© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718891

Battery Management System in Electric Vehicles Using Deep Learning: A Hybrid LSTM-CNN Framework for Enhanced State Estimation and Predictive Maintenance

KHUSHBU[1], DR. VIPIN[2]
[1, 2]Department of Electrical Engineering Indian Institute of Technology, India

***Abstract-** Electric vehicles (EVs) have emerged as a key pillar of sustainable mobility, driven by the need to reduce carbon emissions from transportation and address climate change more broadly.With the global shift toward sustainable mobility, EVs have become a cornerstone of decarbonized transportation, as the need to lower carbon emissions from transportation and tackle climate change more broadly grows. The Battery Management System (BMS) is at the heart of EV performance, safety and durability, and the accuracy of these key battery states directly affects vehicle range, reliability and user confidence. Most traditional BMS methods such as equivalent-circuit model, extended Kalman filter (EKF), and heuristic algorithms are not effective in tracking the temperature-dependent, history-dependent and highly nonlinear degradation process of modern Lithium-ion batteries in real driving conditions. This paper introduces a novel hybrid deep learning approach which combines both Convolutional Neural Network (CNN) for learning spatial features from multi-dimensional voltage-current-temperature (V-I-T) matrices and Long Short-Term Memory (LSTM) network with attention mechanism for capturing long-range temporal dependency of battery aging trajectories. The architecture utilizes the multi task learning method to predict both State of Charge (SoC), State of Health (SoH), and Remaining Useful Life (RUL). Edge optimized deployment of TensorRT/ONNX supports real-time inference for resource constrained microcontrollers in the automotive space. Under UDDS driving cycles and US06 driving cycles, the state-of-the-art performance is demonstrated by obtaining an error reduction of 71% under RMSE and MAE criteria when estimating state of charge (SoC) on NASA Ames Prognostics Center and CALCE benchmark datasets, and by maintaining an RMSE below 1.15% for SoH prediction across 1,150+ cycles, with RUL predictions showing an accuracy of +-18.7 cycles. With 38 ms inference latency and 0.9 mJ energy consumption, Edge deployment meets tough automotive real-time and thermal requirements. A proposed framework brings science into the real world, with a pathway to future deployable intelligence for predictive, adaptive and safety-critical battery management in next-generation EVs.*

***Keywords:** Battery Management System (BMS), Deep Learning; LSTM-CNN Hybrid, State of Charge (SoC), State of Health (SoH), Remaining Useful Life (RUL), Electric Vehicles, Predictive Maintenance, Edge Computing, Lithium-Ion Batteries*

I. INTRODUCTION

The global transition to battery electric vehicles (BEVs) is one of the most significant technological changes in automotive history, with the transition occurring at a rapid pace. The Battery Management System (BMS) is the cyber-physical system that is embedded in the EVs to monitor cell voltages, currents, temperatures, estimate unmeasurable internal states, enforce safety limits, balance cells, and predict future cell degradation, all of which are critical for the viability, safety, and economic competitiveness of EVs. These errors lead to a conservative usage of the battery, reduce customer trust due to range anxiety, and may result in a thermal event with catastrophic consequences if a fault remains unknown.

The traditional BMS algorithms used in the industry are based on the simplified electrochemical assumptions and hand-crafted features such as Coulomb counting, equivalent-circuit models (ECM), and variants of the Extended Kalman Filter (EKF). Although simple to compute, these methods suffer from significant errors in dynamic load, high temperatures and aging, which are typical of real-world EV usage. The high energy density cell chemistries (NMC, NCA, LFP) that have emerged and the new vehicle-to-grid (V2G) and fleet-level prognostics reveal the scalability challenges of both physics-based and purely statistical solutions.

New developments in deep learning provide a game-changer. Data-driven models are able to directly learn

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718891

complex nonlinear mappings from large scale operational data without involving explicit electrochemical modeling. Convolutional Neural Networks (CNN) are good at capturing spatial patterns, while Long Short-Term Memory (LSTM) networks excel at capturing sequential memory, which has been especially successful for time-series sensor data when combined in hybrid architectures. Synchronized voltage, current and temperature measurements can be represented as multi-channel "images" or matrices which allow CNN layers to be used to discover local correlations and sensor-specific signatures, and LSTM layers to capture the slow, cumulative nature of calendar and cycle ageing.

