



(REVIEW ARTICLE)



EDGE AI-BASED REAL-TIME BATTERY MANAGEMENT SYSTEMS: TECHNOLOGY TRENDS AND DESIGN FRAMEWORK

JongMyoung Kim *

Department of Artificial Intelligence and Big Data, Sehan University.

* Corresponding Author

International Journal of Science and Research Archive, 2026, 21(01), 644–651

Publication history: Received on 21 August 2026; revised on 24 September 2026; accepted on 26 September 2026

Article DOI: <https://doi.org/10.30574/ijrsra.2026.21.1.1871>

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Battery management systems (BMSs) increasingly require low-latency, reliable, and intelligent processing capabilities to support the rapid expansion of electric vehicles and stationary energy storage systems. Conventional centralized and cloud-dependent BMS architectures can provide advanced analytics but may face limitations associated with communication latency, network dependence, data transmission overhead, and privacy. Edge artificial intelligence (Edge AI) provides an alternative approach by performing data processing and intelligent inference close to the battery system. This review examines major technology trends in real-time BMSs and proposes a reference design framework for Edge AI-based battery management. Recent developments in battery sensing, artificial intelligence-based state estimation, decentralized computing, lightweight machine learning, digital twins, communication, and cybersecurity are reviewed. Based on these trends, a layered framework is proposed in which battery sensors and embedded controllers acquire operating data, edge processors perform preprocessing and intelligent analysis, and cloud platforms provide long-term data management, model updating, and fleet-level analytics. The framework emphasizes local state estimation, anomaly detection, predictive maintenance, secure edge-cloud communication, and optional digital-twin integration. The review indicates that Edge AI can reduce dependence on continuous cloud connectivity and enable faster decision-making, although practical deployment remains constrained by limited embedded resources, model complexity, cybersecurity risks, heterogeneous battery platforms, and interoperability issues. Future BMS development is expected to combine lightweight AI, adaptive edge-cloud collaboration, explainable decision support, digital twins, and security-by-design principles.

Keywords: Battery management system; Edge AI; Real-time monitoring; Edge computing; Digital twin; State of health

1 INTRODUCTION

Lithium-ion batteries have become a core energy-storage technology for electric vehicles (EVs), renewable-energy systems, portable electronics, and distributed energy-storage applications. Their high energy density, relatively long cycle life, and decreasing cost have supported rapid adoption across multiple industries. However, the performance and safety of lithium-ion batteries are highly dependent on operating conditions such as temperature, charging current, depth of discharge, cell imbalance, and accumulated degradation. Battery management systems therefore play a critical role in monitoring, controlling, and protecting battery packs throughout their operational life.

A conventional BMS typically measures voltage, current, and temperature and uses these measurements to perform functions such as state-of-charge (SOC) estimation, state-of-health (SOH) assessment, cell balancing, thermal management, fault detection, and charging control. As battery packs become larger and more complex, the amount of data generated by the BMS also increases. At the same time, modern applications require increasingly accurate state estimation and faster responses to abnormal operating conditions.

Artificial intelligence (AI) and machine-learning techniques have been introduced into BMSs to improve state estimation and battery health prediction. Deep neural networks have demonstrated the ability to estimate SOC directly from voltage, current, and temperature data [1], while broader data-driven approaches have been investigated for SOC and SOH estimation across different battery operating conditions [2]. Machine learning has also been applied to battery cycle-life prediction and charging optimization [3,4].

Despite these advances, cloud-centered AI architectures are not always suitable for time-sensitive battery applications. Transmitting large volumes of raw battery data to remote servers may introduce communication delays, increase bandwidth usage, and create dependence on network availability. In safety-critical environments, even a short delay between sensing and decision-making can reduce the effectiveness of fault detection or protective control.

Edge AI addresses this limitation by executing AI models directly on embedded controllers, edge processors, or gateways located near the battery system. Instead of continuously transmitting all raw sensor data to the cloud, an Edge AI-enabled BMS can process measurements locally and transfer only selected information, summarized features, alerts, or model updates. This architecture can reduce latency, improve privacy, and maintain core BMS functions even when cloud connectivity is unavailable.

