Estimation of Lithium-Ion Battery State of Charge and Health Using LSTM Networks

Nahed Ghanay
Université de Lorraine, LCOMS
Metz, France
LISIER, University of Tunis
Tunis, Tunisia

Abdelmoudjib Benterki
ESTACA École d'Ingénieurs
Paris, France

Moussa Boukhnifer
Université de Lorraine, LCOMS
Metz, France

Achraf Jabeur Telmoudi
LISIER
University of Tunis
Tunis, Tunisia

***Abstract*—This paper proposes a Long Short-Term Memory (LSTM) based approach for accurate State Of Charge (SOC) and State Of Health (SOH) estimation in lithium-ion batteries, which is critical for improving the safety and longevity of electric vehicles. The nonlinear dynamics of batteries, influenced by factors such as temperature, voltage, current, and SOC, poses significant challenges to traditional estimation methods. Using the Long-Term Degradation dataset, our LSTM model captures temporal dependencies and complex electrochemical interactions to predict SOC and SOH under varying operating conditions. The experimental results demonstrate robust performance, with mean squared errors as low as 5.3121×10^{-5} for the estimation of SOC and 5.572×10^{-5} for the estimation of SOH for different current profiles. The proposed framework provides a scalable solution for real-time battery management systems, reducing the reliance on manual feature extraction and enabling generalization across different battery technologies.**

***Index Terms*—Lithium-ion batteries, long-term degradation dataset, LSTM, SOC estimation, SOH estimation.**

I. Introduction

The global dependence on fossil fuels is expected to account for around 55% of total energy consumption by 2040, with road transport playing a major role in CO_2 emissions. Fuel consumption increased by 8% in 2001 and reached 40% of total energy consumption in 2014. Without intervention, CO emissions $_2$ from the transport sector could increase by 60% by 2050 [1]. As diesel and petrol vehicles emit almost twice as many pollutants as electric vehicles, these fuels are considered inefficient and unsustainable.

The growing demand for electric vehicles highlights the need to develop models capable of simulating and estimating battery parameters in real-time while considering their dynamics and non-linearity. The diagnosis and prognosis of electrochemical phenomena play a key role in defining the state of health of batteries. Batteries can nevertheless suffer premature aging, often due to internal short-circuits (electrode or separator failure) or external short-circuits (immersion or contamination), leading to thermal runaway. This can lead to catastrophic fires or explosions.

To address this issue, researchers have developed frameworks to estimate battery SOC and SOH, enhancing diagnostic and prognostic performance.

Researchers typically divide the estimation of the SOC and SOH of lithium ion batteries into two main categories: classical model-based approaches and AI-based data-driven methods. Classical methods, such as equivalent circuit models (ECMs) and electrochemical models, rely on parameter estimation using measurable signals like voltage, current, and temperature to infer internal battery states [2], [3]. These approaches offer good physical interpretability and are suitable for real-time applications because of their relatively low computational cost. However, parameter inaccuracies, the need for extensive calibration, and difficulties in modeling battery aging and adapting to varying operating conditions often limit these approaches.

Artificial intelligence (AI) has emerged as a transformative technology with applications in various fields. Machine learning and deep learning techniques have been found to have a successful implementation in healthcare, finance, and autonomous systems [4], [5]. These advances enabled the development of intelligent systems capable of performing complex tasks with high accuracy and efficiency.

