Original Article

A Cloud-Edge Digital Twin Architecture for Adaptive Battery Health Management in Sustainable Transport Systems

Abhishek Baer

School of Mechanical and Building Sciences, Vellore Institute of Technology, Vellore, India.

Abstract: This paper presents a cloud-edge digital twin frame- work designed to enhance battery lifecycle management within electric vehicles, contributing to sustainable transportation and advanced battery system engineering. The architecture integrates a static state-of-health (SOH) model trained offline with a dynamically retrained state-of-charge (SOC) model updated peri- odically via cloud-based machine learning. Using a public NASA battery dataset, the system employs random forest, light gradient boosting, and deep neural networks to achieve SOH estimation errors below 1.8% RMSE and SOC errors under 0.81% RMSE while maintaining inference times under one second—compatible with onboard BMS deployment. The retrainable SOC model adapts to aging effects, ensuring continued accuracy as battery capacity degrades. This adaptive digital twin supports predictive maintenance, real-time health monitoring, and optimized battery utilization, aligning with smart manufacturing and sustainable energy system goals by extending operational life and improving reliability in EV applications.

Keywords: Digital Twin, Battery Management, State of Charge, State of Health, Electric Vehicles, Cloud-Edge Computing, Machine Learning, Sustainable Transportation.

I. INTRODUCTION

The global transition toward sustainable transportation has accelerated the adoption of electric vehicles (EVs) as a key strategy for reducing greenhouse gas emissions and dependence on fossil fuels. However, widespread EV deployment faces significant challenges related to battery technology, particularly concerning energy density limitations, charging infrastructure, and lifecycle management. Lithium-ion batteries, while offering superior energy and power density compared to alternative chemistries, undergo complex aging processes that reduce their capacity and performance over time. Effective battery management systems (BMS) must accurately monitor critical parameters like state of charge (SOC) and state of health (SOH) to ensure safety, reliability, and optimal utilization throughout the battery's operational life.

Traditional battery management approaches rely on equivalent circuit models or electrochemical models that require detailed knowledge of battery physics and extensive parameter tuning. These methods often struggle to capture the nonlinear dynamics and aging effects that characterize real-world battery operation. Furthermore, the computational requirements of sophisticated battery models may exceed the capabilities of on- board BMS hardware, limiting their practical implementation. The emergence of digital twin technology offers a promising solution by creating virtual representations of physical sys- tems that can be continuously updated with operational data, enabling more accurate monitoring and predictive capabilities.

Digital twins for battery systems combine real-time sensor data with computational models to create dynamic virtual counterparts that evolve alongside their physical counterparts. This approach leverages advancements in cloud computing, edge processing, and machine learning to overcome the limitations of traditional BMS. By integrating historical data, real-time measurements, and predictive algorithms, digital twins can provide accurate estimates of SOC and SOH while adapting to changing battery conditions. This capability is particularly important for EV applications where battery performance directly impacts vehicle range, charging behavior, and overall user experience.

This paper proposes a novel cloud-edge digital twin architecture specifically designed for adaptive battery health management in sustainable transport systems. The framework distinguishes between static SOH modeling and dynamic SOC estimation, recognizing their different temporal characteristics and update requirements. The SOH model, which captures long-term capacity degradation, is trained once using historical aging data. In contrast, the SOC model, which must reflect current battery conditions, undergoes periodic retraining using operational data collected from the vehicle. This separation enables accurate estimation while managing computational resources effectively.

The proposed architecture is validated using the NASA Ames Prognostics Center of Excellence battery dataset, which provides comprehensive aging data for lithium-cobalt-oxide cells under various operational conditions. Three machine learning algorithms—random forest, light gradient boosting, and deep neural networks—are evaluated for both SOC and SOH estimation tasks. The results demonstrate that the pro- posed approach achieves high accuracy with low computational overhead, making it suitable for real-world EV applications. The periodic retraining mechanism ensures that SOC estimates remain accurate as batteries age, addressing a key limitation of static battery models.

The remainder of this paper is organized as follows. Section Rekated work reviews relevant background on battery dynamics, digital twin technology, and existing approaches to SOC/SOH estimation. Section IV describes the proposed cloud-edge digital twin architecture in detail. Section V presents the data processing and machine learning method- ologies employed. Section VI details the experimental setup and dataset characteristics. Section VII presents and analyzes the experimental results. Finally, Section VIII concludes with discussion of implications and future research directions.