The development of digital-twin technology further expands the potential role of Edge AI in battery management. A battery digital twin creates a continuously updated virtual representation of a physical battery or battery pack. Previous studies have demonstrated cloud-based battery digital twins capable of supporting online SOC and SOH estimation [5], virtual battery experiments [6], and aging analysis [7]. Recent research has also investigated the integration of physical models and data-driven techniques for more adaptive battery digital twins [8].

These developments suggest that next-generation BMS architectures will not rely exclusively on either embedded control or cloud computing. Instead, battery intelligence is likely to be distributed across battery sensors, local edge devices, vehicle or system controllers, and cloud platforms.

This review examines recent technological developments related to Edge AI-based real-time BMSs and presents a reference design framework for their future development. The main contributions of this paper are threefold. First, major trends in battery sensing, AI-based state analysis, edge computing, and decentralized intelligence are summarized. Second, a layered Edge AI-BMS design framework is proposed to organize sensing, local processing, AI inference, communication, cloud services, and digital-twin functions. Third, major deployment challenges and future research directions are discussed, including computational efficiency, cybersecurity, interoperability, explainability, and edge-cloud collaboration.

2 EDGE AI TRENDS IN REAL-TIME BATTERY MANAGEMENT SYSTEMS

2.1 Battery sensing and real-time state monitoring

Accurate battery management begins with reliable sensing. A typical BMS continuously collects cell voltage, pack voltage, current, and temperature information. Advanced BMS architectures may additionally use impedance, pressure, strain, gas, or other sensing information depending on the safety and diagnostic requirements of the application.

Traditional battery monitoring primarily relies on predefined thresholds and mathematical estimation methods. These approaches remain important because they are computationally efficient and can operate on low-cost embedded controllers. However, fixed threshold-based monitoring has limited ability to identify complex degradation patterns or interactions among multiple variables.

The increasing availability of higher-resolution battery data has created opportunities for data-driven state estimation. For example, deep neural networks have been used to estimate SOC using voltage, current, and temperature measurements without requiring a complete electrochemical model [1]. Data-driven approaches can learn nonlinear relationships that may be difficult to describe using fixed analytical equations.

Real-time monitoring becomes particularly important when battery operating conditions change rapidly. EV acceleration, regenerative braking, rapid charging, and variations in ambient temperature can cause significant changes

in electrical and thermal conditions over short periods. An intelligent BMS must therefore distinguish normal dynamic behavior from abnormal behavior without introducing excessive computational delay.

Edge processing allows the first stage of battery-data analysis to be located close to the sensing layer. Filtering, normalization, feature extraction, threshold checking, and lightweight state estimation can be performed before information is transmitted to a higher-level controller or cloud platform. Such local processing reduces communication requirements and enables faster responses to critical battery events.

Wireless battery management systems can further reduce wiring complexity in large battery packs, but wireless communication introduces additional issues related to latency, interference, reliability, and security. Therefore, real-time battery sensing should be considered not only as a measurement problem but also as a distributed data-management problem.

2.2 AI-based SOC and SOH analytics

SOC and SOH are among the most important internal states estimated by a BMS. SOC represents the available charge relative to the battery's usable capacity, while SOH provides an indication of degradation relative to the original battery condition.

Traditional SOC estimation methods include Coulomb counting, open-circuit-voltage methods, equivalent-circuit models, and Kalman-filter-based techniques. These methods are widely used because their behavior can be modeled mathematically and their computational requirements are relatively predictable.

Machine-learning methods provide a complementary approach. Chemali et al. demonstrated deep neural-network-based SOC estimation using measurable battery variables [1]. Data-driven battery research has since expanded to include neural networks, support-vector machines, ensemble learning, recurrent neural networks, and other machine-learning approaches [2].

