wang[1,2]
1 School of Information Science and Technology, ShanghaiTech University, Shanghai, China
2 Shanghai Engineering Research Center of Energy Efficient and Custom AI IC, Shanghai, China
Email: wanghy@shanghaitech.edu.cn

***Abstract*—To ensure safe usage and robust performance of energy storage batteries, accurate state-of-charge (SOC) and state-of-health (SOH) estimations are required. Due to recent breakthroughs in machine learning and artificial intelligence methods, data-driven methods have attracted increased attention. This paper reports state-of-the-art research progress in machine learning-enabled methods for SOC and SOH estimations. Comprehensive comparisons are made in terms of the dataset, estimation accuracy, and battery type to provide a clear picture for SOC and SOH estimation. Moreover, the challenges and research opportunities on future SOC and SOH estimation are disclosed.**

***Index Terms*—Machine learning, deep learning, state of charge (SOC), state of health (SOH).**

I. Introduction

Carbon neutrality requires deep penetration of renewable energy. While the utilization of renewable energy inevitably requires energy storage systems. Lithium-ion battery is one of the most important energy storage components and its performance is always monitored by the battery management system (BMS). The basic structure of BMS is shown in Fig. 1. As indicated, state-of-Charge (SOC) and state-of-health (SOH) are two important parameters reflecting battery performance. Accurate SOC estimation can avoid overcharging and over-discharging to prolong the battery life, while accurate SOH estimation can ensure safe, reliable, and efficient operations of the battery.

Over the past decades, a large number of SOC estimation methods have been investigated. Traditional methods can be divided into the following categories: look-up table-based methods [1], [2], ampere-hour integral methods [3], [4], model-based methods [5]–[8] and data-driven methods [9], [10]. For look-up table-based methods, it suffers from some distinct drawbacks, e.g., it is hard to measure precise OCV in real time, and the measurement of AC impedance can only be conducted offline. So these methods are only suitable for laboratory environment. Ampere-hour integral method is extensively applied to SOC estimation due to its advantages of easy implementation and small computational burden. However, the major drawback is that the initial SOC is hardly accurately acquired and the measurement comes with errors from current sensor, temperature drift and noise. Besides, the maximum available capacity needs to be recalibrated with the variation of the aging levels. To overcome the weakness of the above mentioned estimation methods, various model-based methods have been proposed. The main challenges of current model-based methods lie in identifying all parameters, and the parameters are sensitive to material characteristics and aging levels. An appropriate balance between model fidelity and computational complexity needs to be achieved [11]. Data-driven methods are flexible and battery model-free. However, the accuracy of SOC estimation relies on the quality and quantity of the training dataset, and it is time-consuming.

This work was supported in part by the National Natural Science Foundation of China under Grant 52077140, and in part by the Shanghai Rising Star Program under Grant 20QA1406700.

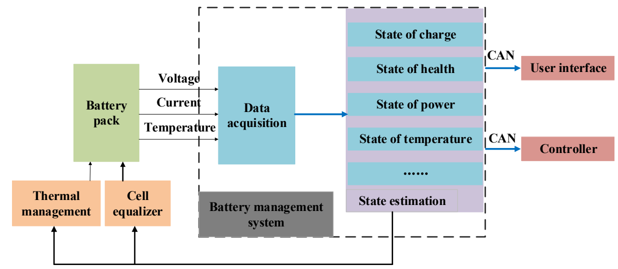

Fig. 1. Block diagram of a typical battery management system.

Similarly, traditional battery SOH estimation methods can also be roughly divided into several categories: direct measurement methods [12]–[14], indirect analysis methods [15]–[17], model-based methods [18], [19] and data-driven methods [20], [21]. However, the first two methods are hard to achieve an accurate measurement result and are limited in real-time estimation. Model-based methods are the commonly used online techniques with high precision for SOH estimation. But it is difficult to establish an accurate battery model due to the complex operations and parameter identification. Data-driven methods avoid the drawbacks of model-based methods, while the accuracy is highly dependent on data.

To date, there have been a few review papers on SOC and SOH estimation. A review of SOC estimation is published in [22]. However, it focuses on the model-based methods, while the data-driven methods are barely covered. In [23], the SOC estimation methods are classified into different categories, and the advantages with disadvantages are discussed. But it does not deliver a detailed description of machine learning (ML) methods. The performance of the different online SOC estimation (such as, extended Kalman filter (EKF), unscented

Kalman filter (UKF), central difference Kalman filter (CDKF), etc.) is discussed, but it concentrates only on model-based methods [24]. In terms of SOH estimation methods, they can be divided into four categories: model-based methods, data-driven methods, hybrid methods, and other methods. However, ML methods are not analyzed in depth [25]. To be specific, non-probabilistic ML-based estimation techniques are reviewed, and the principle of the ML algorithms is introduced. Each ML algorithm is compared to show its advantages and disadvantages [26]. To summarize, these reviews only focus on a single type of estimation, i.e., either SOC or SOH. In [27], battery modeling, state estimation and battery charging are briefly reviewed. However, there is a lack of detailed analysis. In [28], both SOC and SOH estimation methods are reviewed, while the ML methods are only summarized, not detailed. In [29], battery models are reviewed and ML techniques for state prediction are showcased, while the review of ML methods is not comprehensive.