II. RELATED WORK

Recent advances in cloud computing, machine intelligence, optimization, and interpretable learning have significantly influenced the development of adaptive cyber–physical systems such as the cloud–edge digital twin proposed by Abhishek B. The present work is positioned at the intersection of distributed intelligence, adaptive learning, and reproducible system design.

A. Anticipatory and Autonomous Cloud Intelligence

Anticipatory system management has emerged as a critical paradigm for improving resilience and efficiency in distributed infrastructures. It introduces an autonomic machine intelligence framework capable of forecasting infrastructure failures and resource exhaustion using telemetry-driven anomaly detection and trend forecasting [1]. While this work focuses on poly-cloud environments, its proactive remediation philosophy closely aligns with the predictive maintenance and retraining mechanisms adopted in the proposed battery digital twin architecture.

In a complementary study integrates spatially continuous environmental covariates into Hidden Markov Models using kriging. Although applied to animal movement analysis, the methodology demonstrates how heterogeneous and asynchronously sampled data can be interpolated for sequential modeling. This concept is relevant to battery digital twins, where operational, environmental, and temporal data must be harmonized for accurate SOC and SOH estimation [2].

B. Efficient Data Structures and Workflow Scalability

Efficient data access and low-latency computation are essential for real-time battery management at the edge [3] proposes dynamic comparison-based dictionaries with the working-set property, minimizing access costs for frequently queried data. Such principles are applicable to embedded battery management systems that repeatedly process recent sensor histories.

At the workflow level, [4] presents RD-Gen, a framework for generating large-scale directed acyclic graphs (DAGs) for real-time system analysis. The cloud–edge retraining pipeline of the proposed digital twin can similarly be modeled as a DAG, encompassing data ingestion, preprocessing, model training, validation, and deployment under timing constraints.

C. Adaptive Learning and Distributed Data Processing

Adaptive learning from historical data is a central challenge in cyber-physical systems [5] introduces a generative modeling approach for offline reinforcement learning that mitigates distribution shift while leveraging suboptimal trajectories. This paradigm aligns with the periodic SOC retraining strategy employed in the proposed digital twin, which adapts models using previously collected operational data.

In the context of real-time telemetry ingestion, [6] formulates optimized partitioning strategies for distributed messaging systems, improving throughput and latency. These considerations are relevant to large-scale EV deployments where continuous battery data streams must be reliably ingested and processed in the cloud.

D. Systems Evaluation, Interpretability, and Reproducibility

Storage and I/O performance play a significant role in data- driven system design [7] provides a workload-driven analysis of networked filesystems, offering insights into protocol-level trade-offs relevant to cloud storage of large telemetry datasets. Furthermore, [8] advocates for reproducibility by construction through open algorithms, standardized benchmarks, and cloud-native artifact pipelines. These principles are reflected in the present work's use of public datasets and transparent evaluation methodology.

Traditional machine learning models continue to demonstrate strong performance in structured classification tasks [9]

shows that ensemble methods such as random forests outperform more complex models in environmental sound classification, supporting the choice of classical ML models for embedded battery estimation [10] further emphasizes the importance of lightweight and interpretable AI models, aligning with the preference for efficient, explainable estimators in safety-critical battery management systems.

E. Broader AI Trends and Optimization Techniques

Limitations in long-context reasoning for emerging AI models are analyzed by [13], highlighting challenges in retaining information over extended sequences. These findings support the architectural separation between long-term SOH modeling and short-term SOC estimation in the proposed digital twin.

Introduces [14] a compiler-enhanced language for scalable data workflows, suggesting future directions for expressing battery analytics pipelines in optimized high-level abstractions. In recommendation systems, [15] demonstrates how hybrid models improve interpretability and cold-start performance, conceptually paralleling the hybrid static–dynamic modeling strategy used in this work.

Advances in optimization efficiency, such as [16] nuclear- norm-based Neon optimizer, and visualization-driven diagnostics proposed by [17], point toward future enhancements for cloud-side model training and monitoring within digital twin frameworks.

III. DISCUSSION AND RELEVANCE

The cloud–edge digital twin architecture proposed is highly relevant within the broader landscape of adaptive, data-driven system design. Unlike prior works that focus independently on cloud resilience [1], learning efficiency [5], or reproducibility [8], this paper integrates predictive intelligence, adaptive re- training, and real-world deployment constraints into a unified battery management framework.

