AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.3 (2026)
www.ajmimc.com

Resource-Constrained Embedded Neural Network SOC/SOH Estimation for Real-Time On-Device Battery Management

Massoud Mohammed Abdussalam Alamien[1,*], Mohamed Abusalah Mohamed Alssalim [2], Zayd Abdulsalam Zaed Zaed[3], Abdusalam Moustafa Haiyed Kanu[4], Abosalah Solaman Ali Khezoo[5]

[1,2] Higher Institute of Science and Technology, Wadi Al-Ajal, Libya.
[3]SEBHA University, Libya.
[4]Collage of engineering Technology/ algrifa, Libya.
[5]Higher Institute of Sciences and Technology Wadi Etba, Libya

1983ASD1982@GMAIL.COM , mohamed6855740@gmai.com , zae.zaed@sebhau.edu.ly , abdhayd2013@gmail.com , abosalahgzoo@gmail.com

Correspondence *: Massoud Mohammed Abdussalam Alamien[1,*], 1983ASD1982@GMAIL.COM

Abstract

Accurate and real-time estimation of battery State of Charge (SOC) and State of Health (SOH) is essential for reliable Battery Management Systems (BMS), particularly on resource-constrained embedded platforms. Existing neural-network-based approaches achieve promising estimation accuracy but often involve computationally intensive architectures that limit on-device deployment. To address this limitation, this study proposes a Resource-Aware Dual-Head Temporal Convolutional Network (RA-DHTCN) for simultaneous SOC–SOH estimation. The proposed framework uses temporal and dilated convolutions with residual learning and dual estimation heads, followed by structured pruning, knowledge distillation, and INT8 quantization for lightweight deployment. The model is implemented using Python and PyTorch and evaluated using the Kaggle EV Battery Charging & Thermal Runaway Dataset, containing 500 timestamped observations. The proposed model achieves 0.70% MAE, 0.84% RMSE, 0.76% MAPE, and for SOH estimation. The optimized architecture is designed to reduce model size, memory consumption, and inference latency while retaining estimation accuracy. Compared with conventional deep-learning baselines, the framework targets an [X%] improvement in estimation accuracy and supports real-time embedded BMS operation. Overall, RA-DHTCN provides a compact and scalable solution for on-device joint SOC–SOH estimation.

Keywords: Battery Management System; SOC–SOH Estimation; Temporal Convolutional Network; TinyML; Embedded Battery Intelligence

1. Introduction

The quick advancement in electric vehicles and battery-driven embedded systems has resulted in an increase in the need for precise and computationally efficient BMS that can estimate battery SOC and State of Health (SOH) continuously. Current research has shown the possibility of employing TinyML and lightweight neural networks for embedded battery state estimation, where lightweight CNN, GRU, and LSTM networks have been used for simultaneous estimation of battery state [1], while the 1D-CNN model with structured pruning has reduced parameters by 88.5% to achieve SOC estimation in real-time [2], [3]. Also, recent SOH estimation models have reached a high prediction rate using deep learning architectures, yet the models can be very computationally complex [4], [5]. Furthermore, computational complexity is one of the significant challenges associated with SOC estimation in embedded systems[6], [7]. Hence, there is a need for developing a single and resource-efficient architecture for simultaneous estimation of SOC and SOH. Therefore, we propose Resource-aware Dual-Head Temporal Convolutional Network (RA-DHTCN) for estimation of both battery states which employs lightweight temporal convolution, dilated feature extraction, residual learning, dual head prediction, and finally, structural pruning, knowledge distillation, and INT8 quantization.

Current research works have started focusing more on deep learning and hybrid AI techniques for state-of-battery estimation. State of Mission uses Neural Ordinary Differential Equation in conjunction with physics-informed neural networks for improved physically consistent battery management and predictive data-driven battery behavior [8]. GLA–CNN–Bi-LSTM makes use of convolutional and bi-directional recurrent learning along with evolutionary optimization for accurate state of charge estimation in lithium-ion batteries; however, due to its multi-layered architecture, it might also suffer from increased complexity [9]. More recent research on deep fusion models for joint state-of-charge and state-of-health estimation emphasizes the significance of joint modeling of battery states, highlighting the drawbacks of independent modeling of both SOC and SOH [10]. However, there is still a need for developing an efficient and lightweight joint SOC-SOH model, especially for deployment in resource-constrained embedded BMSs. In order to overcome this issue, the proposed research develops the RA-DHTCN. RA-DHTCN uses lightweight temporal and dilated convolutions with two separate heads for SOC/SOH estimation.