With the ever-increasing computing power provided by GPUs and a significant amount of data collected during usage, the data-driven method has gained increased attention. Different ML methods have been proposed and applied to improve BMS performance, such as linear regression (LR), feedforward neural network (FNN), support vector machine (SVM) and recurrent neural network (RNN). Thus, those review papers are not up-to-date. Additionally, understanding ML algorithms is a prerequisite for application. To bridge the research gap, this paper reviews the state-of-the-art ML algorithms for SOC or SOH estimation. Compared with previous work, the main contributions of this review work are as follows:

- Basic principle of ML algorithms is introduced. A comprehensive comparison is made to pave the way for future algorithms improvement.
- SOC and SOH estimation methods in recent years are prudently classified and discussed based on their algorithms and applications. These methods are critically compared and summarized in the aspects of battery types, input and output characteristics, performance, dataset and operating conditions.
- The potential challenges are studied and summarized for the advancement of ML-enabled battery state estimation methods.

The rest of the paper is organized as follows. Section II introduces the basic principle of different ML algorithms as well as their applications in SOC estimation, and the performance of different ML-based methods is compared. Section III presents ML algorithms applied in SOH estimation. Then the comparisons of different ML algorithms and the challenges for ML-based state estimation are shown in Section IV. Finally, the conclusion is drawn in Section V.

II. Battery SOC estimation

SOC is defined as the ratio of the residual capacity of battery to its maximum capacity and is given by

$$SOC(t) = \frac{C_{remain}}{C_{rate}} \times 100\% \tag{1}$$

Conventionally, the value of SOC ranges from 0% to 100%, which represents the fully discharged and fully charged conditions of battery. In essence, accurate SOC estimation is the foundation for other state estimations and it is extremely non-trivial. Thus, a considerable effort has been done to explore the advanced methods for SOC estimation.

A. Linear Regression

Generally, SOC estimation methods have more than one input. Thus, multiple linear regression (MLR) is commonly used to analyze the relationship between the dependent variable and the desired output. The expression is given by

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + \epsilon \tag{2}$$

where Y is the output, x_i is the input variables, β_0 represents constant, the other parameters β_i are partial regression coefficent, and ϵ means the random error.

MLR is used to build model in [30]. In comparison, the MLR approach produces a better prediction than the SVM and ANN. To improve the estimation performance of MLR, spline interpolation (SPL-MLR) and the Generalized Linear Model (GLM) are introduced in [31]. Different optimization techniques used for parameter optimization have been studied.

B. Feedforward Neural Network

Feedforward neural network is one of the simplest neural networks. Each neuron is arranged in layers, and each neuron is only connected to the neuron of the previous layer to receive the output of the previous layer and output to the next layer. The structure of the FNN is shown in Fig. 2.

In [32], an improved SOC estimation method based on a back-propagation neural network (BPNN) is proposed. To improve the learning ability and accuracy of BPNN, wavelet neural network (WNN) is introduced. The network has advantages of the wavelet transform in denoising, background deduction and recovery of characteristic information. Based on this, in [33], single-layer and two-hidden-layer WNNs-based Levenberg-Marquardt algorithms are employed in SOC estimation. Compared with BPNN and EKF, the proposed method is better than the others in terms of mean absolute error (MAE).

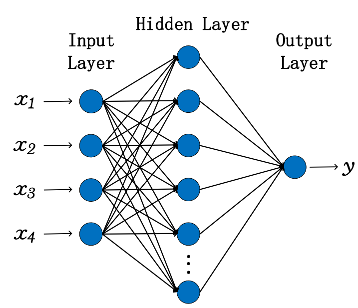

Fig. 2. Architecture of feedforward neural network.

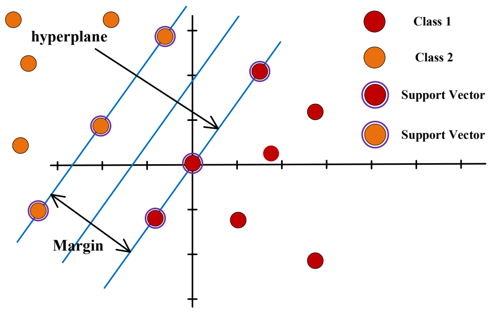

Fig. 3. Hyperplane to separate samples from two classes [38].

C. Radial Basis Function Neural Network

A radial basis function neural network (RBFNN) is a single hidden layer feedforward neural network based on function approximation. RBF is usually defined as monotone function of the Euclidean distance between any point and the center point, which is used as the activation function of the neurons. Compared with the traditional BPNN, RBFNN has SEVERAL advantagess, such as simple structure, fast convergence speed, and strong classification capacity.

in [34], RBFNN is used for SOC estimation in battery degradation process. A Lithium Ion battery is chosen as a device under test. Aging cycle tests and driving operation tests under different temperatures are conducted. Compared with the conventional NN, the proposed method exhibits better accuracy and robustness. The MAE of estimation under different temperatures and aging cycles is all below 5%.

D. Extreme Learning Machine

An extreme learning machine (ELM) can be seen as a special FNN. However, compared with FNN, ELM does not need a backpropagation algorithm to adjust the weight but set the weight through Moore-Penrose generalized inverse [35].

The performance of ELM is highly dependent on training accuracy and the number of neurons in a hidden layer. Hence, in [36], a gravitational search algorithm (GSA) is applied to improve the ELM computational intelligence by searching for the optimal value of hidden layer neurons. Higher accuracy and less computational cost are achieved. Under different operating conditions, ELM-GSA has a fairly low root-mean-square error(RMSE) of less than 2%.

E. Support Vector Machine

Support vector machine is a binary classification model. Its basic principle is to find the optimal separation hyperplane in the feature space to maximize the positive and negative sample interval on the training set [37].

In [39], SVM is used to build a SOC predictive model for battery cell. Using this model, the maximum error (MAX) is constrained below 6% and the RMSE is 0.71%. In [40], the least squares support vector machine (LS-SVM) is implemented to predict the available capacity online. The optimal parameters are obtained by applying the genetic algorithm (GA). SOC estimation is based on the dual AEKF, which can adaptively update the noise covariance matrixes. The numeric results of the estimated SOC based on the joint method show that RMSE is 2.84%.