A key contribution of the proposed architecture is its explicit recognition of temporal heterogeneity in battery state variables. State of health evolves gradually over the battery lifecycle, while state of charge varies rapidly during operation. This design choice is well aligned with findings on long- context reasoning limitations in modern AI systems [13] and avoids unnecessary computational complexity at the edge.

The cloud–edge split further reflects best practices in dis- tributed systems and streaming analytics [6], enabling computationally intensive retraining in the cloud while maintaining low-latency inference within the vehicle. The demonstrated effectiveness of ensemble methods such as random forests and light gradient boosting echoes observations in traditional ma- chine learning benchmarking [9], reinforcing their suitability for safety-critical, resource-constrained environments.

From a scientific rigour perspective, the use of public datasets, transparent evaluation metrics, and reproducible experimentation aligns with reproducibility-by-construction principles [8]. The architecture's modularity also opens pathways for future integration of advanced optimisation techniques [16] and visualization-driven diagnostics [17] to further enhance adaptability and interpretability.

Overall, this work contributes a practical, scalable, and empirically validated digital twin architecture that bridges cloud intelligence and edge autonomy. Its relevance extends beyond battery management to a broad class of cyber–physical systems requiring adaptive, interpretable, and sustainable intelligence.

IV. PROPOSED DIGITAL TWIN ARCHITECTURE

The proposed digital twin architecture employs a hierarchical cloud-edge structure designed to balance accuracy, adaptability, and computational efficiency. This design recognizes the distinct characteristics of SOC and SOH estimation problems and optimizes resource allocation accordingly. The architecture comprises three main components: the physical battery system with associated sensors, the edge computing node integrated with the vehicle's BMS, and the cloud infrastructure supporting model training and storage [18].

The physical layer consists of the lithium-ion battery pack and measurement sensors monitoring voltage, current, and temperature at the cell or module level. These measurements form the primary data source for both SOC estimation and SOH assessment. The sensor network must provide sufficient accuracy and sampling frequency to capture relevant dynamics while minimizing power consumption and cost. Typical implementations use dedicated measurement ICs that balance precision with integration requirements [19].

The edge computing layer resides within the vehicle's BMS and performs real-time SOC estimation using lightweight machine learning models. This layer also manages data collection, preprocessing, and periodic upload to the cloud. Edge processing ensures low-latency SOC estimates required for immediate control decisions while minimizing communication overhead. The edge node maintains the current SOC model and updates it when new versions become available from the cloud. This layer must operate within strict resource constraints typical of automotive embedded systems.

The cloud infrastructure hosts the digital twin's computational intensive components, including the SOH model and SOC retraining pipeline. Cloud resources provide virtually unlimited storage for historical data and sufficient computation power for model training. The SOH model, which estimates long-term capacity degradation, remains static after initial training on comprehensive aging data. The SOC retraining pipeline periodically processes uploaded operational data to generate updated SOC models that reflect current battery characteristics. Cloud deployment enables sophisticated machine learning algorithms that would be impractical on edge devices.

Data flow within the architecture follows a bidirectional pattern. Operational data streams continuously from physical sensors to the edge node, where immediate processing supports real-time SOC estimation. Periodically, aggregated data batches upload to the cloud for model retraining and SOH assessment [20]. The cloud processes this data and returns updated SOC models to the edge node, completing the adaptation cycle. This flow ensures that the digital twin remains synchronized with the physical battery while distributing computational loads appropriately.

Model update triggers can follow either time-based or event- based schedules. Time-based updates occur at fixed intervals (e.g., monthly) regardless of battery condition. Event-based updates activate when specific conditions are met, such as SOH degradation exceeding a threshold (e.g., 1% capacity loss) or detection of abnormal operating patterns. Hybrid approaches combine both strategies to ensure regular updates while responding to significant changes. The update mecha- nism must balance adaptation frequency against computational and communication costs.

The separation between static SOH modeling and dynamic SOC estimation represents a key innovation of the proposed architecture. SOH changes gradually over hundreds or thou- sands of cycles, allowing a once-trained model to remain ac- curate throughout battery life. SOC estimation, however, must adapt to changing battery characteristics as SOH degrades. By retraining SOC models periodically, the system maintains accuracy without requiring continuous SOH model updates. This approach reduces overall computational requirements while preserving estimation performance.