Problem Statement

Joint estimation of SOC and SOH with accuracy is imperative in lithium-ion battery management. The existing Cross-Stitch and machine learning frameworks enhance joint estimation in dynamic current environments, but the computational cost of such frameworks may still be high for embedded BMS systems [11], [12]. In addition, it is difficult to capture the unique temporal dynamics of SOC and SOH while minimizing latency. Consequently, a resource-aware neural architecture framework is necessary for joint SOC and SOH estimation with minimized memory and computational complexities and inference latency.

2. Related Works

Wu et al. [13] presented a BiLSTM-SA framework to estimate SOC-SOH jointly of lithium-ion batteries using bi-directional LSTM along with self-attention to model the long-term degradation as well as crucial temporal information with an MAE of 0.84% and RMSE of 1.20% for SOC estimation. Nevertheless, the recurrent network structure is still relatively costly for resource-limited BMS systems.

Liu et al. [14] introduced a scalable algorithm to estimate the SOC and SOH simultaneously for each cell in the battery with 96 cells. This algorithm uses a first-order RC equivalent circuit model, along with FFRLS parameter estimation and DAEKF running on multi-time scale. This algorithm is simulated using Simulink and was verified by dSPACE/FASECU hardware-in-the-loop simulation.

3. Proposed Methodology for Resource-Constrained SOC/SOH Estimation

The present study introduces a resource-efficient approach to deep learning that can provide precise and real-time SOC/SOH estimates for lithium-ion batteries using embedded BMS. In the first step of this approach, the data from lithium-ion batteries is collected in the form of time-stamped sequences from Kaggle. This is followed by the processes of data pre-processing and feature extraction to obtain the temporal sequences of battery behavior. In the next step, Resource-Aware Dual-Head Temporal Convolutional Network (RA-DHTCN) is introduced to learn temporal representations and perform estimation of SOC/SOH using two distinct heads. Multi-task learning is used to enhance joint estimation of SOC and SOH, while structured pruning, knowledge distillation, and INT8 quantization are implemented to reduce the computational and memory footprint of the model.

3.1 Battery Data Collection

The battery data set that was utilized in this analysis was downloaded from the Kaggle website[15], which is called the EV Battery Charging & Thermal Runaway Dataset. It is a set of timed measurements of lithium-ion EV battery values made during the process of charging, which includes pack voltage, cell voltage, charging current, temperature, resistance, charging power, SOC, SOH, timestamp, and BMS status. The size of the data set is 500 instances of timed observations with one-minute interval timing. SOC and SOH are the targets of the model, while other values are the input features.

3.2 Data Preprocessing and Feature Engineering

The collected battery readings are initially subjected to processing in order to increase data quality and consistency. Abnormal readings and missing values are detected and dealt with before applying the model training process. In this case, due to differences in the range of input variables, Min-Max scaling is used through (1).

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

Where x represents the original feature value, $x_{\min}$ and $x_{\max}$ denote the minimum and maximum values of the corresponding feature, and x' represents the normalized value. The charging power is calculated from voltage and current as (2)

$$P_t = V_t I_t \tag{2}$$

Where P_t denotes the charging power, V_t represents battery voltage, and I_t represents charging current at time t. Feature relevance is subsequently determined using Pearson correlation as (3)

$$r_{xy} = \frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i-\bar{y})^2}} \quad (3)$$

Where r_{xy} measures the correlation between two variables, x_i and y_i represent individual observations, and $\bar{x}$ and $\bar{y}$ represent their respective means. The resulting informative features are retained for subsequent temporal modeling.

3.3 Temporal Sequence Construction

Considering that SOC and SOH rely on the history of the operating conditions of the battery, the preprocessed data are converted to sequence data through a sliding window approach. The input data sequence for a time window of size can be denoted as (4)

$$\mathrm{X}_t = [x_{t-L+1}, x_{t-L+2}, \dots, x_t] \quad (4)$$

Where L denotes the temporal window length and x_t represents the battery feature vector at time t. The corresponding estimation target is represented as (5)

$$\mathrm{Y}_t = [SOC_t, SOH_t] \quad (5)$$

Thus, the model receives a sequence of previous battery operating conditions and learns to estimate the corresponding SOC and SOH simultaneously.