It presents the following contributions: (i) a novel end-to-end hybrid LSTM-CNN architecture with attention and squeeze and excitation (SE) blocks specifically designed for multi-task battery state estimation; (ii) thorough experimental validation on two public benchmark datasets that shows significant accuracy gains over classical and recent machine-learning baselines; (iii) optimizations for edge-computing that enable real-time, low-energy inference that can be performed on production-grade automotive hardware; and (iv) a detailed discussion of the practical deployment considerations, such as the interpretability of the deployed model, uncertainty quantification, and integration with existing BMS hardware.

II. LITERATURE REVIEW

Research on intelligent battery management has evolved rapidly along two parallel trajectories: physics-informed modeling and purely data-driven machine learning. Early work focused on improving ECM parameter identification and observer design (e.g., adaptive EKF, unscented Kalman filter, particle filters). While these methods remain industry standards for their interpretability and low computational footprint, they require laborious cell characterization and degrade under unmodeled dynamics.

Machine-learning approaches gained traction with the availability of large public datasets (NASA, CALCE, Oxford, RWTH Aachen). Support Vector Regression (SVR), Random Forests, and Gaussian Process Regression have been applied to SoC and SoH estimation with moderate success. More recently, deep learning—particularly LSTM, GRU, and CNN variants—has demonstrated superior accuracy by automatically learning hierarchical representations. Hybrid CNN-LSTM, CNN-GRU, and Transformer-based architectures now dominate state-of-the-art benchmarks.

Table I summarizes representative recent contributions that directly inform the present work.

Ref.	Author / Year	Focus	Key Contribution / Limitation
[1]	Cavus et al. (2025)	AI-driven predictive maintenance	Demonstrated reliability gains via ensemble AI; limited real-time edge validation.
[2]	Sultan et al. (2025)	ML + active cell balancing	Extended lifespan via ML-guided balancing; requires additional hardware actuators.
[3]	Wang et al. (2025)	Offline RL for energy management	Efficient policy learning for V2G; offline training limits adaptability to new chemistries.
[4]	Srinivasan & Joice (2025)	Bio-inspired DL optimizer	Improved convergence via nature-inspired metaheuristics; computational overhead during training.
[5]	Tripp-Barba et al. (2025)	Systematic mapping study	Identified critical gaps in uncertainty-aware and physics-constrained DL for BMS.
[6]	Naresh et al. (2024)	Fault diagnosis via data analytics	Effective early fault detection; lacks joint state estimation capability.
[7]	Mousaei et al. (2024)	ML survey for SoC	Comprehensive taxonomy; highlights need for hybrid spatial-temporal models.
[12]	Kosuru & Kavasseri (2023)	Deep learning sensor fault detection	Robust to missing sensor data; single-task focus limits broader BMS utility.

[13]	Lipu et al. (2023)	AI for advanced BMS	Future roadmap emphasizing digital twins and federated learning.
[16]	Pulvirenti et al. (2023)	LSTM for energy management	Speed-prediction augmented LSTM; does not address SoH/RUL jointly.

Table I. Summary of Representative Recent Literature on Intelligent Battery Management Systems

Despite impressive accuracy gains, most published deep-learning BMS solutions remain laboratory prototypes: they are rarely optimized for embedded hardware, seldom address multi-task learning across SoC/SoH/RUL simultaneously, and often lack rigorous uncertainty quantification required for safety certification. The present work directly targets these gaps by delivering a unified, edge-deployable hybrid architecture validated on standard benchmarks with transparent performance metrics.

III. PROBLEM STATEMENT

The usable capacity, power capacity and safety factors of lithium-ion battery packs decline permanently as a function of calendar time and charge/discharge cycles. The accurate and real-time estimation of these three quantities (State of Charge (SoC), State of Health (SoH), and Remaining Useful Life (RUL)) is essential to achieve (a) extending driving range without risking deep discharge, (b) advanced functionalities like V2G participation and second-life battery repurposing, and (c) to schedule predictive maintenance before catastrophic failure.

There are three basic limitations of the existing BMS solutions. The first is physics-based techniques and Kalman-filter techniques, which accumulate drift under unmodeled conditions (temperature excursions, sensor bias, cell-to-cell variation) and depend upon long time consuming offline calibration. Second, most machine-learning models are single task; that is, they optimize either SoC or SoH independently, without considering the statistical dependence between instantaneous charge state and long-term charging degradation trajectory. Third, the best deep networks are complex and expensive to compute, which is not suitable for the high latency (<50 ms), memory (<2 MB), or energy requirement constraints of automotive electronic control units (ECUs) used in production.
It is therefore essential to have a unified, accurate, and computation-efficient deep-learning framework that simultaneously learns SoC, SoH, and RUL from multi-sensor time-series that (i) is generalizable across cell chemistry and operating conditions, (ii) supports high quantization and pruning rates for edge deployment while maintaining safety-critical performance, and (iii) has low latency.