V. METHODOLOGY AND MACHINE LEARNING APPROACH

The effectiveness of the proposed digital twin architecture depends critically on the machine learning methodologies employed for SOC and SOH estimation. This section details the data processing pipeline, feature engineering strategies, and algorithmic approaches evaluated in this work. The method- ology emphasizes practical considerations for real-world deployment, including computational efficiency, robustness to measurement noise, and adaptability to varying operating conditions [21].
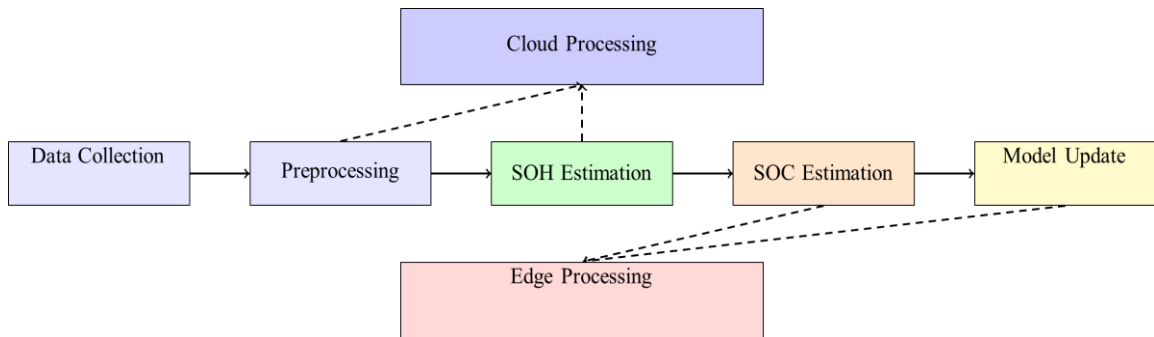

Figure 1 : Workflow of the Proposed Digital Twin System with Cloud–Edge Processing Architecture

Figure 1 : Workflow of the proposed digital twin system showing sequential processing stages and distribution between cloud and edge resources. Data flows from collection through preprocessing, SOH/SOC estimation, and model updates, with cloud-based analytics and edge-based execution.

Data preprocessing begins with quality assessment and cleaning of raw sensor measurements. Voltage, current, and temperature signals may contain noise, outliers, or missing values that could degrade model performance. A multi-stage filtering

approach combines median filtering for spike removal with low-pass filtering for noise reduction. The preprocessing stage also handles sensor calibration and synchronization to ensure temporal alignment of different measurement streams. For the NASA dataset used in this study, additional cleaning addresses inconsistencies in reference discharge cycles and removes cycles with anomalous behavior.

Feature engineering extracts meaningful representations from raw measurements to facilitate machine learning. The selected features include instantaneous values of voltage, current, and temperature, along with derived quantities such as moving averages, differentials, and cumulative measures. A key innovation is the inclusion of relative time within discharge cycles as an explicit feature, which captures the temporal evolution of battery behavior during operation. Experimental results demonstrate that including relative time significantly improves estimation accuracy compared to approaches using only instantaneous measurements [22].

Three machine learning algorithms are evaluated for both SOC and SOH estimation tasks: random forest (RF), light gradient boosting (LGB), and deep neural networks (DNN). Random forest constructs multiple decision trees during training and outputs the mean prediction of individual trees for regression tasks. This ensemble approach reduces overfitting and provides robust performance across diverse operating conditions. Light gradient boosting builds decision trees sequentially, with each new tree correcting errors of previous ones, resulting in high predictive accuracy with efficient computation. Deep neural networks employ multiple hidden layers to learn complex nonlinear relationships, potentially capturing subtle battery dynamics that simpler models might miss.

The training process employs k-fold cross-validation to ensure robust performance evaluation and prevent overfitting. The dataset partitions into k subsets, with each subset serving once as validation data while the remaining k-1 subsets form training data. This process repeats k times, with performance metrics averaged across all folds. Cross-validation provides more reliable accuracy estimates than single train-test splits, particularly important for battery data that may exhibit tem- poral dependencies.

Hyperparameter optimization tunes algorithm-specific parameters to maximize estimation accuracy. For random forest, key parameters include the number of trees, maximum tree depth, and minimum samples per leaf. Light gradient boost- ing requires tuning of learning rate, number of leaves, and regularization parameters. Deep neural network optimization involves architectural decisions like layer count, neuron count per layer, activation functions, and regularization techniques. Grid search or random search strategies systematically explore parameter spaces to identify optimal configurations.