Machine learning-based battery capacity estimation techniques have attracted a great deal of interest as an alternative strategy in recent years. These techniques are simple to use and require no prior understanding of the complex physical concepts underlying batteries. Several techniques have been used effectively for battery capacity estimation, including semi-supervised transfer component analysis [6], and support vector regression (SVR) [7]. In [6], a semi-supervised method is proposed using transfer component analysis to estimate the SOH of lithium-ion batteries. Reduce reliance on labeled data by aligning data distributions and preserving key characteristics, achieving an accuracy (under error 2. 5%. Althoughugh it adapts well to different battery conditions, it depends the quality of thequality of the source data and may face computational challenges in real-time use. In [7], an open-circuit voltage model is combined with a Thevenin equivalent circuit model. A particle swarm optimization-based SVR approach is employed to validate the state-of-health estimation through various tests. However, the method is limited by high computational complexity, dependence on accurate models, and sensitivity to initial conditions. In [8], a Gaussian Process Regression (GPR) model with an Automatic Relevance Determination (ARD) kernel was applied for the calendar aging prediction of lithium-ion batteries. It enabled uncertainty estimation and feature relevance without relying on physical models. Limitations included high computational cost and dependence on training data quality. According to [9], a Support Vector Machine (SVM)-based method was applied for online SOH estimation of lithium-ion batteries using a partial charging segment. However, the method was limited by its sensitivity to the quality of the charging segment and the training data configuration. Typically, machine learning methods rely heavily on diverse training data and often struggle to adapt to new conditions, reducing accuracy in real-world scenarios with varied uses and environments.

To overcome the precision limitations of the previously discussed methods, AI-based techniques—particularly deep learning models, such as recurrent neural networks (RNN) [10], random forests [11], and convolutional neural networks (CNNs) [12], have been used to automatically extract features from the data to alleviate this bottleneck. In [10], Elman's neural network was used to predict the battery's capacity for subsequent cycles in real time until the end of its lifetime. The network inputs were limited to the charged capacities of previous cycles. In [11], a random forest regression model was trained using the load capacity recorded in a particular voltage region as input data. However, this method is limited by its sensitivity to temperature changes and the need for retraining on different battery types. In addition, [12] trained a CNN model for capacity estimation, directly using the voltage, current, and load capacitance of each cycle as inputs. However, CNNs are primarily designed for image processing, while this case involves temporal signals. This necessitates the prior conversion of data into matrices, complicating the process. In this regard, an RNN well suited for sequential data—offers a more appropriate alternative."

In this context, an LSTM-based recurrent neural network is a more suitable choice for estimating the battery's SOC and SOH, and this selection will be further justified in Section III.
This article is structured as follows: Section 2 describes the selection of attributes and the database, Section 3 details the LSTM-based approach for estimating the state of charge (SOC) and state of health (SOH), as well as the results obtained, and Section 4 presents the conclusion of the article.

II. Selection of features and dataset

A. Selection of features

A battery's state of health (SOH) is strongly influenced by several factors, including temperature, voltage, current, and state of charge (SOC), all of which interact to affect its long-term performance and durability.

1) Temperature: High temperatures accelerate chemical reactions in the battery, which may increase short-term charge capacity, but also reduce discharge time. Prolonged exposure to temperatures above 45°C causes accelerated degradation of the battery materials, reducing its lifespan. This can also lead to serious risks such as internal short-circuits, explosions or fires due to thermal runaway. Conversely, low temperatures, below 0°C, reduce battery capacity by reducing the amount of usable energy, and slow down electrochemical reactions. Internal resistance also increases at low temperatures, causing voltage loss and reducing battery efficiency, increasing charge times and reducing discharge performance.

2) Voltage: The voltage of a battery is a direct indicator of its state of charge (SOC). Too high or too low a voltage can put excessive stress on the battery cells, resulting in faster degradation of electrode materials and shorter battery life. Excessive voltage, which can lead to overcharging, can damage the battery's chemical structure, cause overheating, and, in extreme cases, lead to explosions or fire. Similarly, too deep a discharge, when voltage falls below a critical threshold, can cause irreversible damage to the battery's internal chemistry, making it difficult to store and restore energy. Appropriate voltage management is therefore essential to maintain the battery's chemical stability and extend its life.