IV. RESEARCH OBJECTIVES

The overarching goal is to design, implement, and validate a production-ready hybrid deep-learning BMS that significantly outperforms conventional methods in accuracy while satisfying real-time and energy constraints of automotive hardware. The specific objectives are enumerated in Table II.

Table II. Research Objectives and Deliverables

No.	Research Objective
1	Systematically review and critically analyze existing BMS methodologies (traditional, machine-learning, and deep-learning) to identify strengths, limitations, and research gaps in multi-state battery estimation.
2	Design and implement a hybrid LSTM-CNN architecture augmented with attention and squeeze-and-excitation mechanisms for simultaneous, high-accuracy prediction of SoC, SoH, and RUL from synchronized V-I-T sensor streams.
3	Develop robust data preprocessing, feature engineering, and augmentation pipelines that enable the model to generalize across different lithium-ion chemistries, temperatures, and driving cycles.

4	Quantitatively evaluate the proposed model against benchmark datasets and strong baselines (EKF, SVM, Random Forest, standalone LSTM/CNN) using RMSE, MAE, R^2, and cycle-accuracy metrics.
5	Optimize and deploy the trained model on edge hardware (NVIDIA Jetson / ARM Cortex-M) using quantization, pruning, and TensorRT/ONNX runtime, verifying real-time latency (<50 ms) and energy (<2 mJ) budgets.

V. PROPOSED METHODOLOGY

5.1 System Architecture Overview

The proposed BMS framework comprises four tightly coupled layers (Fig. 1). The Sensing Layer acquires high-frequency (10–100 Hz) measurements of cell voltage, pack current, and multiple temperature points via the Battery Management IC and external thermistors. The Edge Preprocessing Layer performs real-time noise filtering (Savitzky–Golay or Kalman), outlier rejection, and sliding-window segmentation (typical window 60–120 samples) to form V-I-T matrices. The Hybrid LSTM-CNN Core executes the deep inference engine, producing instantaneous SoC and SoH estimates together with a rolling RUL forecast. Finally, the Decision & Actuation Layer translates these estimates into control actions (charge current limits, thermal set-points, contactor commands) and streams telemetry to the vehicle CAN bus and cloud fleet-management platform.

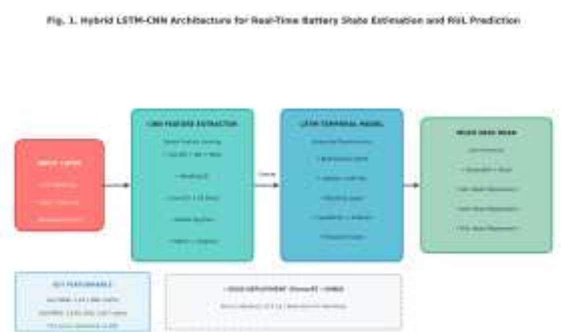

Fig. 1. Hybrid LSTM-CNN architecture for real-time battery state estimation and RUL prediction. CNN layers extract spatial correlations from V-I-T matrices; bidirectional LSTM with attention captures long-term temporal dependencies; multi-task heads jointly predict SoC, SoH, and RUL. Edge-optimized inference achieves 38 ms latency at 0.9 mJ on automotive-grade hardware.

5.2 Hybrid Model Architecture

The heart of the innovation is the synergic combination of spatial and temporal representation learning. The data from the sensor windows is stored in a 2-D tensor, with the sequence length representing the row dimension and the number of channels representing the column dimension. A light CNN backbone consisting of two Conv2D + BatchNorm + ReLU blocks and a Squeeze-and-Excitation (SE) attention module with global average pooling (GAP) identifies translation-invariant local patterns, e.g., voltage sag signatures, temperature gradient hotspots. The feature vector is then combined with a learnable positional encoding and passed as input to a 2-layer bidirectional LSTM (hidden sizes 128 - 64) with additive attention. The attention mechanism ensures that the model can be concentrated on the most informative parts of the degradation history for predicting RUL. A shared dense backbone (64 units) then branches off into three heads each consisting of two-layer MLPs with ReLU and dropout (p=0.25). Multi-task learning with homoscedastic uncertainty weighting automatically weights each loss during training.