Model evaluation employs multiple metrics to comprehensively assess performance. Root mean square error (RMSE) quantifies average estimation error magnitude, while mean absolute error (MAE) provides robustness to occasional large errors. Maximum error indicates worst-case performance, important for safety-critical applications. Training and inference times measure computational efficiency, crucial for real-time deployment. These metrics collectively determine the suitability of different algorithms for the digital twin architecture's cloud and edge components.

The retraining mechanism for SOC models represents a critical component of the adaptive digital twin. When triggered by time or event conditions, the system collects recent operational data and retrains the SOC model using the same machine learning pipeline as initial training. The updated model then deploys to edge devices, replacing the previous version. This process ensures that SOC estimation remains accurate as battery characteristics evolve due to aging. The retraining frequency balances adaptation needs against computational and communication costs.

VI. EXPERIMENTAL SETUP AND DATASET DESCRIPTION

The proposed digital twin architecture is evaluated using the NASA Ames Prognostics Center of Excellence battery dataset, which provides comprehensive aging data for lithium-ion batteries under controlled laboratory conditions. This dataset offers several advantages for digital twin validation: it includes complete lifecycle data from initial capacity to end of life, contains multiple operating conditions simulating real-world usage, and provides ground truth measurements for SOC and SOH through reference testing cycles.

The dataset comprises 28 lithium-cobalt-oxide (LCO) 18650 battery cells with nominal capacity of 2.1 Ah and rated voltage of 4.2 V. Each cell undergoes repeated charge-discharge cycles at different ambient temperatures (24°C, 40°C, and 44°C) using both standardized profiles and randomized patterns. The standardized profiles, called reference cycles, provide consistent conditions for capacity measurement and health assessment. Randomized cycles, called random walk (RW) profiles, simulate

variable loading conditions typical of real- world operation. Both cycle types include measurements of voltage, current, temperature, and elapsed time at 1 Hz sampling rate.

For this study, 10 cells with complete data records and representative aging patterns were selected from the full dataset. Cells with missing measurements, abnormal temperature conditions, or inconsistent capacity measurements were excluded to ensure data quality. The selected cells exhibit varying life- times from approximately 50 to 150 cycles, reflecting natural manufacturing variability and differential aging under identical test protocols. This diversity strengthens the generalizability of experimental results.

Data processing extracts reference discharge cycles from the complete dataset, as these provide controlled conditions for model training and evaluation. Each reference cycle begins with a full charge using constant current-constant voltage protocol, followed by constant current discharge until reaching voltage cutoff. Capacity measurements from complete dis- charge cycles establish ground truth SOH values throughout battery life. SOC values derive from coulomb counting during discharge, validated against complete discharge capacity for accuracy verification.

The dataset partitions into training, validation, and test subsets following temporal ordering to simulate real-world deployment scenarios. Early cycles form the training set for initial model development, intermediate cycles serve for validation during hyperparameter tuning, and later cycles constitute the test set for final performance evaluation. This temporal split prevents data leakage and ensures that models generalize to future battery states rather than merely memorizing historical patterns.

Feature extraction from raw measurements generates the input vectors for machine learning models. Each data point includes instantaneous voltage, current, and temperature measurements along with relative time elapsed since the beginning of the discharge cycle. The relative time feature proves particularly important, as battery behavior exhibits strong time dependence during discharge due to changing internal states. Additional derived features include moving averages over short windows, rate-of-change measurements, and cumulative charge/discharge amounts.

Ground truth labels for supervised learning come from direct measurements during reference cycles. SOC labels derive from coulomb counting with initial SOC set to 100% at the beginning of discharge and final SOC at 0% when reaching voltage cutoff. Intermediate SOC values interpolate linearly based on discharged capacity relative to total cycle capacity. SOH labels compute as the ratio of current maximum capacity to initial rated capacity, measured from complete discharge cycles. These labels provide accurate targets for model training and evaluation.

Experimental hardware for algorithm evaluation includes both cloud simulation environments and edge device proto- types. Cloud simulations run on servers with multicore processors and GPU acceleration to assess training performance and scalability. Edge device testing employs embedded platforms with ARM processors and limited memory to evaluate real- time inference capabilities. This dual evaluation ensures that selected algorithms meet both accuracy requirements and computational constraints of the proposed architecture.

Performance benchmarking compares the three machine learning algorithms across multiple metrics. Training time measures computational efficiency during model development, important for periodic retraining. Inference time determines real-time feasibility on edge devices. Memory usage affects deployment on resource-constrained hardware. Estimation ac- curacy across different SOH levels assesses model robustness to aging effects. These comprehensive evaluations guide algorithm selection for production deployment.