3) Current: The charging and discharging current has a significant impact on battery performance and longevity. Too high a charge current can lead to degradation of the battery's internal materials, particularly the electrodes, reducing its ability to hold a charge. Similarly, too high a discharge current can lead to overheating, resulting in faster capacity loss. It's crucial to respect manufacturers' recommended current limits to avoid overcharging or over discharging. Excessively high currents can also increase battery temperature, which, combined with other factors, can accelerate material degradation and increase the risk of failure.

4) SOC: The SOC determines the amount of energy available in the battery and directly affects its ability to deliver energy optimized. An SOC that is too low or too high can reduce battery efficiency and accelerate degradation. At very low SOCs, internal resistance increases, which can lead to power loss and reduced energy efficiency. On the other hand, high SOCs, especially when combined with high temperatures, can accelerate undesirable chemical reactions and cause irreversible damage to the electrodes. Frequent charge/discharge cycles at extreme SOCs are particularly harmful and accelerate the degradation of internal materials, reducing battery life. Maintaining the SOC within an optimum range is therefore crucial to preserve battery capacity and avoid premature degradation.

The impact of each of the above parameters on battery performance is supported by previous research using these same factors, and the importance of the database, which will be detailed in the next section, justifies the choice of these parameters for battery health prognosis.

B. Selection of dataset

A database freely available on the Internet was used to develop the prognostic model. A choice was made in order to select the database best suited to the context of the study. The different protocols used by datasets are explained in Table I.

The Prognostics Data Repository, made available in 2009 by NASA's Prognostics Center of Excellence (PCoE), is one of the most cited datasets in the literature for the development of prognostics algorithms [13]. This dataset contains data from 34 NCA-type lithium batteries, with a nominal capacity of 2 Ah, subjected to charge/discharge cycles up to 70% of their initial capacity, under a variety of temperature conditions.

The Cycle Life Prediction dataset, designed by researchers at Stanford University and MIT, is based on CC-CV discharge protocols and brings together the results of tests carried out on 135 Li-ion cells of the LFP type, with a nominal capacity of 1.1 Ah, as mentioned in [14]. These tests were carried out in a thermal chamber maintained at 30°C, until the cells reached 80% of their initial capacity. However, this dataset does not include driving cycle protocols and does not allow the influence of temperature to be modeled.

Mention should also be made of the Oxford Battery Degradation dataset, designed by researchers at Oxford University [15]. This dataset covers 8 NMC-type Li-ion cells, with a capacity of 740 mAh, tested in a thermal chamber maintained at 40°C. Like the previous dataset, it does not include driving cycle protocols, and the tests were carried out at a single temperature.

The Panasonic 18650PF NMC dataset concerns an NMC battery, with a capacity of 2.9 Ah, which is subject to various discharge protocols and driving cycles varying temperature conditions. The battery is recharged at a current of 1C up to a voltage of 4.2 V.

The LG 18650HG2 dataset [16] covers an NMC battery with a capacity of 3 Ah, tested in a thermal chamber at McMaster University. The tests include a series of various protocols: an HPPC (Hybrid Pulse Power Characterization) test, various DC-DC charge/discharge protocols, and a series of 12 driving cycles. The drive cycles used are as follows: HWFET, UDDS, LA92, and US06.

Finally, the Long-Term Degradation dataset [17] concerns 22 NMC cells with a capacity of 3 Ah, which are cycled up to 80% of their initial capacity. The charge/discharge protocols follow a CC-CV mode at currents varying between 0.5C and 3C, and the tests are carried out at temperatures of 15 ° C, 25 ° C, and 35 ° C.

The aim of this work is to predict the SOC and, SOH of Li-ion batteries using an LSTM neural network. For this, a database meeting the following criteria is required:

- The use of recent Li-ion battery technology.
- Availability of a sufficient quantity of data (several tens or hundreds of cycles), enabling exploitation for SOH estimation.
- Tests to be carried out under various temperature conditions.

Taking into account these criteria and the comparison of the different datasets presented in Table II, it can be concluded that the Long-Term Degradation database is the most appropriate.