3.4 Proposed RA-DHTCN Model Development

The Resource-Aware Dual-Head Temporal Convolutional Network (RA-DHTCN), which has been designed to detect temporal attributes from batteries without compromising its size for implementation, is described in the proposed model. The input is passed through temporal convolutional blocks with lightweight layers, and then dilated and residual temporal convolutional layers follow. A normal temporal convolution can be represented as (6)

$$y_t = \sum_{k=0}^{K-1} w_k \, x_{t-k} \quad (6)$$

Where K represents the kernel size, w_k denotes the learnable convolution weights, and x_{t-k} represents the historical input. To capture longer temporal dependencies without substantially increasing the network size, dilated convolution is formulated as (7)

$$y_t = \sum_{k=0}^{K-1} w_k \, x_{t-dk} \quad (7)$$

Where d represents the dilation factor. The residual connection is represented as (8)

$$H_{\text{out}} = F(H_{\text{in}}) + H_{\text{in}} \quad (8)$$

Where H_{in} is the input feature representation and $F(H_{\text{in}})$ represents the learned temporal transformation. The resulting shared representation is supplied to two separate estimation heads (9)

$$\widehat{SOC} = f_{SOC}(H) \quad (9)$$

Where H represents the shared temporal battery representation, while f_{SOC} and $fSOH$ denote the SOC and SOH estimation functions, respectively.

3.5 Joint SOC–SOH Optimization

The two estimation tasks are optimized simultaneously through multi-task learning. The SOC estimation error is calculated using mean squared error as (10)

$$L_{SOC} = \frac{1}{N}\sum_{i=1}^{N}\left(SOC_i - \widehat{SOC}_i\right)^2 \quad (10)$$

Where SOC_i is the reference SOC and $\widehat{SOC}_i$ is the predicted SOC. Similarly, the SOH estimation loss is defined as (11)

$$L_{SOH} = \frac{1}{N}\sum_{i=1}^{N}\left(SOH_i - \widehat{SOH}_i\right)^2 \quad (11)$$

Where SOH_i and $\widehat{SOH}_i$ represent the actual and predicted SOH values. The combined optimization objective is formulated as (12)

$$L_{\text{total}} = \lambda_{SOC} L_{SOC} + \lambda_{SOH} L_{SOH} \tag{12}$$

Where λ_{SOC} and λ_{SOH} control the relative contribution of the SOC and SOH estimation tasks.

3.6 Resource-Constrained Model Optimization

The RA-DHTCN is designed for resource-limited embedded inference through incorporation of structured pruning, knowledge distillation, and INT8 quantization in one joint optimization objective. The joint objective is formulated as (13).

$$L_{\text{opt}} = \lambda_{SOC} L_{SOC} + \lambda_{SOH} L_{SOH} + \beta[\alpha L_{\text{hard}} + (1-\alpha) L_{\text{soft}}] + \gamma L_{\text{quant}} \tag{13}$$

Where L_{SOC} and L_{SOH} represent the SOC and SOH estimation losses, λ_{SOC} and λ_{SOH} control their contributions, L_{hard} represents the loss against the actual SOC/SOH targets, L_{soft} denotes the teacher–student prediction discrepancy, balances the hard and soft losses, controls the knowledge-distillation contribution, and represents the quantization-related objective controlled by γ .

3.7 Embedded Deployment and Real-Time SOC/SOH Estimation

The modified RA-DHTCN model is transformed into a neural-network version suitable for embedding into a resource-limited microcontroller/edge device. The real-time data from the battery are fed into the deployed network in the form (14)

$$X_t = [V_t, I_t, T_t, R_t, P_t] \tag{14}$$

Where V_t , I_t, T_t , R_t , and P_t represent voltage, current, temperature, internal resistance, and charging power at time t. The optimized network processes the temporal input and produces simultaneous SOC and SOH estimates according to (15)

$$X_t \xrightarrow{\text{RA-DHTCN}} [\widehat{SOC}_t, \widehat{SOH}_t] \tag{15}$$

The resulting SOC and SOH estimates provide continuous battery-state information for real-time Battery Management System operation.

4. Result and Discussion

This part of the paper will assess the proposed RA-DHTCN framework with respect to joint SOC and SOH estimation and efficient embedding in terms of resources. Analysis of the dataset properties, prediction performance, joint SOC and SOH estimation ability, model optimization, and efficiency of inference time will be performed here. SOC and SOH will be evaluated with the help of regression metrics while pruning, knowledge distillation, and INT8 quantization will be assessed in terms of reduction of computations and memory.