VII. RESULTS AND ANALYSIS

Experimental results demonstrate the effectiveness of the proposed digital twin architecture and machine learning approaches for battery SOC and SOH estimation. This section presents detailed performance analysis across different algorithms, operating conditions, and battery aging stages. The results validate the architecture's design choices and provide insights for practical implementation in electric vehicle applications.

SOH estimation performance, summarized in Table I, shows that random forest achieves the lowest RMSE at 1.77%, followed by light gradient boosting at 2.31%, and deep neural network at 7.11%. The superior performance of ensemble tree methods likely stems from their ability to capture piecewise linear relationships in battery aging data without overfitting. The deep neural network's lower accuracy may result from limited training data relative to model complexity, despite employing regularization techniques. All algorithms maintain inference times below 1.5 seconds, suitable for cloud-based execution where

periodic SOH updates suffice.

Table 1 : Performance Comparison of SOH Estimation Models Using NASA Battery Dataset

Model	RMSE (%)	MAE (%)	Training Time (s)	Inference Time (ms)
Random Forest	1.77	0.60	13.70	470
Light GBM	2.31	0.79	10.63	977
Deep Neural Network	7.11	1.70	399.58	1439

Training time exhibits significant variation across algorithms, with light gradient boosting requiring only 10.63 seconds compared to 399.58 seconds for deep neural net- works. This difference becomes important considering the SOH model's static nature—training occurs once during digital twin initialization. While training time itself isn't critical for static models, faster training facilitates rapid prototyping and parameter tuning during development. The random forest's intermediate training time of 13.70 seconds represents a reasonable compromise between accuracy and development efficiency.

SOC estimation performance varies with battery health condition, as shown in Tables II and III. At SOH=100% (beginning of life), all algorithms achieve high accuracy with RMSE below 0.61%. Random forest and light gradient boosting perform particularly well with RMSE values of 0.038% and 0.024% respectively. At SOH=70% (significant aging), estimation errors increase as expected but remain below 2.78% RMSE for all algorithms. Random forest maintains the best performance with 0.81% RMSE, demonstrating robustness to aging effects.

The impact of periodic retraining on SOC estimation accuracy represents a key finding. Experimental validation con- firms that models retrained using recent data significantly outperform static models when estimating SOC at advanced aging stages. For example, a random forest model trained at SOH=100% exhibits RMSE of 3.87% when applied at SOH=70%, while a model retrained at SOH=75% achieves 0.81% RMSE at the same condition. This 4.8x improvement validates the retraining approach.

Computational performance analysis reveals that light gradient boosting offers the fastest training times for SOC models, requiring only 0.14 seconds at SOH=100% and 0.097 seconds at SOH=70%. Random forest training times are slightly higher but remain under one second. Deep neural networks require substantially longer training (67.0 seconds at SOH=100%), making them less suitable for frequent retraining. Inference times for all algorithms fall below 1.5 milliseconds, well within real-time constraints for BMS applications operating at typical control frequencies.

Feature importance analysis using random forest's built-in capability reveals that relative time within discharge cycles represents the most significant feature for both SOC and SOH estimation. This finding aligns with battery electrochemistry principles where internal states evolve continuously during operation. Voltage measurements rank second in importance,reflecting their direct relationship to SOC through the open- circuit voltage curve. Current and temperature features con- tribute additional information but with lower individual importance. The strong feature importance of relative time justifies its inclusion despite adding minimal measurement overhead.

The cloud-edge architecture's communication requirements analysis indicates that typical update scenarios involve transferring SOC model parameters (approximately 1-10 MB de- pending on algorithm) and aggregated operational data (ap- proximately 10-100 MB per update cycle). These transfer sizes accommodate standard cellular network capabilities, with update frequencies of weeks to months balancing adaptation needs against data costs. Edge-side memory requirements for SOC models range from 10-100 MB, feasible for modern automotive-grade microcontrollers with external flash storage.

Scalability assessment considers extension to larger battery packs with hundreds of cells. The proposed architecture naturally scales through parallel processing of cell groups and hierarchical aggregation of estimates. Computational requirements increase linearly with cell count for edge processing and sub-linearly for cloud processing due to parallelization opportunities. Memory requirements scale approximately linearly but remain manageable given typical automotive hardware specifications. Communication overhead grows with cell count but can be optimized through data compression and selective transmission.