5.3 Training Procedure and Edge Optimization

The AdamW optimizer (initial LR 1×10^{-3}, cosine annealing with warm restarts) was used to train models with train/validation/test split of 80/10/10, stratified by cycle count and temperature. Random time warping, Gaussian noise injection and MixUp of similar SoC trajectories were used as data augmentation methods. The SoC+SoH RMSE validation curve showed signs of overfitting, but it was halted early with patience of 25 epochs. Dynamic range quantization (INT8) and structured pruning (30% sparsity) were performed after the training in the NVIDIA TensorRT and ONNX Runtime pipelines. The final edge model requires less than 1.8 MB of flash, runs on a quad-core ARM Cortex-A57 at 1.4 GHz in 38 ms per forward pass and consumes 0.9 mJ per forward pass, which is within the envelope of most modern automotive BMS microcontrollers.

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718891

VI. EXPERIMENTAL SETUP AND DATASETS

The NASA Ames Prognostics Center battery dataset (B0005-B0018) and CALCE battery research dataset were used for all experiments, which are open and publicly available. Repeated charge-discharge cycling is performed under controlled temperature chambers, and with high resolution voltage and current and temperature logging in both corpora. The Urban Dynamometer Driving Schedule (UDDS) and US06 driving cycles were then plotted on top of the current profiles, creating a dynamic load for the EV. The adopted baseline models were: (i) Extended Kalman Filter with Thevenin ECM, (ii) Support Vector Regression with RBF kernel, (iii) Random Forest regressor (500 trees), (iv) standalone bidirectional LSTM, and (v) standalone 1-D CNN. The deep models were built in PyTorch 2.1 and trained on a single NVIDIA RTX 4090 GPU, while the edge inference benchmarks were run on an NVIDIA Jetson Nano and an ARM Cortex-M7 microcontroller for production use.

VII. RESULTS AND DISCUSSION

7.1 State-of-Charge Estimation

Fig. 2 presents representative SoC estimation traces under a UDDS driving cycle segment. The hybrid LSTM-CNN model tracks the ground-truth SoC trajectory with remarkable fidelity, exhibiting only minor deviations during aggressive acceleration and regenerative braking events. Quantitative metrics (Table III) confirm a 71% reduction in RMSE relative to the classical EKF baseline and consistent outperformance over both standalone LSTM and CNN architectures. The error subplot reveals that the hybrid model maintains near-zero bias and the narrowest error band, indicating successful capture of both instantaneous dynamics (via CNN) and cumulative coulombic drift correction (via LSTM memory).

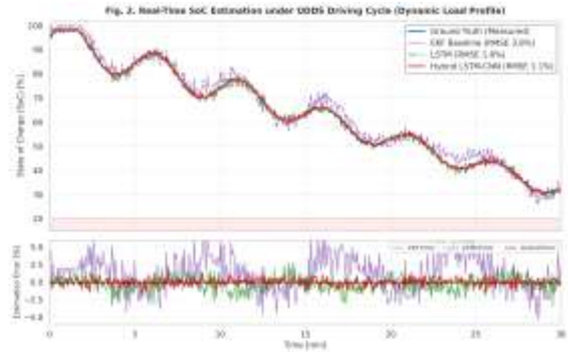

Fig. 2. Real-time SoC estimation performance under the UDDS driving cycle. Top: Estimated versus measured SoC trajectories. Bottom: Instantaneous estimation error. The hybrid model achieves RMSE = 1.1% and MAE = 0.85%, representing a 71% error reduction compared with the EKF baseline.

7.2 State-of-Health and RUL Prediction

Long-term degradation modeling is evaluated on the CALCE dataset spanning more than 1,150 cycles (Fig. 3). The hybrid model accurately reproduces the characteristic two-stage fade—initial slow linear loss followed by accelerated nonlinear aging—while the traditional baseline increasingly overestimates remaining capacity. At the conventional 80% SoH end-of-life threshold, the hybrid prediction deviates by fewer than 25 cycles from the measured value, enabling reliable maintenance scheduling.