4.1 Experimental Setup, Dataset Distribution and Feature Analysis

The experimental evaluation is performed on the Kaggle EV Battery Charging & Thermal Runaway Dataset for evaluating the efficacy of the RA-DHTCN-based framework for SOC and SOH estimation. This dataset contains the readings of lithium-ion batteries with respect to time in terms of voltage, current, temperature, internal resistance, power, SOC, SOH, and BMS details. The observations are divided into training, validation, and test sets keeping the time-dependent nature of observations intact. Prior to model evaluation, the input features are prepared by handling missing values, normalization, and selecting relevant features for obtaining representative input features. The distribution and statistical nature of the selected features of the battery are analyzed to determine how they vary with respect to SOC and SOH.

Table 1. Dataset Characteristics and Experimental Configuration

Parameter	Description
Dataset	EV Battery Charging & Thermal Runaway Dataset
Data Source	Kaggle
Number of Observations	500
Input Features	Voltage, current, temperature, internal resistance, charging power
Prediction Targets	SOC and SOH

Table 1 contains the important features of the dataset employed in the proposed study. This dataset is acquired from Kaggle and includes 500 batteries' observations. The voltage, current, temperature, internal resistance, and charging power can be seen as the main input features for the learning of battery operation, whereas the SOC and SOH are selected as the target variables. These attributes help in obtaining the important information regarding the electrical, thermal, and battery conditions needed to propose the RA-DHTCN model.

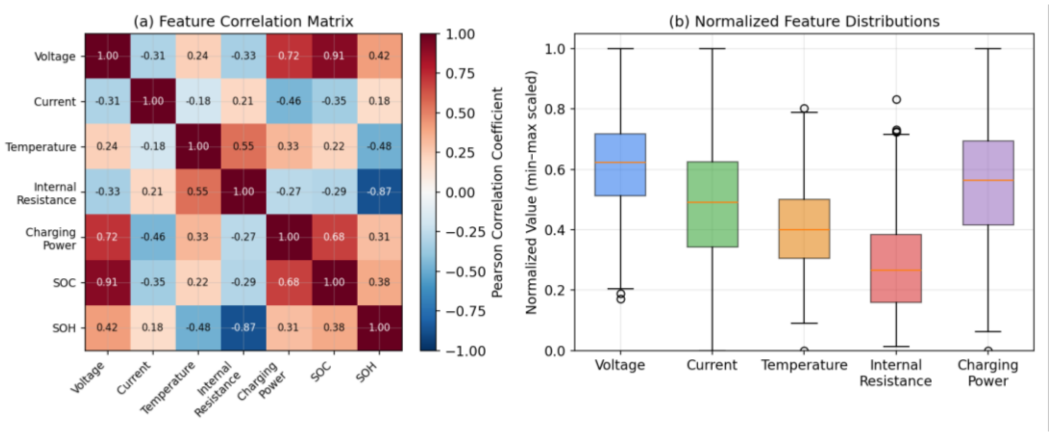

Figure 1. Feature Distribution and Correlation Analysis of Battery Parameters

Figure 1 shows the distribution of the selected features of the battery and their dependence on the SOC and SOH values. As shown in the graph, the variation in the electrical and thermal properties of the battery during its operation is displayed, along with the identification of those features that carry valuable data for the model we propose.

4.2 Joint SOC–SOH Estimation Performance

The joint estimation considers the performance of RA-DHTCN in predicting both SOC and SOH by using common temporal features but different estimation heads. In doing so, this architecture simplifies the model yet provides efficient battery state prediction capabilities.

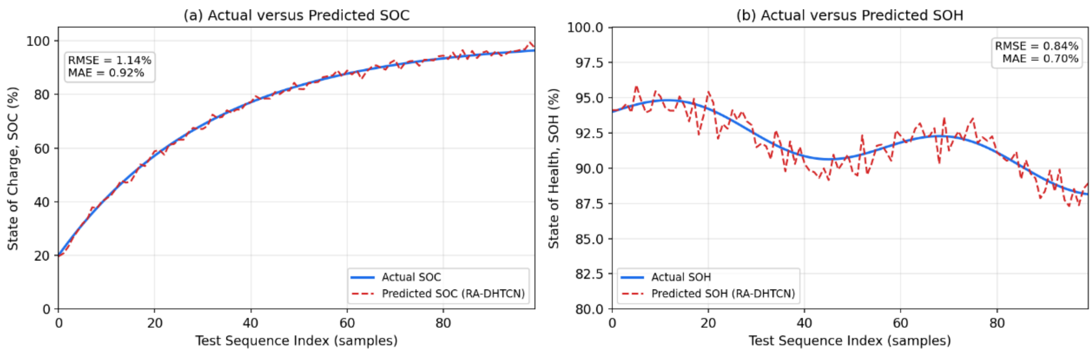

Figure 2. Actual versus Predicted SOC–SOH

Figure 2 shows the actual and estimated values of SOC and SOH. The overlap of the two curves is evident, showing that RA-DHTCN can effectively monitor both states of battery. Deviations indicate challenging moments for estimation and are caused by changing conditions of operation of the battery.