III. LSTM SOC AND SOH ESTIMATION

There are several reasons for choosing a model based on LSTM neural networks to estimate the SOC and SOH of batteries. Firstly, battery data such as voltage, current, temperature, and capacity are time-dependent and often non-linear because of complex electrochemical processes. LSTMs are designed to capture long-term temporal dependencies and can effectively model non-linear dynamics through their deep architecture and activation functions. They also handle noisy data well and can integrate multiple input variables to reflect interactions between different battery parameters. Supported by recent studies, LSTMs provide accurate diagnostics and reliable future predictions, making them a strong choice for this application.

A. LSTM architecture

LSTMs are designed to handle sequential data while overcoming the problem of gradient disappearance or explosion encountered in traditional recurrent neural networks. The idea behind this choice of neural network architecture is to divide the signal between what is important in the short term, through the hidden state, and what is important in the long term, through the cell state as shown in Fig. 1

An LSTM consists of three main gates [18], each with a distinct role in memory management in each time step:transforms the output into a probability

- **Forget Gate:** determines which information from the previous state is retained or deleted. As shown in equation (1) :

$$f(t) = \sigma(W_f \cdot [h(t-1), x(t)] + b_f) \tag{1}$$

  The sigmoid function σ provides a probability for each memory element, based on the previous hidden state h(t-1), current input x (t), weights W_f, and bias b_f.
- **Input Gate:** decides which new information should be added to the memory state in two steps. First, it computes the importance of the incoming information:

$$i(t) = \sigma(W_i \cdot [h(t-1), x(t)] + b_i) \tag{2}$$

  Then, it generates candidate values to potentially add to memory:

$$g(t) = \tanh(W_g \cdot [h(t-1), x(t)] + b_g) \tag{3}$$

  In (2) and (3), the weights W_i, W_g and the biases b_i, W_g are learned during training. The hyperbolic tangent function, tanh, is used to constrain the values between -1 and 1. This gate combines i(t) and g(t) to update the memory state.

TABLE I: The various protocols used by datasets

Protocol	Definition
CC-CV	A common charging method that switches between constant current (CC) charging and constant voltage (CV) charging, depending on the voltage of the rechargeable battery.
HPPC	A technique based on intermittent current pulses followed by rest periods.
HWFET (Highway Fuel Economy Test)	A protocol simulating highway driving conditions at a speed of 100 km/h.
UDDS (Urban Dynamometer Driving Schedule)	A dynamometer driving schedule simulating urban driving conditions for a light-duty vehicle.
LA92 (Los Angeles)	Same principle as the UDDS protocol, but adapted to the specific conditions of Los Angeles.
US06	A test simulating driving conditions for electric vehicles with long acceleration phases.

TABLE II: Advantages and disadvantages of different databases

Dataset	Advantages	Disadvantages
Prognostics data repository (PCoE)	Useful for SOC and SOH estimation	Old battery technology Low quantity of usable data Only 3 different temperatures
Cycle Life Prediction Dataset - Stanford et MIT	Useful for SOC and SOH estimation Several batteries are tested	Only one temperature condition No driving cycle One discharge protocol used
Oxford Battery Degradation	Useful for SOC and SOH estimation	Only one temperature condition No driving cycle
Panasonic 18650PF NMC	Several temperature conditions Multiple driving cycles	Impedance data not available Single battery used for all protocols
LG 18650HG2	Several driving cycles Several temperature conditions useful for SOC estimation	One battery used for all protocols
Long-Term Degradation	Various temperature conditions useful for SOC and SOH estimation	

- **Output Gate:** controls which information from the current memory state is important to generate the hidden state h(t)

$$o(t) = \sigma(W_o \cdot [h(t-1), x(t)] + b_o) \tag{4}$$

$$h(t) = o(t) \odot \tanh g(t) \tag{5}$$

In (4), o(t) is calculated using the weights W_o and b_o that are learned during training and the hidden state from the previous step, h(t-1) . The current hidden state, h(t), is then updated on the basis of o(t) as shown in (5)