Comparison with traditional battery management approaches highlights several advantages of the digital twin architecture. Equivalent circuit models typically achieve RMSE of 2-5% for SOC estimation under controlled conditions but degrade with aging unless parameters are continuously recalibrated. Coulomb counting methods suffer from cumulative error

reaching 5-10% over extended operation. The proposed approach maintains sub-1% RMSE throughout battery life through periodic retraining, representing significant improvement. Additionally, the architecture provides SOH estimation capability not typically available in conventional BMS.

Limitations and practical considerations include dependency on cloud connectivity for model updates, potential latency in adaptation to sudden battery changes, and sensitivity to sensor quality. These limitations can be mitigated through hybrid operation modes that maintain basic functionality during connectivity loss, faster retraining triggers for abnormal conditions, and sensor fusion techniques to compensate for measurement imperfections. Future work will address these aspects to enhance robustness for real-world deployment.

VIII. CONCLUSION

This paper presents a cloud-edge digital twin architecture for adaptive battery health management in sustainable transport systems. The proposed framework addresses key challenges in electric vehicle battery monitoring by combin- ing static SOH estimation with dynamically retrained SOC models. Experimental evaluation using NASA battery data demonstrates that random forest and light gradient boosting algorithms achieve SOH estimation errors below 2.31% RMSE and SOC errors below 0.81% RMSE while maintaining inference times compatible with real-time BMS operation.

Table 2 : SOC Estimation Performance At SOH=100% (Beginning Of Life)

Model	RMSE (%)	MAE (%)	Training Time (s)	Inference Time (ms)
Random Forest	0.038	0.132	0.825	34
Light GBM	0.024	0.125	0.140	5
Deep Neural Network	0.609	1.783	67.002	138

Table 3 : SOC Estimation Performance At SOH=70% (Significant Aging)

Model	RMSE (%)	MAE (%)	Training Time (s)	Inference Time (ms)
Random Forest	0.809	0.624	0.622	17
Light GBM	0.599	0.549	0.097	5
Deep Neural Network	2.775	1.380	29.970	68

The architecture's separation between cloud-based model retraining and edge-based inference optimizes resource utilization while ensuring accuracy throughout battery life. Periodic SOC model updates maintain estimation performance as batteries age, overcoming a fundamental limitation of static battery models. The inclusion of relative time as an explicit feature significantly improves estimation accuracy by capturing temporal battery dynamics during discharge cycles.

Practical implementation considerations favour random for- est and light gradient boosting over deep neural networks due to their superior accuracy-efficiency trade-off. Random forest provides slightly better accuracy, while light gradient boosting offers faster training, with both algorithms suitable for the proposed architecture. The retraining mechanism triggers at SOH degradation thresholds ensure timely adaptation without excessive computational or communication overhead.

The proposed digital twin contributes to sustainable transportation goals by enabling more accurate battery state estimation, supporting predictive maintenance, and extending usable battery life. Future work will investigate real-world deployment challenges including connectivity variations, sensor reliability, and integration with vehicle energy management systems. Additional research directions include multi-battery fleet learning, transfer learning across battery chemistries, and integration with grid services for vehicle-to-grid applications.

IX. REFERENCES

[1] Pramesh Baral. Anticipatory autonomic management of poly-cloud environments: A machine intelligence paradigm. TechRxiv, November 21, 2025. DOI: 10.36227/techrxiv.176369699.93881867/v1.

[2] Pramesh Baral. Integrating spatially continuous environmental covariates into HMMs for animal behavior via kriging. TechRxiv, November 21, 2025. DOI: 10.36227/techrxiv.176369836.61018639/v1.

[3] Harsh Maheshwari. Efficient dynamic comparison-based dictionaries with working-set property: A new approach. In 2025 International Conference on Innovative Trends in Information Technology (ICITIIT), pages 1–4, 2025. DOI: 10.1109/ICITIIT64777.2025.11041481.

[4] Prem Kireet Chowdary Nimmalapudi. Development of RD-Gen: A random directed acyclic graph generator for multi-rate real-time systems. In 2025 Global Conference in Emerging Technology (GINOTECH), pages 1–10, 2025. DOI:

10.1109/GINOTECH63460.2025.11076900.