F. Gaussian Process Regression

Gaussian process regression (GPR) is a non-parametric Bayesian approach to regression problems. It can capture a wide variety of relations between inputs and outputs by utilizing a theoretically infinite number of parameters and letting the data determine the level of complexity through the means of Bayesian inference [41].

In [42], an autoregressive model is constructed to further improve estimation accuracy and confidence. The test platform is chosen as a battery pack with four individual cells. For the proposed autoregessive model, the initial SOC needs to be known, otherwise the error would increase to a certain extent. The estimation errors of the autoregressive GPR model are lower than 3% for different dynamic cycles and aging states.

G. Deep Neural Network

Deep neural network (DNN) is a neural network with much more hidden layers. Thus, it has better expression ability and generalization ability. In [43], DNN with four hidden layers is introduced. It has a higher computational speed than combined model with EKF, and offers competitive estimation with MAE below 1%.

For DNN, two latent problems limit flexible and scalable applications. First, the data should follow similar distributions. Second, large-scale training data is required to ensure the success of the DNN. However, practically, it is difficult to collect sufficient data. To address this issue, many improved deep-learning methods have been proposed.

H. Convolutional Neural Network

Different from fully connected DNN, a convolutional neural network (CNN) consists of convolutional layers, pooling layers, and fully connected (FC) layers. The convolutional layer contains a set of filters whose parameters need to be learned. Each filter is convolved with the input volume to compute an activation map made of neurons. The pooling layers are designed to downsample along the spatial dimensionality of the given input, which further reduces the number of parameters within that activation. The FC layers perform the same duties found in standard DNN and are used for classification. Its basic architecture is shown in Fig. 4.

In [44], a CNN-based battery capacity estimation method is proposed. A novel data segmentation and time series-to-image

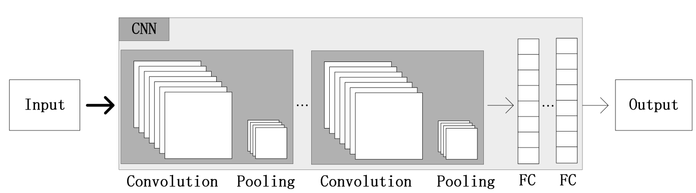

Fig. 4. A simple CNN architecture.

transformation method are introduced to realize capacity estimation with limited data. Compared with traditional DNN, it can achieve better accuracy with much fewer parameters.

I. Recurrent Neural Network

Recurrent neural network is a kind of recursive neural network that takes sequence data as input and recurses in sequence evolution direction with all nodes chaining together. However, training traditional RNN has been proven to be problematic because the backpropagated gradients either grow or shrink at each time step. This means over many time steps they typically explode or vanish [47]. To deal with this problem, many optimization schemes are proposed.

Long short-term memory (LSTM) and gated-recurrent unit are two typical novel RNNs. The corresponding recurrent unit of LSTM contains three gated units: input gate, output gate and forget gate which set up a self-loop, while the unit of GRU contains only two gates: update gate and reset gate.

In [45], a single LSTM-RNN is proposed for SOC estimation of a Panasonic 18650 battery cell. It investigates the effect of network depth and the number of training data on performance, which achieves 0.774% MAE at 25°C.

In [46], GRU is applied for battery SOC estimation. The author tests the estimation performance of GRU under varying temperatures and different battery materials. This method can also work in case the initial SOC is unknown. The proposed method provides an estimation performance with 2.15% RMSE and 1.65% MAE.

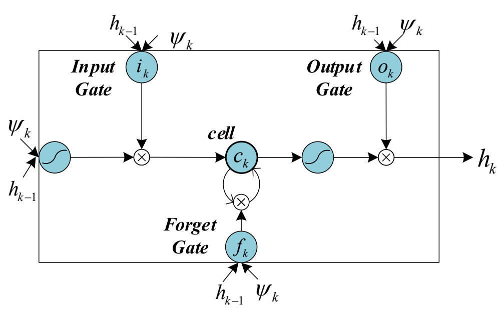

Fig. 5. LSTM cells [45].

J. Comparisons of SOC Estimation Methods

An extensive comparison in terms of data profile, inputs, and outputs of network, battery type for experiments, and test conditions is conducted. The comparison results are summarized in Table. I. Most of the methods utilize similar parameters such as inputs (current, voltage, and temperature), while some of them use derivatives and integrals of these parameters or other parameters for special applications. The datasets achieved by lab experiments are usually automotive dynamic profiles, while some of them use constant current (CC) or CC-CV charge/discharge tests and pulse tests. Multi-temperature validation has been carried out in the literature, but validation at negative temperatures is rarely reported.

III. Battery SOH estimation

Generally, battery aging often refers to degraded capacity and increased internal resistance, which is directly employed as deterioration parameters of battery electrical performance. SOH can be determined as the ratio of the current values to the initial values of parameters of a battery, such as battery capacity and internal resistance.

$$SOH = \frac{C_i}{C_0} \times 100\% \tag{3}$$

$$SOH = \frac{R_i}{R_0} \times 100\% \tag{4}$$

where C_i and R_i represent the capacity and internal resistance at the i_{th} cycle, C_0 and R_0 represent nominal capacity and internal resistance of the new battery without being used.

A. Linear Regression

In [48], a comprehensive dataset consisting of 124 commercial cells is generated, which is open access. LR is introduced to predict and classify cells by cycle life. This work highlights the promise of combining data generation with data-driven modeling to predict the behaviors of complex systems.