Machine learning has also been applied to long-term battery degradation. Severson et al. showed that early-cycle data can contain information relevant to predicting battery cycle life before significant capacity degradation becomes visible [3]. Such approaches illustrate how battery data can be used not only for present-state estimation but also for predictive maintenance and lifecycle management.

However, a model that performs well in a laboratory environment may not be directly suitable for an embedded BMS. Deep learning models can require significant memory, computational power, and energy consumption. Edge AI therefore requires a balance between predictive capability and resource efficiency.

Several general AI optimization techniques can support this objective. Quantization can reduce computational complexity by replacing high-precision operations with lower-precision integer operations [9]. Lightweight neural-network architectures demonstrate that carefully designed networks can achieve useful inference performance with lower computational requirements [10]. Although many lightweight architectures were originally developed for computer vision, their underlying principles of model compression, efficient operations, and reduced parameter counts are applicable to embedded battery analytics.

For Edge AI-based BMSs, model selection should therefore consider more than estimation accuracy. Inference latency, processor utilization, memory footprint, energy consumption, and robustness should also be treated as important design criteria.

2.3 Edge and decentralized intelligence

A centralized architecture places most intelligent processing in a single controller or cloud platform. This simplifies model management but can create bottlenecks when the number of battery cells, packs, vehicles, or energy-storage units increases.

Edge computing distributes computation closer to the data source. In a battery system, intelligence can be located at several levels, including cell-monitoring units, module controllers, pack controllers, vehicle gateways, or local energy-management servers.

A distributed architecture has several potential advantages. First, safety-related decisions can be made locally without waiting for cloud communication. Second, only necessary information needs to be transmitted, reducing network traffic. Third, local processing can improve privacy because raw operating data may remain on the device.

Edge and cloud computing should not necessarily be regarded as competing approaches. A hybrid architecture can divide responsibilities according to time sensitivity and computational requirements. Real-time fault detection and protective decisions can be executed at the edge, whereas long-term degradation analysis, fleet-level comparison, and large-scale model training can remain in the cloud.

Federated learning provides another possible mechanism for distributed battery intelligence. Federated learning enables multiple devices to participate in model training without directly centralizing all raw data [11]. In an EV application, individual vehicles could potentially train local models using operational battery data and share model updates rather than complete driving histories.

Such an approach could support fleet-level improvement while reducing the amount of sensitive information transferred to a central server. Nevertheless, practical implementation would require solutions for heterogeneous batteries, non-identically distributed operating data, communication constraints, and secure aggregation.

3 REFERENCE DESIGN FRAMEWORK FOR EDGE AI-BASED BMS

3.1 Layered Edge AI-BMS architecture

Based on current technology trends, an Edge AI-based BMS can be organized into five functional layers: the physical battery layer, sensing and acquisition layer, edge intelligence layer, communication and coordination layer, and cloud or digital-twin layer.

The physical battery layer consists of individual cells, modules, battery packs, power electronics, thermal-management components, and protection devices. These components represent the physical system that must be monitored and controlled.

The sensing and acquisition layer collects voltage, current, temperature, and other available battery signals. Initial signal conditioning, filtering, synchronization, and validity checking are performed at this stage. This layer should ensure that higher-level algorithms receive reliable input data.

The edge intelligence layer is the core of the proposed framework. It performs time-sensitive data processing close to the battery system. Functions may include SOC estimation, SOH screening, anomaly detection, thermal-risk assessment, fault classification, and short-term prediction. Because this layer operates on embedded hardware, algorithms should be selected according to both analytical performance and computational requirements.

The communication and coordination layer connects distributed battery controllers, edge gateways, vehicle networks, and cloud platforms. Depending on the application, communication may use CAN, Ethernet, wireless sensor networks, cellular communication, or other protocols.

The cloud and digital-twin layer performs computationally intensive or long-term functions. These can include historical data storage, fleet analytics, model training, model-version management, degradation analysis, and digital-twin simulation.