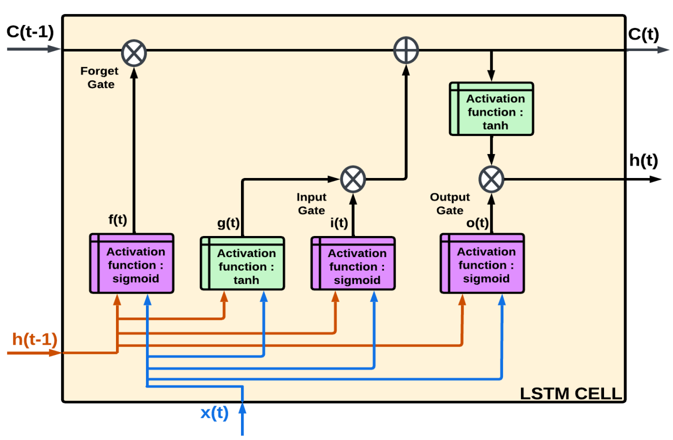


Fig. 1: Architecture of the LSTM RNN

B. LSTM training

In the chosen database, several charging and discharging protocols were tested at different temperatures to analyze battery behavior, capacity, and impact on its longevity. In this article, we will focus on three current profiles, namely **0.5C**, **0.5C-1C**, and **0.5C-2C**, at a temperature of 25 ° C.

IV. Results

Data pre-processing is first carried out, including resampling to a period of T =30s. The current battery capacity $Ccurr$ is then determined by current integration. The total capacity $Cfull$ is estimated using linear interpolation between peaks. SOC and SOH are then calculated according to equations (6) and (7). Finally, attributes are extracted for model training, followed by a data normalization step.

$$SOC = C_{curr}/C_{full} \tag{6}$$

$$SOH = C_{full}/C_{nom} \tag{7}$$

TABLE III: LSTM model hyperparameters

Number of hidden layers	1
Neurons number	100
Epochs number	100
Mini-batch size	64
Sequence number	300
Sequence length	300
Optimizer	Adam
Learning rate	0.001

Then, for SOC estimation, current, voltage, and current capacity were considered as model inputs. For SOH estimation, the same inputs as for SOC were used, but with the SOC value added as an additional parameter.

For both SOC and SOH estimation tasks, 80% of the data was used for training and 20% for validation. The