B. Feedforward Neural Network

In [49], a combination of a prior knowledge-based neural network (PKNN) and Markov chain model is established to estimate the Lithium-ion battery SOH. The prior knowledge is transformed into constraints to optimize the traditional NN to improve the generalization ability. And the output of PKNN is modified by the Markov correction to obtain a more accurate and reliable result. A feed-forward migration NN that merges the NN and the input-output slope and bias correction (SBC) model migration is proposed in [50]. By integrating the SBC migration model into the NN, the knee point dose not need to be covered by the online accumulated training dataset and the nonlinear fitting performance is significantly improved.

C. Extreme Learning Machine

To relieve the dependence of the estimation performance on the initial data, a metabolic ELM (MELM) framework is proposed to obtain accurate SOH estimation for batteries with unclear usage levels using a few historical data [51]. In [52], a hierarchical ensemble model of ELM is developed to build the underlying relationship between the extracted indicator and SOH. This method is tested under three different datasets and the influence of starting voltage is also analyzed. The experimental results exhibit better performance compared with other ML methods.

D. Support Vector Machine

In [53], the fixed size LS-SVM is employed to estimate the SOH with less computation intensity. The charging time variation within a fixed voltage range in each cycle is extracted as the health indicator to quantify capacity degradatio. In [54], SVR is used to build a state-space model to describe the degradation parameters. And an impedance aging-based degradation

TABLE I
COMPARISON OF SOC ESTIMATION METHODS.

Machine Learning Methods	Lowest error	Data Profiles	Input	Output	Battery	Multi-Temperature
MLR [31]	1.36%(MAE)@US06 1.66(MAE)%@BJDST	CALCE	V, I	SOC	INR 18650-20R (Li-NiMnCoO2/Graphite)	0°C, 25°C 45°C
BPNN [32]	0.82%(MAE)@FUDS 0.59%(MAE)@US06 0.55%(MAE)@BJDST	CALCE	I,V,T, $\int idt$,$\int vdt$ di, d^2v	SOC	18650 NMC	25°C
RBFNN [34]	2.4%(MAE)@UDDS 2.1(MAE)%@ECE	Lab experiments	V,I,C	SOC	Lithium-ion battery	10°C, 25°C, 40°C
ELM [36]	0.55%(MAE)@BJDST 1.67(MAE)%@US06	Lab experiments	V,I,T	SOC	18650 NMC	25°C, 40°C
SVM + AEKF [40]	2.84%(RMSE)@UDDS 2.76(RMSE)%@Hybrid	Lab experiments	T, R_{int}, R_p $V, \Delta V$	C	Li-ion FP3391146A	52°C,25°C 0°C,-12°C
GPR [42]	2.99%(RMSE)@DST 1.91(RMSE)%@FUDS 2.92(RMSE)%@DST	Lab experiments	V_{peak}, I, V_{cell}, T	SOC	$LiNi_{0.8}Co_{0.1}Mn_{0.1}O_2$ -graphite pack	0°C,25°C
DNN [43]	1.06%(MAE)@HWFET 1.59%(MAE)@US06	Lab experiments	V_{avg}, I, V, I_{avg}	SOC	Panasonic 18650 PF	-20°C,-10°C, 0°C,10°C,25°C
CNN [44]	2.93%(RMSE)@Oxford 0.95%(RMSE)@cells	Oxford 124 commercial cells	V, I ,T	C	Kokam pouch cells APR18650M1A	30°C,40°C
LSTM-RNN [45]	0.688%(MAE)@CC charge 0.774%(MAE)@Combined	Lab experiments	V, I, T	SOC	Panasonic NCR18650PF	0°C, 10°C, 25°C
GRU-RNN [46]	NMC:0.77%(MAE)@DST LFP:1.72(MAE)%@DST	Lab experiments	V, I, T	SOC	BAK B18650CD NMC A123 18650 LFP	0°C, 10°C,20°C, 30°C,40°C,50°C, 27°C (RT)

parameter model is established to improve the robustness of the SOH estimation. Based on this, a SOH prognosis and remaining useful life (RUL) prediction framework is proposed, which can provide accurate estimation results.

E. Gaussian Process Regression

In [55], the author first utilize the GPR with automatic relevance determination (ARD) kernel to perform battery calendar aging prognosis. In [56], a novel hybrid method by fusion of partial incremental capacity and GPR is proposed and dual GPR models are employed to forecast battery health conditions. The Pearson correlation analysis method is used to screen the feature variables for high-quality input, which is used for GPR-based model. Moreover, the long-term GPR-based models for battery RUL prediction are established using the data of SOH estimation.

F. Deep Neural Network

In [57], a five-layer network is employed to improve the estimation accuracy of battery pack when applied to actual driving cycle. The raw data used for training and testing is collected by Shanghai Electric Vehicle Public Data Collecting, Monitoring and Research Center (SHEVDC). The estimation results show high accuracy to the actual value and the maximum error is below 4.5%.

G. Recurrent Neural Network

Due to the sampling error and measure interference, data loss often occurs during acquisition. In [58], an ELM method is leveraged to forecast the entire temperature variation during the constant current charging process based on randomly discontinuous short-term charging data. Based on this, GRU is employed to predict SOH. The validation results highlight that the proposed method can accurately estimate SOH through incomplete charging data.

In [59], the authors first build an approximate battery SOH degradation model for real vehicle operation. Except for directly measurable battery parameters, some parameters concerning weather and driving behaviors, as well as the inconsistencies of cell voltages and probe temperatures, are simultaneously added to the inputs for collaborative training.

H. Comparisons of SOH Estimation Methods

Similar comparisons are made in Table II. Almost all of the proposed methods can estimate SOH accurately, showing their promising application. However, only a few works use the data of real-world EVs to train and test the model. As we know, data is one of the most important factors affecting the performance of machine learning methods. Most of the literature uses publicly available datasets (Oxford dataset [60], NASA dataset [61] and CALCE dataset [62]), but those datasets are only appropriate for preliminary studies.