This layered architecture allows functions to be distributed according to their urgency. Protective control should remain close to the physical battery, whereas non-critical long-term analytics can be performed remotely.

A key design principle is graceful degradation. The battery system should remain safe even when cloud communication becomes temporarily unavailable. Therefore, core protection, state estimation, and fault-handling logic should not depend entirely on an external server.

3.2 Edge data processing and lightweight AI

The effectiveness of Edge AI depends strongly on how raw battery data are processed. Continuously transmitting every sensor sample may be unnecessary and inefficient. Instead, local processing can extract meaningful indicators.

Typical processing may include noise filtering, missing-data detection, normalization, statistical feature extraction, trend estimation, and event detection. Depending on available resources, the edge processor may run conventional algorithms and AI models in parallel.

For example, a conventional estimator can provide a baseline SOC value, while a machine-learning model can identify nonlinear patterns or compensate for model errors. Combining analytical and data-driven methods can improve robustness while limiting reliance on a single black-box model.

AI models deployed on edge hardware should also be optimized for resource constraints. Quantization, pruning, feature reduction, and compact neural architectures can reduce memory and computational requirements. Integer-based inference has been shown to significantly improve efficiency in constrained devices [9].

Model complexity should be adapted to the decision type. A safety-critical fast anomaly detector may use a simple lightweight classifier, while a more complex SOH prediction model may be executed less frequently or on a higher-level gateway.

Another important consideration is model updating. Battery characteristics change over time because of aging, environmental conditions, and usage patterns. Static AI models may gradually lose accuracy. An Edge AI framework should therefore provide mechanisms for periodic recalibration or model updates.

Updates can be generated centrally from aggregated battery data and distributed to edge devices. Alternatively, selected parameters may be updated locally. The update mechanism should include version control, rollback capability, authentication, and compatibility verification.

Explainability is also becoming increasingly relevant. Battery operators and engineers may require information about why an AI model classifies a particular condition as abnormal. Explainable AI approaches can provide information about influential variables and support human interpretation of model decisions [12].

3.3 Edge-cloud communication and cybersecurity

The introduction of network connectivity creates additional security requirements for BMSs. Battery operating data may contain information about vehicle usage, charging patterns, device condition, and energy consumption. Unauthorized modification of BMS data may also create safety risks.

Therefore, communication between edge devices and cloud services should include authentication, encryption, integrity verification, secure update mechanisms, and access control.

Not all data should be transmitted with the same priority. Safety alarms and critical state changes require immediate communication, whereas historical monitoring data can be transferred periodically or aggregated.

This distinction supports efficient bandwidth utilization. The edge node can classify information according to urgency and transmit only necessary data in real time.

Data integrity is particularly important in distributed battery architectures. If sensor data or historical records are modified, state analysis and maintenance decisions may become unreliable. Blockchain-based or cryptographic integrity mechanisms have been proposed as possible methods for verifying BMS data without storing all raw data directly on a distributed ledger [13].

However, security mechanisms also introduce computational and communication overhead. Edge BMS design must therefore balance security strength with latency, energy use, and hardware capability.

Security should be considered from the initial architecture stage rather than added after system deployment. A security-by-design approach should include device authentication, secure boot, encrypted communication, protected model storage, update verification, anomaly monitoring, and audit mechanisms.

4 APPLICATIONS, LIMITATIONS, AND FUTURE DIRECTIONS

4.1 Electric vehicles and energy-storage systems

EVs represent one of the most demanding BMS applications because battery operating conditions change continuously. High charge and discharge currents, wide temperature variations, regenerative braking, and fast charging can produce complex battery behavior.

Edge AI can support EV battery management by enabling local processing of high-frequency operating data. Potential applications include SOC estimation, SOH monitoring, early fault identification, thermal-state assessment, adaptive charging recommendations, and predictive maintenance.

The same architecture can be applied to stationary battery energy-storage systems. Large-scale storage installations may contain thousands of cells and multiple racks. Monitoring all raw data centrally can create significant communication and storage requirements.