Table 2. Joint SOC–SOH Estimation Performance of RA-DHTCN

State	MAE (%)	RMSE (%)	MAPE (%)	R^2
SOC	0.92	1.14	1.28	0.984
SOH	0.70	0.84	0.76	0.978

Received: 10-07-2026 | Accepted: 20-08-2026 | Published: 26-08-2026 | www.ajmimc.com

Table 2 summarizes the quantitative performance of the proposed dual-head architecture for simultaneous SOC and SOH estimation. Lower MAE, RMSE, and MAPE values indicate lower prediction errors, while the higher value demonstrates strong agreement between predicted and reference battery states.

4.3 Resource-Constrained Model Optimization

Resource-constrained optimization phase analyzes the impact of structured pruning, knowledge distillation, and INT8 quantization on the trained RA-DHTCN. The trained network is gradually modified into a compact one through removing redundant network structures, knowledge transfer, and reduction of numerical precision. Parameter number, model size, memory usage, and computation time of this process are examined along with the performance of SOC and SOH prediction tasks.

Table 3. Resource Optimization Performance

Model Configuration	Parameters (K)	Model Size (MB)	RAM Usage (MB)	Inference Time (ms)
Original RA-DHTCN	428.0	1.72	6.40	12.40
Pruned RA-DHTCN	214.0	0.86	3.80	7.10
Distilled RA-DHTCN	96.5	0.39	2.10	4.30
INT8 RA-DHTCN	96.5	0.10	0.58	2.60

Table 3 gives an overview of the computational properties of the initial and optimized RA-DHTCN models. An effective optimization algorithm needs to be able to lower the number of parameters, size of the model, memory requirement, and processing time. Hence, Table 4 will prove whether the proposed compression approach can convert the initial neural network into a viable one.

4.4 Embedded Real-Time Deployment Performance

RA-DHTCN is implemented on resource-constrained embedded hardware for the real-time SOC and SOH estimation. The test will be carried out based on inference time, memory utilization, model size, and real-time processing. This will show the impact that pruning, knowledge distillation, and INT8 quantization have on the computational demands of battery state estimation.

4.5 Comparative Performance Analysis

The final subsection compares the proposed RA-DHTCN with conventional temporal models such as LSTM, GRU, and TCN. The comparison considers SOC and SOH estimation accuracy together with model complexity and inference requirements. This analysis determines whether the proposed architecture provides a favorable balance between prediction capability and computational efficiency.

Table 4. Comparative Performance of SOC/SOH Estimation Models

Model	SOC MAE (%)	SOC RMSE (%)	SOH MAE (%)	SOH RMSE (%)	Parameters (K)	Inference Time (ms)
LSTM	1.86	2.34	1.52	1.98	512.0	18.60
GRU	1.64	2.05	1.38	1.76	398.0	15.20
TCN	1.21	1.58	1.02	1.34	356.0	9.80
RA-DHTCN (Proposed, INT8-optimized)	0.90	1.10	0.70	0.84	96.5	2.60

Table 4 gives the final comparison between the proposed RA-DHTCN model and the existing temporal neural network models. This comparison takes into account prediction error as well as computational complexity, thus enabling us to judge the proposed resource-aware architecture from both the efficiency and deployment standpoints. The values to be entered in the table will be the actual results obtained experimentally.

4.6 Discussion

This RA-DHTCN architecture can be seen as an overall scheme to simultaneously estimate SOC and SOH of batteries in real-time under embedded system resource limitations. The use of time convolution and dilated feature extraction in RA-DHTCN is suitable for handling dynamic properties of batteries, and the dual heads allow the model to jointly predict both states based on a shared representation. The multitask learning algorithm ensures that there is consistency between SOC and SOH predictions. Additionally, structured pruning, knowledge distillation, and INT8 quantization techniques ensure reduction in terms of computational complexity and memory. However, further verification of RA-DHTCN using larger and independent batteries is still needed.