IV. DISCUSSION AND CHALLENGES

A. Comparisons of ML Methods

Many machine learning algorithms have been applied to SOC and SOH estimation. Based on the advantages and

TABLE II
COMPARISON OF SOH ESTIMATION METHODS.

Machine Learning Methods	Lowest error	Data Profiles	SOH approach	Battery	Multi-Temperature
LR [48]	9.1%@commercial	124 commercial cells	Based on the cycle	APR18650M1A	30°C
PKNN [49]	0.3448%(MAE)@NASA	NASA dataset commercial battery dataset	Directly estimate SOH	18650 Li-ion IFP1865140	24°C
ELM [52]	2.04%(RMSE)@NASA 0.46%(RMSE)@Oxford 1.68%(RMSE)@CALCE	NASA dataset Oxford dataset CALCE dataset	Based on the current capacity	18650 Li-ion batteries Kokam pouch cells $LiCoO_2$ battery	24°C,40°C
SVM [54]	0.79%(RMSE)@Oxford 1.58%(RMSE)@NASA	Oxford dataset NASA dataset	Based on the impedance	Kokam pouch cells 18650 cells	24°C,40°C
GPR [56]	0.15%(MAE)@NASA	NASA dataset	Based on the capacity	Lithium-ion 18650	24°C
DNN [57]	0.45%(MAE)@SHEVDC	SHEVDC	Based on the current capacity	LiNMC oxide battery	0°C,45°C
LSTM [59]	0.232%(RMSE)@real-world	SMC-EV dataset	Based on the internal resistance	Li-ion battery	-20°C,0°C,10°C, 20°C,40°C,55°C

disadvantages of each algorithm, there is a need to analyze the application of the algorithm in specific scenarios.

When the amount of data is small, SVM is a candidate for modeling. Modifying the kernel function can improve the quality of fit for scarce data or high dimensional models and nonlinear models. However, the computational burden also increases, and SVM is sensitive to outliers.

For simple algorithms, such as FNN, SVM and ELM, ensemble learning (EL) is a general meta-approach to obtain better performance by combining several ML models, which are known as weak learners. The prediction result of EL is computed by combining the results from all the weak learners, and each uncorrelated model may solve one weakness of the other. Thus, the accuracy and generalization can be significantly improved.

Deep learning (DL) is a promising candidate for battery state estimation with the ever-increasing computational power and a significant amount of data. DL algorithms have obvious advantages when the data is of high quality. Specifically, CNN is suitable for multi-dimension data and RNN is suitable for time-series data. Though it needs more computational efforts than other algorithms, it can be mitigated with the development of cloud computing and big data.

B. Challenges

1) Battery Pack Estimation: Compared with battery cell estimation, battery pack estimation is more complex. Due to the influence of different manufacturing processes and the environment, there are differences in the dynamic characteristics among battery cells. When performing battery pack state estimation, many other factors should be accounted, Those factors include inconsistency in cell capacity, temperature, voltage and so on [63], [64]. Furthermore, the data from he battery pack is enormous. And as the 800V system is becoming mainstream, packs will be more and more complex. Thus, the application of machine learning algorithms in battery pack state estimation has a wide range of research prospects.

2) Combined Methods: To overcome the drawbacks of a single method, some work combines the single methods to obtain better estimation accuracy and less computational burden. Many combined methods have been proposed. For example, in [59], LSTM and MLR are combined to improve learning and memorizing ability. Data-driven methods and model-based methods are also combined to simplify the tedious procedure of tuning the parameters of the network, with the advantages of increased model freedom [65], [66].

3) Joint Estimation: Battery state estimations are coupled and they interact with each other. For example, the accuracy of SOC estimation is affected by SOH estimation, such as capacity and resistance. While these parameters used for SOH estimation are limited by the accuracy of SOC estimation. Thus, to do single state estimation while ignoring others may only achieve relatively accurate results. A few joint estimation methods have been proposed, such as joint estimation of SOC and SOH [67]–[69], joint estimation of SOC and state of power (SOP) [70], and joint estimation of SOC and state of temperature (SOT) [71], [72]. However, multi-scale state estimation requires great computational efforts, it limits the joint estimation methods with more than three battery states. Thus, improving the computation efficiency is one of the feasible methods for the co-estimation of more than three states.

4) Data: The dataset plays an important role in machine learning algorithms to train and validate. However, obtaining sufficient high-quality data may take a long time. Though more and more datasets have been generated, the quality and efficiency need further improvement. On the other hand, most recent literature does not use data obtained from real-world applications, but rather publicly available datasets. Thus, it is only suitable for preliminary studies. There is a need for more research in real-world applications, where the influencing factors, such as ambient temperature, weather conditions, and driving operations, should be considered.

V. Conclusion

This paper reviews the state-of-the-art machine learning-enabled methods for SOC and SOH estimation and makes a comprehensive comparison of different methods.

Existing machine learning-enabled methods still have still a long way from practical usage. On the one hand, these methods are generally developed from laboratory data, which causes a gap between lab and practice. The influencing factors in the real world are also important. Furthermore, the computational power also limits the practical application, due to a large amount of data, especially for the battery pack. The feasible estimation method must be a tradeoff of accuracy, computational effort and robustness.

Based on the research status of battery state estimation, this paper summarizes potential challenges, which can benefit future work and development, especially the applications in the real world.

References

[1] G. Dong, J. Wei, C. Zhang, and Z. Chen, "Online state of charge estimation and open circuit voltage hysteresis modeling of LiFePO4 battery using invariant imbedding method," *Appl. Energy*, vol. 162, pp. 163–171, Jan. 2016.