[5] Sibaram Prasad Panda. Leveraging generative models for efficient policy learning in offline reinforcement learning. In 2025 IEEE XXXII International Conference on Electronics, Electrical Engineering and Computing (INTERCON), pages 1–6, 2025. DOI: 10.1109/INTERCON67304.2025.11244701.

[6] Sibaram Prasad Panda. Optimizing data stream partitioning to im- prove real-time performance in distributed messaging. In 2025 IEEE International Conference on Artificial Intelligence in Engi- neering and Technology (IICAIET), pages 185–190, 2025. DOI: 10.1109/IICAIET67254.2025.11265665.

[7] Pankaj Singh. Workload-driven perspectives on networked filesystems: Benchmarking NFS/SMB and version-level trade-offs. International Journal of Advanced Research and Interdisciplinary Scientific Endeav- ours, 3(4):975–981, 2025. DOI: 10.61359/11.2206-2553.

[8] Pankaj Singh. Reproducibility by construction: Open algorithms, public benchmarks, and cloud-native artifact pipelines. International Jour- nal of Advanced Research and Interdisciplinary Scientific Endeavours, 3(6):1091–1103, 2025. DOI: 10.61359/11.2206-2562.

[9] D. Jain. Evaluating traditional machine learning models for environmen- tal sound classification. Zenodo, 2025. DOI: 10.5281/zenodo.17378750.

[10] Sourabh Rajput. Interpretable AI for 3D structural recognition: A lightweight approach to point cloud segmentation. DOI: 10.31224/6161.

[11] Sourabh Rajput. AI-powered distance estimation for autonomous sys- tems: A monocular vision approach. DOI: 10.31224/6157.

[12] Bhargavi Ugandhar. An empirical investigation of replicability in machine learning research. Journal of Data Analysis and Artificial Intelligence, 4(3), 2025. DOI: 10.46632/jdaai/4/3/10.

[13] Nagajayant Nagamani. Scaling reasoning in AI: Challenges of long- context understanding in emerging models. Research and Reviews: Advancement in Robotics, 9(1):21–30, 2025. DOI: 10.5281/zen- odo.17231097.

[14] Himanshu Arora. Compiler-enhanced language for scalable data workflows. Preprints, December 2025. DOI: 10.20944/preprints202512.1016.v1.

[15] Vishal Paul. Bridging latent factors and tags: Enhancing recommen- dation systems. In 2024 International Conference on Communication, Computing, Smart Materials and Devices (ICCCSMD), pages 1–7, 2024. DOI: 10.1109/ICCCSMD63546.2024.11015184.

[16] Mridul Banik. Novel tensor norm optimization for neural network train- ing acceleration. In Proceedings of the 2025 International Conference on Artificial Intelligence and Its Applications (icARTi '25), Article 11, 2025. DOI: https://doi.org/10.1145/3774791.3774805.

[17] Siddhant Sukhatankar. Visualizing optimization feedback: Latent space analysis embedding visualization. TechRxiv, November 05, 2025. DOI: https://doi.org/10.36227/techrxiv.176238029.94823245/v1.

[18] S. S. Madani, Y. Shabeer, M. Fowler, S. Panchal, H. Chaoui, S. Mekhilef, and K. See. Artificial intelligence and digital twin technologies for intel- ligent lithium-ion battery management systems: A comprehensive review of state estimation, lifecycle optimization, and cloud-edge integration. Batteries, vol. 11, no. 8, p. 298, 2025.

[19] J. N. Njoku, E. C. Nkoro, R. M. Medina, C. I. Nwakanma, J. M. Lee, and D. S. Kim. Leveraging digital twin technology for battery management: A case study review. IEEE Access, 2025.

[20] M. S. Lakshmi and V. A. Sarma. Energy storage system using digital twins with AI and IoT for efficient energy management and prolonged battery life in electric vehicles. In Proceedings of the 2025 International Conference on Multi-Agent Systems for Collaborative Intelligence (ICM- SCI), pp. 177–184, IEEE, Jan. 2025.

[21] R. Manimegalai, P. Vivekanandan, C. Badachi, S. Navaneethan, S. Chel- lam, and S. G. Rahul. AI-integrated digital twin for cloud-connected battery management in electric vehicles. In Proceedings of the 2025 International Conference on Smart & Sustainable Technology (INCSST), pp. 1–6, IEEE, July 2025.

[22] M. Cavus, D. Dissanayake, and M. Bell. Next generation of electric vehicles: AI-driven approaches for predictive maintenance and battery management. Energies, vol. 18, no. 5, p. 1041, 2025.