Distributed edge processing can reduce this burden by analyzing data at the module or rack level and transmitting summarized health indicators or anomaly information to the supervisory system.

In both EV and stationary-storage applications, the objective of Edge AI is not necessarily to replace existing BMS algorithms. Instead, Edge AI can supplement established control and protection methods with more adaptive analysis capabilities.

4.2 Integration with digital twins

Digital twins provide a promising extension to Edge AI-based BMS architectures. A digital twin is a virtual representation that is synchronized with information from the physical battery.

Li et al. proposed a cloud battery management approach incorporating online SOC and SOH estimation within a digital-twin architecture [5]. Subsequent research has extended digital-battery-twin concepts to virtual experiments, aging analysis, and data-mechanism fusion [6-8].

Combining Edge AI and digital twins creates a division of responsibilities. The edge device can provide real-time sensing and fast local decisions, while the digital twin can maintain a more comprehensive representation of battery condition over the lifecycle.

For example, the edge BMS may detect an unusual temperature-voltage relationship and generate an event. The cloud-based digital twin can then compare this event with historical degradation trends, similar battery units, or simulated operating conditions.

This architecture may support condition-based maintenance and long-term asset management without requiring complex models to operate continuously on embedded hardware.

Digital twins can also assist model development. Virtual experiments may generate additional scenarios that are difficult or expensive to reproduce on physical batteries. However, the usefulness of a digital twin depends on synchronization quality, model fidelity, data quality, and computational resources.

Therefore, future Edge AI-BMS designs should treat digital twins as complementary analytical components rather than replacements for direct battery sensing and safety control.

4.3 Scalability, security, interoperability, and future research

Several barriers remain before Edge AI can be widely adopted in production BMSs.

The first challenge is computational limitation. Automotive-grade and industrial BMS controllers are designed for reliability, deterministic behavior, and low power consumption. They may not provide the computational resources available in high-performance cloud servers.

Consequently, AI models must be optimized for embedded deployment. Lightweight architectures, quantization, hardware acceleration, and selective execution will become increasingly important.

The second challenge is generalization. Battery behavior differs according to chemistry, manufacturer, cell format, operating environment, and aging condition. A model trained on one dataset may not perform equally well on another battery system.

Future research should therefore focus on transferable and adaptive models rather than only maximizing performance on a single dataset.

The third challenge is cybersecurity. Increasing connectivity expands the attack surface of battery systems. Secure communication, trusted model updates, integrity verification, and intrusion monitoring must be integrated into Edge AI architectures.

The fourth challenge is interoperability. Different manufacturers use different battery hardware, communication protocols, data structures, and control strategies. Without standardized interfaces, AI models and digital-twin services may be difficult to transfer across platforms.

Common data models and standardized interfaces could support broader adoption of intelligent BMS services.

The fifth challenge is explainability and trust. In safety-critical systems, an AI-generated decision should ideally provide sufficient information for validation by engineers. Explainable AI can help identify the measurements or trends responsible for a prediction and can improve diagnostic transparency [12].

Finally, future BMS architectures are likely to become increasingly collaborative. Local edge controllers can handle immediate safety and control tasks, while cloud platforms perform model training and fleet-level analytics. Federated learning may allow distributed systems to improve models without centralizing all raw data [11].

Digital twins can provide a lifecycle-level representation, while secure data-integrity mechanisms can improve trust in the information exchanged among system components.

The long-term direction is therefore not simply an "AI-based BMS" but a distributed intelligent battery-management ecosystem combining sensing, embedded intelligence, communication, digital twins, cloud analytics, explainability, and cybersecurity.

5 CONCLUSION

The increasing complexity of electric vehicles and energy-storage systems is changing the role of battery management systems. Conventional BMSs remain essential for measurement, protection, state estimation, and control, but the growing volume of battery data and the demand for faster, more adaptive decisions are creating opportunities for Edge AI.