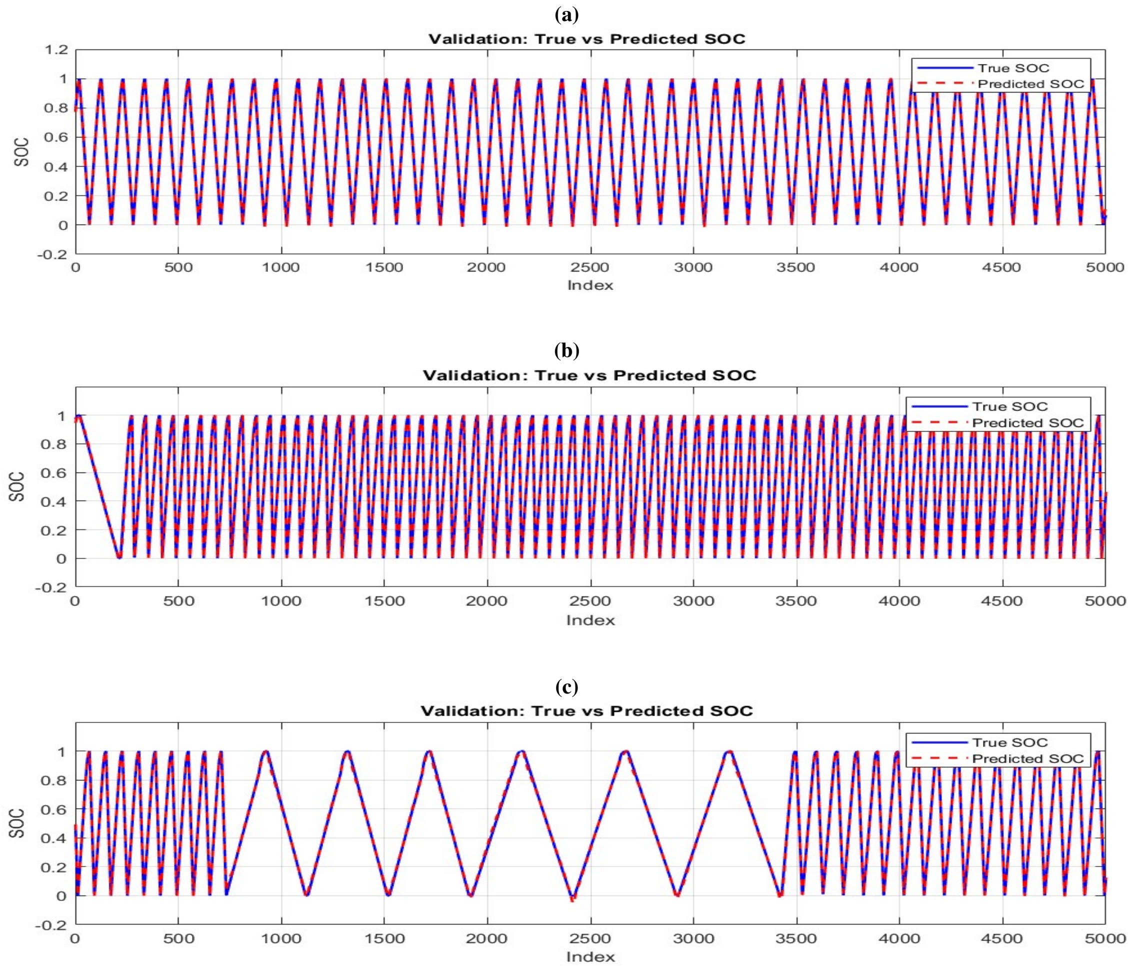


Fig. 2: SOC estimation based on LSTM. (a) Current profile at 0.5C, (b) Current profile at 0.5C-2C, and (c) Current profile at 0.5C-1C.

LSTM model's performance was assessed under three different current profiles, with the resulting curves shown in Fig. 2 and Fig. 3. For SOC estimation, the 0.5C–1C profile yielded the highest accuracy, with a mean squared error of 5.3121×10^{-5}, although it required a longer training time as illustrated in table IV .

TABLE IV: Mean squared errors and training durations for different current profiles for SOC estimation.

Current profile	Mean squared error	Training duration
0.5C	7.022×10^{-5}	38 hours
0.5C-2C	$9,8606 \times 10^{-5}$	54 hours
0.5C-1C	5.3121×10^{-5}	88 hours

For SOH estimation , the 0.5C profile resulted in the lowest error 7.51×10^{-6} and the shortest training duration as shown in table V.

These results indicate that more complex current profiles improve the estimation accuracy, likely due to the richer dynamics they provide during training. However, they also demand longer training times. The variation in MSE across profiles highlights the impact of load conditions on model learning.