[2] H. He, X. Zhang, R. Xiong, Y. Xu, and H. Guo, "Online model-based estimation of state-of-charge and open-circuit voltage of lithium-ion batteries in electric vehicles," *Energy*, vol. 39, no. 1, pp. 310–318, Mar. 2012.

[3] Z. Lei, T. Liu, X. Sun, H. Xie, and Q. Sun, "Extended state observer assisted Coulomb counting method for battery state of charge estimation," *Int. J. Energy Res.*, vol. 45, no. 2, pp. 3157–3169, Feb. 2021.

[4] L. Zhao, M. Lin, and Y. Chen, "Least-squares based coulomb counting method and its application for state-of-charge (SOC) estimation in electric vehicles," *Int. J. Energy Res.*, vol. 40, no. 10, pp. 1389–1399, Aug. 2016.

[5] K. Yang, Y. Tang, and Z. Zhang, "Parameter Identification and State-of-Charge Estimation for Lithium-Ion Batteries Using Separated Time Scales and Extended Kalman Filter," *Energies*, vol. 14, no. 4, p. 1054, Feb. 2021.

[6] F. Guo, G. Hu, S. Xiang, P. Zhou, R. Hong, and N. Xiong, "A multi-scale parameter adaptive method for state of charge and parameter estimation of lithium-ion batteries using dual Kalman filters," *Energy*, vol. 178, pp. 79–88, Jul. 2019.

[7] C. Jiang, S. Wang, B. Wu, C. Fernandez, X. Xiong, and J. Coffie-Ken, "A state-of-charge estimation method of the power lithium-ion battery in complex conditions based on adaptive square root extended Kalman filter," *Energy*, vol. 219, p. 119603, Mar. 2021.

[8] Z. Deng, X. Hu, X. Lin, Y. Kim, and J. Li, "Sensitivity Analysis and Joint Estimation of Parameters and States for All-Solid-State Batteries," *IEEE Trans. Transp. Electrif.*, vol. 7, no. 3, pp. 1314–1323, Sept. 2021.

[9] P. Khumprom and N. Yodo, "A Data-Driven Predictive Prognostic Model for Lithium-ion Batteries based on a Deep Learning Algorithm," *Energies*, vol. 12, no. 4, p. 660, Feb. 2019.

[10] T. Zahid, K. Xu, W. Li, C. Li, and H. Li, "State of charge estimation for electric vehicle power battery using advanced machine learning algorithm under diversified drive cycles," *Energy*, vol. 162, pp. 871–882, Nov. 2018.

[11] Z. Wang, G. Feng, D. Zhen, F. Gu, and A. Ball, "A review on online state of charge and state of health estimation for lithium-ion batteries in electric vehicles," *Energy Rep.*, vol. 7, pp. 5141–5161, Nov. 2021.

[12] Y. H. Chiang, W. Y. Sean, and J. C. Ke, "Online estimation of internal resistance and open-circuit voltage of lithium-ion batteries in electric vehicles," *J. Power Sources*, vol. 196, no. 8, pp. 3921–3932, Apr. 2011.

[13] W. Waag, S. Käbitz, and D. U. Sauer, "Experimental investigation of the lithium-ion battery impedance characteristic at various conditions and aging states and its influence on the application," *Appl. Energy*, vol. 102, pp. 885–897, Feb. 2013.

[14] D. I. Stroe, M. Swierczynski, S. K. Kær, and R. Teodorescu, "Degradation Behavior of Lithium-Ion Batteries During Calendar Ageing - The Case of the Internal Resistance Increase," *IEEE Trans. Ind. Appl.*, vol. 54, no. 1, pp. 517–525, Jan. 2018.

[15] S. Qu, Y. Kang, P. Gu, C. Zhang, and B. Duan, "A Fast Online State of Health Estimation Method for Lithium-Ion Batteries Based on Incremental Capacity Analysis," *Energies*, vol. 12, no. 17, p. 3333, Aug. 2019.

[16] B. Jiang, H. Dai, and X. Wei, "Incremental capacity analysis based adaptive capacity estimation for lithium-ion battery considering charging condition," *Appl. Energy*, vol. 269, p. 115074, Jul. 2020.

[17] X. Hu, S. E. Li, Z. Jia, and B. Egardt, "Enhanced sample entropy-based health management of Li-ion battery for electrified vehicles," *Energy*, vol. 64, pp. 953–960, Jan. 2014.

[18] C. Burgos-Mellado, M. E. Orchard, M. Kazerani, R. Cárdenas, and D. Sáez, "Particle-filtering-based estimation of maximum available power state in Lithium-Ion batteries," *Appl. Energy*, vol. 161, pp. 349–363, Jan. 2016.

[19] Y. Hua, A. Cordoba-Arenas, N. Warner, and G. Rizzoni, "A multi time-scale state-of-charge and state-of-health estimation framework using nonlinear predictive filter for lithium-ion battery pack with passive balance control," *J. Power Sources*, vol. 280, pp. 293–312, Apr. 2015.

[20] A. A. Hussein, "Capacity Fade Estimation in Electric Vehicle Li-Ion Batteries Using Artificial Neural Networks," *IEEE Trans. Ind. Appl.*, vol. 51, no. 3, pp. 2321–2330, May. 2015.

[21] X. Hu, Y. Che, X. Lin, and S. Onori, "Battery Health Prediction Using Fusion-Based Feature Selection and Machine Learning," *IEEE Trans. Transp. Electrif.*, vol. 7, no. 2, pp. 382–398, Jun. 2021.

[22] R. Xiong, J. Cao, Q. Yu, H. He, and F. Sun, "Critical Review on the Battery State of Charge Estimation Methods for Electric Vehicles," *IEEE Access*, vol. 6, pp. 1832–1843, Dec. 2017.