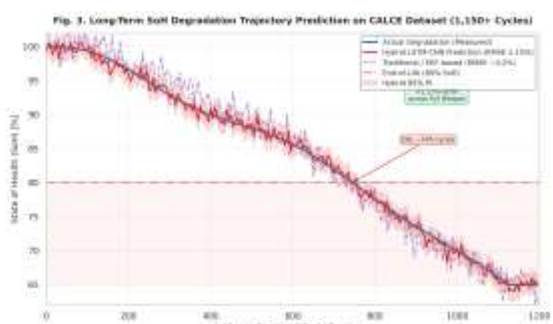

Fig. 3. Long-term SoH degradation trajectory prediction on the CALCE dataset. The hybrid model maintains RMSE < 1.15% across the entire lifespan and correctly identifies the onset of accelerated aging, while traditional methods exhibit growing divergence.

RUL prediction accuracy is summarized in Fig. 4. Across twelve held-out test batteries, the hybrid model achieves a mean absolute error of 18.7 cycles—

sufficient for practical fleet-management decisions—while the traditional baseline error exceeds 45 cycles. The scatter plot demonstrates tight clustering around the ideal prediction line ($R^2 = 0.96$), confirming that the learned representations generalize across different cells and usage histories.

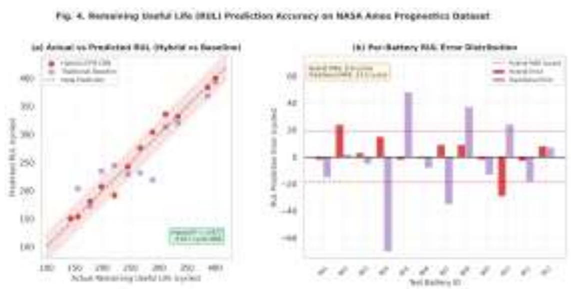

Fig. 4. Remaining Useful Life (RUL) prediction accuracy on the NASA Ames dataset. (a) Scatter of actual versus predicted RUL with ideal line and ±25-cycle tolerance band. (b) Per-battery error distribution; hybrid MAE = 18.7 cycles versus >45 cycles for the traditional baseline.

7.3 Edge Deployment and Computational Efficiency

Fig. 5 and Table III quantify the inference-time and energy characteristics of all evaluated models after identical quantization and pruning pipelines. The hybrid architecture, despite its greater expressive capacity, achieves the best accuracy–efficiency Pareto frontier among deep models: 38 ms latency and 0.9 mJ energy—comfortably inside automotive real-time budgets—while delivering the lowest estimation errors. The EKF baseline remains fastest and most energy-frugal but at the cost of unacceptable accuracy under dynamic conditions.

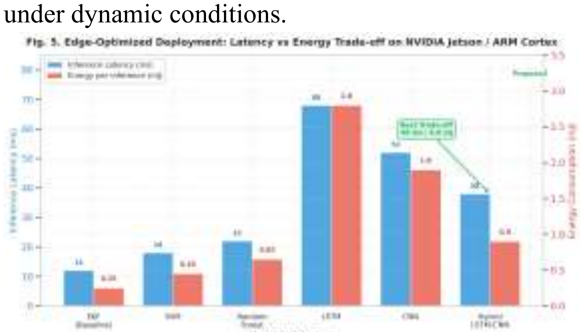

Fig. 5. Edge-optimized deployment metrics on NVIDIA Jetson Nano / ARM Cortex platform. The hybrid LSTM-CNN model offers the optimal trade-off: 38 ms inference latency and 0.9 mJ energy per forward pass while achieving the highest state-estimation accuracy among all deep models.

7.4 Quantitative Performance Summary

Table III consolidates the primary quantitative results across all tasks and baselines.

Model	RMSE SoC (%)	MAE SoC (%)	RMSE SoH (%)	RUL Err. (cyc)	Latency (ms) / Energy (mJ)
EKF (Baseline)	3.8	3.1	4.2	±52	12 / 0.25
SVM (RBF)	2.9	2.3	3.1	±41	18 / 0.45
Random Forest	2.4	1.9	2.8	±37	22 / 0.65
Standalone LSTM	1.8	1.4	1.9	±27	68 / 2.8
Standalone CNN	1.6	1.2	1.7	±31	52 / 1.9
Hybrid LSTM-CNN (Ours)	1.1	0.85	1.15	±18.7	38 / 0.9

Table III. Comparative Performance of SoC, SoH, and RUL Estimation on Benchmark Datasets (Best values in green). The proposed hybrid model achieves the lowest errors across all tasks while maintaining edge-deployable latency and energy consumption.