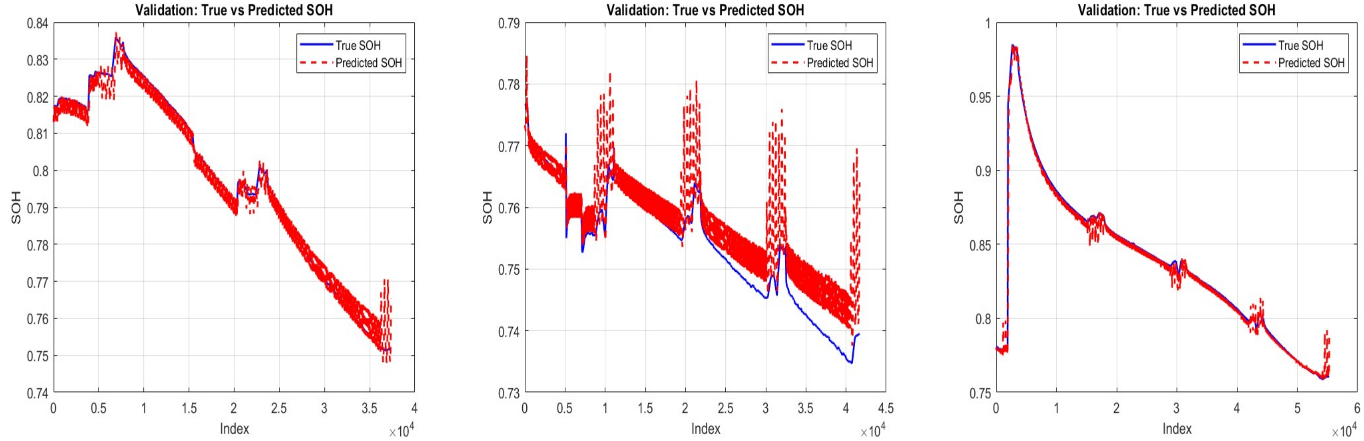
TABLE V: Mean squared errors and training durations for different current profiles for SOH estimation.

Current profile	Mean squared error	Training duration
0.5C	7.51×10^{-6}	34 hours
0.5C-2C	5.3×10^{-3}	42 hours
0.5C-1C	3.572×10^{-5}	80 hours

However, the model presents some limitations, including long training durations and reduced generalization to operating conditions not seen during training.

V. Conclusion

In this work, we developed an LSTM-based framework for SOC and SOH estimation in lithium-ion batteries, addressing key limitations of conventional machine learning methods. By leveraging the temporal learning capabilities of LSTMs, our model effectively captures complex, nonlinear relationships among critical battery parameters such as temperature, voltage, current, and SOC. The proposed framework demonstrates high estimation accuracy under dynamic current profiles, with low mean square errors and reduced training times, highlighting its practical feasibility. Despite



(a) **(b)** **(c)**

Fig. 3: SOH estimation based on LSTM : (a) Current profile at 0.5C, (b) Current profile at 0.5C-2C, and (c) Current profile at 0.5C-1C.

these promising results, several challenges remain. These include reducing computational overhead for deployment in embedded systems, mitigating the influence of measurement noise particularly in SOH estimation, and enhancing generalization across varying usage conditions. Importantly, future work will incorporate temperature variation more comprehensively, not just as an input feature but as a dynamic condition affecting degradation mechanisms. This would provide deeper insight into thermal effects and improve robustness under real-world operating environments. Future developments will explore hybrid architectures that combine LSTMs with attention mechanisms, physics-informed layers, and optimization algorithms to improve interpretability and adaptability further. Broadening the diversity of datasets including extreme temperatures and aging scenarios will be essential for wider applicability. These efforts aim to accelerate the deployment of intelligent, data-driven battery management strategies to support the transition to scalable, sustainable electric mobility.

REFERENCES

[1] R. Sim, S. Lee, W. Li, and A. Manthiram, "Influence of Calendering on the Electrochemical Performance of $LiNi_{0.9}Mn_{0.05}Al_{0.05}O_2$ Cathodes in Lithium-Ion Cells," *ACS Appl. Mater. Interfaces*, vol. 13,pp. 42898–42908, Aug. 2021.

[2] Meng, J., Boukhnifer, M., Diallo, D., & Wang, T. (2020). "A new cascaded framework for lithium-ion battery state and parameter estimation". Applied Sciences, 10(3), 1009.