[23] J. P. Rivera-Barrera, N. Muñoz-Galeano, and H. O. Sarmiento-Maldonado, "SoC Estimation for Lithium-ion Batteries: Review and Future Challenges," *Electronics*, vol. 6, no. 4, p. 102, Nov. 2017.

[24] J. Meng, M. Ricco, G. Luo, M. Swierczynski, D. I. Stroe, A. I. Stroe, and R. Teodorescu, "An Overview and Comparison of Online Implementable SOC Estimation Methods for Lithium-Ion Battery," *IEEE Trans. Ind. Appl.*, vol. 54, no. 2, pp. 1583–1591, Jan. 2018.

[25] H. Tian, P. Qin, K. Li, and Z. Zhao, "A review of the state of health for lithium-ion batteries: Research status and suggestions," *J. Clean. Prod.*, vol. 261, p. 120813, Jul. 2020.

[26] X. Sui, S. He, S. B. Vilsen, J. Meng, R. Teodorescu, and D. I. Stroe, "A review of non-probabilistic machine learning-based state of health estimation techniques for Lithium-ion battery," *Appl. Energy*, vol. 300, p. 117346, Oct. 2021.

[27] K. Liu, K. Li, Q. Peng, and C. Zhang, "A brief review on key technologies in the battery management system of electric vehicles," *Front. Mech. Eng.*, vol. 14, no. 1, pp. 47–64, Apr. 2018.

[28] J. Hou, T. Li, F. Zhou, D. Zhao, Y. Zhong, L. Yao, and L. Zeng, "A Review of Critical State Joint Estimation Methods of Lithium-Ion Batteries in Electric Vehicles," *World Electr. Veh. J.*, vol. 13, no. 9, p. 159, Aug. 2022.

[29] M.-F. Ng, J. Zhao, Q. Yan, G. J. Conduit, and Z. W. Seh, "Predicting the state of charge and health of batteries using data-driven machine learning," *Nat. Mach. Intell.*, vol. 2, no. 3, pp. 161–170, Mar. 2020.

[30] V. Chandran, C. K. Patil, A. Karthick, D. Ganeshaperumal, R. Rahim, and A. Ghosh, "State of Charge Estimation of Lithium-Ion Battery for Electric Vehicles Using Machine Learning Algorithms," *World Electr. Veh. J.*, vol. 12, no. 1, p. 38, Mar. 2021.

[31] D. Castanho, M. Guerreiro, L. Silva, J. Eckert, T. A. Alves, Y. d. S. Tadano, S. L. Stevan, H. V. Siqueira, and F. C. Corrêa, "Method for SoC Estimation in Lithium-Ion Batteries Based on Multiple Linear Regression and Particle Swarm Optimization," *Energies*, vol. 15, no. 19, p. 6881, Sept. 2022.

[32] M. S. Hossain Lipu, M. A. Hannan, A. Hussain, and M. H. Saad, "Optimal BP neural network algorithm for state of charge estimation of lithium-ion battery using PSO with PCA feature selection," *J. Renew. Sustain. Energy*, vol. 9, no. 6, p. 064102, Dec. 2017.

[33] B. Xia, D. Cui, Z. Sun, Z. Lao, R. Zhang, W. Wang, W. Sun, Y. Lai, and M. Wang, "State of charge estimation of lithium-ion batteries using

optimized Levenberg-Marquardt wavelet neural network," *Energy*, vol. 153, pp. 694–705, Jun. 2018.

[34] L. W. Kang, X. Zhao, and J. Ma, "A new neural network model for the state-of-charge estimation in the battery degradation process," *Appl. Energy*, vol. 121, pp. 20–27, May. 2014.

[35] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, Dec. 2006.

[36] M. S. Hossain Lipu, M. A. Hannan, A. Hussain, M. H. Saad, A. Ayob, and M. N. Uddin, "Extreme learning machine model for state-of-charge estimation of lithium-ion battery using gravitational search algorithm," *IEEE Trans. Ind. Appl.*, vol. 55, no. 4, pp. 4225–4234, Jul. 2019.

[37] J. Nalepa and M. Kawulok, "Selecting training sets for support vector machines: a review," *Artif Intell Rev*, vol. 52, no. 2, pp. 857–900, Jan. 2018.

[38] J. N. J. Hu, J. N. J. Hu, H. B. Lin, X. P. Li, C. L. Jiang, X. H. Qiu, and W. S. Li, "State-of-charge estimation for battery management system using optimized support vector machine for regression," *J. Power Sources*, vol. 269, pp. 682–693, Dec. 2014.

[39] J. C. Alvarez Anton, P. J. Garcia Nieto, C. Blanco Viejo, and J. A. Vilan Vilan, "Support vector machines used to estimate the battery state of charge," *IEEE Trans. Power Electron.*, vol. 28, no. 12, pp. 5919–5926, 2013.

[40] Z. Deng, L. Yang, Y. Cai, H. Deng, and L. Sun, "Online available capacity prediction and state of charge estimation based on advanced data-driven algorithms for lithium iron phosphate battery," *Energy*, vol. 112, pp. 469–480, Oct. 2016.

[41] E. Schulz, M. Speekenbrink, and A. Krause, "A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions," *J Math Psychol*, vol. 85, pp. 1–16, Aug. 2018.

[42] 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, p. 118000, Aug. 2020.

[43] E. Chemali, P. J. Kollmeyer, M. Preindl, and A. Emadi, "State-of-charge estimation of Li-ion batteries using deep neural networks: A machine learning approach," *J. Power Sources*, vol. 400, pp. 242–255, Oct. 2018.

[44] Y. Li, K. Li, X. Liu, and L. Zhang, "Fast battery capacity estimation using convolutional neural networks," *Trans. Inst. Meas. Control.*, Nov. 2020.