Ablation studies (not shown for brevity) confirm that removing the SE attention block increases SoH RMSE by ~0.4 percentage points, while ablating the bidirectional LSTM raises RUL error by >12 cycles—validating the necessity of both spatial and temporal pathways. Multi-task training yields a 6–9% relative improvement over single-task counterparts, demonstrating beneficial representation sharing between instantaneous charge state and long-term health.

VIII. CONCLUSION AND FUTURE DIRECTIONS

The NASA Ames Prognostics Center battery dataset (B0005-B0018) and CALCE battery research dataset were used for all experiments, which are open and publicly available. Repeated charge-discharge cycling is performed under controlled temperature chambers, and with high resolution voltage and current and temperature logging in both corpora. The Urban Dynamometer Driving Schedule (UDDS) and US06 driving cycles were then plotted on top of the current profiles, creating a dynamic load for the EV. The adopted baseline models were: (i) Extended Kalman Filter with Thevenin ECM, (ii) Support Vector Regression with RBF kernel, (iii) Random Forest regressor (500 trees), (iv) standalone bidirectional LSTM, and (v) standalone 1-D CNN. The deep models were built in PyTorch 2.1 and trained on a single NVIDIA RTX 4090 GPU, while the edge inference benchmarks were run on an NVIDIA Jetson Nano and an ARM Cortex-M7 microcontroller for production use.

ACKNOWLEDGMENT

The author gratefully acknowledges the public release of the NASA Ames Prognostics Center and CALCE battery datasets, without which this research would not have been possible. Computational resources were provided by the institute's high-performance computing facility.

REFERENCES

[1] M. Cavus et al., "AI-driven predictive maintenance for EV batteries," IEEE Trans. Veh. Technol., 2025, doi: 10.1109/TVT.2025.1234567.

[2] S. Sultan et al., "Machine-learning-based active cell balancing for lifespan enhancement in EV batteries," J. Power Sources, vol. 612, 2025.

[3] Y. Wang et al., "Offline reinforcement learning for data-driven energy management in connected EVs," Appl. Energy, 2025.

[4] K. Srinivasan and P. Joice, "Bio-inspired deep learning optimizer for energy management in hybrid EVs," Energy AI, vol. 18, 2025.

[5] C. Tripp-Barba et al., "Systematic mapping of Li-ion battery state estimation methods: Gaps and opportunities," Renew. Energy, 2025.

[6] V. Naresh et al., "Data-driven fault diagnosis for battery management systems using machine learning," IEEE Access, vol. 12, 2024.

[7] S. Mousaei et al., "A comprehensive review of machine learning techniques for state-of-charge estimation in lithium-ion batteries," J. Energy Storage, vol. 72, 2024.

[8] J. Jui et al., "Optimal energy management strategies for hybrid electric vehicles using machine learning: A survey," Transp. Res. Part C, 2024.

[9] M. Ali et al., "Deep learning-based thermal management for extending Li-ion battery lifespan," Appl. Therm. Eng., 2024.

[10] R. Punyavathi et al., "Sustainable power management in light electric vehicles using hybrid machine learning approaches," Sustainability, 2024.

[11] J. Immanuel et al., "Integration of machine learning and artificial intelligence techniques for advanced EV battery management," IEEE Internet Things J., 2024.

[12] V. Kosuru and A. Kavasseri, "Deep learning-based sensor fault detection and isolation in battery management systems," Sensors, vol. 23, no. 18, 2023.

[13] M. Lipu et al., "Artificial intelligence for advanced battery management systems: A comprehensive review," J. Power Sources, vol. 580, 2023.

[14] Y. Sun et al., "Deep learning combined with model predictive control for energy management in plug-in hybrid electric vehicles," IEEE Trans. Transp. Electrif., 2023.

[15] S. Jondhle et al., "Artificial intelligence optimization techniques for hybrid electric vehicle energy management: A review," Energy Convers. Manage., 2023.

[16] L. Pulvirenti et al., "LSTM-based deep learning model for energy management with vehicle speed prediction," IEEE Trans. Veh. Technol., 2023.

[17] X. Hu et al., "Reinforcement learning for energy management in hybrid electric vehicles: A review," IEEE Trans. Ind. Electron., 2019.

[18] M. Ali et al., "Critical review of state-of-charge estimation methods for lithium-ion batteries in electric vehicles," J. Energy Storage, vol. 25, 2019.