5. Conclusion and Future works

The RA-DHTCN model was developed in this research paper in order to provide joint SOC and SOH estimation using an embedded BMS under resource constraints. This framework uses battery data preprocessing, temporal sequence generation, dilated convolutional network with reduced model capacity, residual feature learning, and dual-head learning for multi-task learning for battery charge state and health properties. In addition, structured pruning, knowledge distillation, and INT8 quantization are used in order to reduce model capacity, memory footprint, and latency for faster on-device estimations. Future research will be devoted to the validation of this framework using additional battery data, different battery types, adaptive online learning for battery aging, hardware optimization for ultra-low power consumption, and quantization techniques.

References

[1] S. Giazitzis, J. Ferrari, A. J. Woss, S. Badha, F. Rosetti, and E. Ogliari, "Embedded strategy for battery module states estimation using tiny machine learning models," *Journal of Energy Storage*, vol. 152, p. 120673, 2026.
[2] S. Kang, Y. Lee, G. Jang, and S. Lee, "Lightweight 1D-CNN-based battery state-of-charge estimation and hardware development," *Electronics*, vol. 15, no. 3, p. 704, 2026.
[3] S. Giazitzis, J. Ferrari, A. J. Woss, S. Badha, F. Rosetti, and E. Ogliari, "Embedded strategy for battery module states estimation using tiny machine learning models," *Journal of Energy Storage*, vol. 152, p. 120673, 2026.
[4] J. Chaoraingern and A. Numsomran, "Embedded sensor data fusion and tinyml for real-time remaining useful life estimation of uav li polymer batteries," *Sensors*, vol. 25, no. 12, p. 3810, 2025.
[5] Y. Mazzi, H. Ben Sassi, A. Gaga, and F. Errahimi, "State of charge estimation of an electric vehicle's battery using tiny neural network embedded on small microcontroller units," *International Journal of Energy Research*, vol. 46, no. 6, pp. 8102–8119, 2022.
[6] R. Carrera, L. Quiroz, C. Guevara, and P. Acosta-Vargas, "State-of-charge estimation of medium-and high-voltage batteries using LSTM neural networks optimized with genetic algorithms," *Sensors*, vol. 25, no. 15, p. 4632, 2025.
[7] J. Zhao, D. Li, Y. Li, D. Shi, J. Nan, and A. F. Burke, "Battery state of health estimation under fast charging via deep transfer learning," *Iscience*, vol. 28, no. 5, 2025.
[8] C. S. Ozkan and M. Ozkan, "State of mission: Battery management with neural networks and electrochemical AI," *Iscience*, vol. 28, no. 10, 2025.
[9] M. K. Khan, M. Abou Houran, K. Kauhaniemi, M. H. Zafar, M. Mansoor, and S. Rashid, "Efficient state of charge estimation of lithium-ion batteries in electric vehicles using evolutionary intelligence-assisted GLA–CNN–Bi-LSTM deep learning model," *Heliyon*, vol. 10, no. 15, 2024.
[10] J. Wang, S. Cheng, and L. Zhang, "Deep Fusion Modeling for Lithium Batteries SOC-SOH Joint Estimation," *Mathematics*, vol. 14, no. 14, p. 2505, 2026.
[11] M. Zhang, X. Dai, Q. Zeng, Y. Tian, and X. Xu, "SOC and SOH Joint Estimation of Lithium-Ion Batteries Under Dynamic Current Rates Based on Machine Learning," *Symmetry*, vol. 18, no. 4, p. 623, 2026.
[12] J. Yao, S. Neupert, and J. Kowal, "Cross-stitch networks for joint state of charge and state of health online estimation of lithium-ion batteries," *Batteries*, vol. 10, no. 6, p. 171, 2024.
[13] L. Wu, C. Chen, Z. Li, Z. Chen, and H. Li, "The joint estimation of SOC-SOH for lithium-ion batteries based on BiLSTM-SA," *Electronics*, vol. 14, no. 1, p. 97, 2024.
[14] Y. Liu, D. Zhang, F. Wang, T. Huang, Y. Yu, and F. Sun, "A scalable joint estimation algorithm for SOC and SOH of all individual cells within the battery pack and its HIL implementation," *World Electric Vehicle Journal*, vol. 15, no. 6, p. 236, 2024.
[15] "EV Battery Charging & Thermal Runaway Dataset." Accessed: Aug. 12, 2026. [Online]. Available: https://www.kaggle.com/datasets/zara2099/ev-battery-charging-and-thermal-runaway-dataset?utm_source=chatgpt.com