[45] E. Chemali, P. J. Kollmeyer, M. Preindl, R. Ahmed, and A. Emadi, "Long Short-Term Memory Networks for Accurate State-of-Charge Estimation of Li-ion Batteries," *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6730–6739, Aug. 2018.

[46] F. Yang, W. Li, C. Li, and Q. Miao, "State-of-charge estimation of lithium-ion batteries based on gated recurrent neural network," *Energy*, vol. 175, pp. 66–75, May. 2019.

[47] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May. 2015.

[48] K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis, M. Z. Bazant, S. J. Harris, W. C. Chueh, and R. D. Braatz, "Data-driven prediction of battery cycle life before capacity degradation," *Nat. Energy*, vol. 4, no. 5, pp. 383–391, Mar. 2019.

[49] H. Dai, G. Zhao, M. Lin, J. Wu, and G. Zheng, "A Novel Estimation Method for the State of Health of Lithium-Ion Battery Using Prior Knowledge-Based Neural Network and Markov Chain," *IEEE Trans. Ind. Electron.*, vol. 66, no. 10, pp. 7706–7716, Oct. 2019.

[50] 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. Transp. Electrif.*, vol. 6, no. 2, pp. 363–374, Jun. 2020.

[51] L. Chen, H. Wang, B. Liu, Y. Wang, Y. Ding, and H. Pan, "Battery state-of-health estimation based on a metabolic extreme learning machine combining degradation state model and error compensation," *Energy*, vol. 215, p. 119078, Jan. 2021.

[52] W. Liu, Y. Xu, and X. Feng, "A Hierarchical and Flexible Data-Driven Method for Online State-of-Health Estimation of Li-Ion Battery," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14 739–14 748, Dec. 2020.

[53] 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, p. 117957, Aug. 2020.

[54] J. Wei, G. Dong, and Z. Chen, "Remaining Useful Life Prediction and State of Health Diagnosis for Lithium-Ion Batteries Using Particle Filter and Support Vector Regression," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5634–5643, Jul. 2018.

[55] K. Liu, Y. Li, X. Hu, M. Lucu, and W. D. Widanage, "Gaussian Process Regression with Automatic Relevance Determination Kernel for Calendar Aging Prediction of Lithium-Ion Batteries," *IEEE Trans. Ind. Inform.*, vol. 16, no. 6, pp. 3767–3777, Jun. 2020.

[56] 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. 56–67, May. 2019.

[57] L. Song, K. Zhang, T. Liang, X. Han, and Y. Zhang, "Intelligent state of health estimation for lithium-ion battery pack based on big data analysis," *J. Power Storage*, vol. 32, p. 101836, Dec. 2020.

[58] Z. Chen, H. Zhao, Y. Zhang, S. Shen, J. Shen, and Y. Liu, "State of health estimation for lithium-ion batteries based on temperature prediction and gated recurrent unit neural network," *J. Power Sources*, vol. 521, p. 230892, Feb. 2022.

[59] J. Hong, Z. Wang, W. Chen, L. Wang, P. Lin, and C. Qu, "Online accurate state of health estimation for battery systems on real-world electric vehicles with variable driving conditions considered," *J. Clean. Prod.*, vol. 294, p. 125814, Apr. 2021.

[60] C. Birkl, "Oxford battery degradation dataset 1," 2017.

[61] B. Saha and K. Goebel, "Battery data set," *NASA AMES prognostics data repository*, 2007.

[62] 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. 10 314–10 321, Dec. 2011.

[63] L. Song, K. Zhang, T. Liang, X. Han, and Y. Zhang, "Intelligent state of health estimation for lithium-ion battery pack based on big data analysis," *J. Power Storage*, vol. 32, p. 101836, Dec. 2020.

[64] 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, p. 118000, Aug. 2020.

[65] F. Yang, S. Zhang, W. Li, and Q. Miao, "State-of-charge estimation of lithium-ion batteries using LSTM and UKF," *Energy*, vol. 201, p. 117664, Jun. 2020.

[66] Y. Tian, R. Lai, X. Li, L. Xiang, 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, p. 114789, May. 2020.

[67] X. Li, Z. Wang, and L. Zhang, "Co-estimation of capacity and state-of-charge for lithium-ion batteries in electric vehicles," *Energy*, vol. 174, pp. 33–44, May. 2019.

[68] Z. Song, J. Hou, X. Li, X. Wu, X. Hu, H. Hofmann, and J. Sun, "The sequential algorithm for combined state of charge and state of health estimation of lithium-ion battery based on active current injection," *Energy*, vol. 193, p. 116732, Feb. 2020.

[69] Y. Song, D. Liu, H. Liao, and Y. Peng, "A hybrid statistical data-driven method for on-line joint state estimation of lithium-ion batteries," *Appl. Energy*, vol. 261, p. 114408, Mar. 2020.

[70] X. Lai, L. He, S. Wang, L. Zhou, Y. Zhang, T. Sun, and Y. Zheng, "Co-estimation of state of charge and state of power for lithium-ion batteries based on fractional variable-order model," *J. Clean. Prod.*, vol. 255, p. 120203, May. 2020.

[71] F. Feng, S. Teng, K. Liu, J. Xie, Y. Xie, B. Liu, and K. Li, "Co-estimation of lithium-ion battery state of charge and state of temperature based on a hybrid electrochemical-thermal-neural-network model," *J. Power Sources*, vol. 455, p. 227935, Apr. 2020.

[72] W. Li, Y. Xie, X. Hu, Y. Zhang, H. Li, and X. Lin, "An Online SOC-SOTD Joint Estimation Algorithm for Pouch Li-Ion Batteries Based on Spatio-Temporal Coupling Correction Method," *IEEE Trans. Power Electron.*, vol. 37, no. 6, pp. 7370–7386, Jun. 2022.