This review examined major technology trends related to real-time BMSs and proposed a layered reference framework for Edge AI-based battery management. The framework distributes functions across physical battery components, sensing systems, edge intelligence, communication networks, and cloud or digital-twin services.

The primary advantage of Edge AI is its ability to process battery data close to the source. This can reduce communication latency, decrease dependence on continuous cloud connectivity, and support faster local responses. Cloud platforms remain valuable for long-term analytics, large-scale model training, fleet management, and digital-twin functions.

However, the transition toward Edge AI-enabled BMSs introduces important challenges. Embedded computing resources are limited, battery systems are heterogeneous, cybersecurity risks increase with connectivity, and AI models require validation across diverse operating conditions. Interoperability and explainability will also become increasingly important as intelligent BMS architectures become more complex.

Future battery management systems are expected to combine lightweight AI, adaptive edge-cloud coordination, secure communication, explainable decision support, digital twins, and standardized interfaces. Such integration can support the transition from conventional monitoring-oriented BMSs toward distributed and intelligent battery-management platforms capable of supporting safer, more reliable, and more sustainable energy-storage applications.

REFERENCES

- [1] Chemali E, Kollmeyer PJ, Preindl M, Emadi A. State-of-charge estimation of Li-ion batteries using deep neural networks: A machine learning approach. *Journal of Power Sources*. 2018;400:242-255. doi:10.1016/j.jpowsour.2018.06.104.
- [2] Ng MF, Zhao J, Yan Q, Conduit GJ, Seh ZW. Predicting the state of charge and health of batteries using data-driven machine learning. *Nature Machine Intelligence*. 2020;2:161-170.
- [3] Severson KA, Attia PM, Jin N, Perkins N, Jiang B, Yang Z, et al. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*. 2019;4:383-391. doi:10.1038/s41560-019-0356-8.
- [4] Attia PM, Grover A, Jin N, Severson KA, Markov TM, Liao YH, et al. Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature*. 2020;578:397-402. doi:10.1038/s41586-020-1994-5.
- [5] Li W, Rentemeister M, Badede J, Jöst D, Schulte D, Sauer DU. Digital twin for battery systems: Cloud battery management system with online state-of-charge and state-of-health estimation. *Journal of Energy Storage*. 2020;30:101557. doi:10.1016/j.est.2020.101557.
- [6] Schmitt J, Horstkötter I, Bäker B. State-of-health estimation by virtual experiments using recurrent decoder-encoder based lithium-ion digital battery twins trained on unstructured battery data. *Journal of Energy Storage*. 2023;58:106335. doi:10.1016/j.est.2022.106335.

- [7] Guo W, Li Y, Sun Z, Vilsen SB, Stroe DI. A digital twin to quantitatively understand aging mechanisms coupled effects of NMC battery using dynamic aging profiles. *Energy Storage Materials*. 2023;63:102965.
- [8] Lyu C, Xu S, Li J, Pecht M. Digital twin modeling method for lithium-ion batteries based on data-mechanism fusion driving. *Green Energy and Intelligent Transportation*. 2024;3(5):100162. doi:10.1016/j.geits.2024.100162.
- [9] Jacob B, Kligys S, Chen B, Zhu M, Tang M, Howard A, et al. Quantization and training of neural networks for efficient integer-arithmetic-only inference. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018:2704-2713.
- [10] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018:4510-4520.
- [11] McMahan HB, Moore E, Ramage D, Hampson S, Agüera y Arcas B. Communication-efficient learning of deep networks from decentralized data. *Proceedings of Machine Learning Research*. 2017;54:1273-1282.
- [12] Kim JM. Explainable artificial intelligence in battery management systems: A comprehensive review. *International Journal of Science and Technology Research Archive*. 2025;8(2):14-26. doi:10.53771/ijstra.2025.8.2.0034.
- [13] Kim JM. Blockchain-based architecture for ensuring BMS data integrity. *Journal of Advanced Technology Convergence*. 2025;4(2):21-30. doi:10.23152/JATC.2025.04.02.021.