[3] Meng, J., Boukhnifer, M., & Diallo, D. (2020, September)."Lithium-ion battery monitoring and observability analysis with extended equivalent circuit model". In 2020 28th Mediterranean Conference on Control and Automation (MED) (pp. 764-769). IEEE.

[4] Benterki, A., Boukhnifer, M., Judalet, V., & Maaoui, C. (2020). "Artificial intelligence for vehicle behavior anticipation: Hybrid approach based on maneuver classification and trajectory prediction". IEEE Access, 8, 56992-57002.

[5] Benterki, A., Judalet, V., Choubeila, M., & Boukhnifer, M. (2019, October). "Long-term prediction of vehicle trajectory using recurrent neural networks". In IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society (Vol. 1, pp. 3817-3822). IEEE.

[6] Y. Li, H. Sheng, Y. Cheng, D.-I. Stroe, and R. Teodorescu, "State-of-health estimation of lithium-ion batteries based on semi-supervised transfer component analysis," *Applied Energy*, vol. 278, p. 115703, 2020, Elsevier.

[7] D. Yang, Y. Wang, R. Pan, R. Chen, and Z. Chen, "State-of-health estimation for the lithium-ion battery based on support vector regression," *Applied Energy*, vol. 230, pp. 234–242, 2018, Elsevier.

[8] 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 Transactions on Industrial Informatics*, vol. 15, no. 9, pp. 5537–5546, Sep. 2019.

[9] X. Feng, C. Weng, X. He, X. Han, L. Lu, D. Ren, and M. Ouyang, "Online state-of-health estimation for Li-ion battery using partial charging segment based on support vector machine," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3549–3557, Apr. 2019.

[10] G. W. You, S. Park, and D. Oh, "Diagnosis of electric vehicle batteries using recurrent neural networks," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 10, pp. 8191–8199, Oct. 2017.

[11] Y. Li, C. Zou, M. Berecibar, E. Nanini-Maury, J. C. W. Chan, P. Van den Bossche, and J. Van Mierlo, "Random forest regression for online capacity estimation of lithium-ion batteries," *Applied Energy*, vol. 228, pp. 1782–1790, 2018, Elsevier.

[12] Y. Li and J. Tao, "CNN and transfer learning based online SOH estimation for lithium-ion battery," Ningbo Institute of Technology, Zhejiang University, Ningbo, China, and College of Control Science and Engineering, Zhejiang University, Hangzhou.

[13] K. Goebel, B. Saha, A. Saxena, J. R. Celaya, and J. P. Christophersen, "Prognostics in battery health management," *IEEE Instrumentation and Measurement Magazine*, vol. 11, pp. 25–31, Aug. 2008.

[14] 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," *Nature Energy*, vol. 4, pp. 383–391, May 2019.

[15] C. R. Birkl, "Oxford Battery Degradation Dataset 1," University of Oxford, 2017. [Online]. Available: https://ora.ox.ac.uk/objects/uuid:03ba4b01-cfed-46d3-9b1a-7d4a7bdf6fac

[16] P. Kollmeyer, C. Vidal, M. Naguib, and M. Skells, "LG 18650HG2 Li-ion Battery Data and Example Deep Neural Network xEV SOC Estimator Script," Mendeley Data, V3, Mar. 2020. [Online]. Available: https://data.mendeley.com/datasets/cp3473x7xv/3

[17] S. National Laboratories, "Degradation of Commercial Lithium-ion Cells as a Function of Chemistry and Cycling Conditions," BatteryArchive.org. [Online]. Available: https://www.batteryarchive.org/snl_study.html

[18] Kamilya Smagulova and Alex Pappachen James, "A survey on LSTM memristive neural network architectures and applications". *The European Physical Journal Special Topics*, vol. 228, no. 10, pp. 2313–2324, 2019. urlhttps://doi.org/10.1140/epjst/e2019-900046-x