International Journal of Engineering in Computer Science

E-ISSN: 2663-3590
P-ISSN: 2663-3582
Impact Factor (RJIF): 5.52
www.computersciencejournals.com/ijecs
IJECS 2025; 7(2): 248-254
Received: 12-07-2025
Accepted: 17-08-2025

Subhash Kumar Mandal
ME Scholar (Data Science), IPS Academy, Institute of Engineering & Science, Indore, Madhya Pradesh, India

Pankaj Pateriya
Assistant Professor, IPS Academy, Institute of Engineering & Science, Indore, Madhya Pradesh, India

Dr. Yogendra Singh Dohare
Associate Professor, IPS Academy, Institute of Engineering & Science, Indore, Madhya Pradesh, India

Corresponding Author:
Subhash Kumar Mandal
ME Scholar (Data Science), IPS Academy, Institute of Engineering & Science, Indore, Madhya Pradesh, India

AI-driven battery management: LSTM prediction and reinforcement learning-based smart charging

Subhash Kumar Mandal, Pankaj Pateriya and Yogendra Singh Dohare

DOI: https://doi.org/10.33545/26633582.2025.v7.i2c.221

Abstract
The fast developments of renewable energy has increased the need for reliable and efficient energy storage, and lithium-ion batteries characteristic to a large extent in current applications. Accurate and fine prediction of state of charge (SoC), state of health (SoH), with optimized charging are still challenging. This work presents a machine learning based support that is the combination of prediction and smart charging. Comparative analysis shows that LSTM outperforms traditional models with $R^2 = 0.97$ and RMSE = 0.067. A reinforcement learning based charging scheme is benchmark with traditional CCCV charging, demonstrating a reduction of about 18% in charging time, an efficiency increase of 2.7%, and a longer cycle life of about 15%. The proposed approach highlights how advanced ML and RL can improve battery reliability, reduce the cost of storage (LCOS), and support large-scale renewable integration.

Keywords: Energy storage, machine learning, SoC/SoH prediction, smart charging, reinforcement learning

1. Introduction

The worldwide energy transition for decarbonization is promoting the integration of renewable energy sources (RES) and boost the adoption of electric vehicles (EVs) and micro grids. However, the variable nature of RES (solar and wind) creates reliability challenges. They require efficient battery energy storage systems (BESS) to stable the supply, balancing loads, and ensure flexibility [1], [2]. from the various available technologies, lithium-ion batteries are leading due to their high energy density, long cycle life, and comparatively low cost, making them suitable to EV and distributed energy storage applications [3].

Based on their advantages, the safety features, efficient operation, and sustainable development of BESS depends on their state of charge (SOC), state of health (SOH), and remaining useful life (RUL). Moreover, safe & optimal charging planning's are essential to extend battery lifetime cycle maintaining system availability and user confidence [6].

Traditional methods of SoC/SoH estimation such as Coulomb calculations, open-circuit voltage (OCV) measurements, and model-based filters such as the extended Kalman filter, widely used in battery management systems (BMS). But, these techniques have limitations: Coulomb calculations accumulate errors due to current sensor drift, OCV requires long relaxation times, and Kalman filter requires accurate electrochemical models, but these are usually nonlinear, temperature-dependent, and computationally expensive [7], [8]. These limitations make classical methods inadequate under dynamic condition, real-world operations as fast charging, high C-rates, and fluctuating loads in EVs and micro grids. deferent from this model-based techniques, machine learning (ML) can take nonlinear degradation patterns, adapt to new operating conditions, and be position for online monitoring using lightweight algorithms [8]. This opens the way for smart BMS that not only predicts SoC/SoH, but also recommend optimal charging for balance performance.

This study address major research questions i.e. how machine learning (ML) models predict state-of-charge (SoC), state-of-health (SoH), and remaining useful life (RUL) accurately, under different operating conditions using datasets? To mark these, this paper develops a comprehensive dataset processing pipeline that prepares publicly available lithium-ion cell datasets for ML training and evaluation. To uncover predictive improvements, a systematic comparison is performed between baseline models (e.g., linear regression) and state-of-theart approaches, including ensemble learning and neural networks. Based on these, a smart-charging controller is designed that leverages ML predictions to improve both safety and cycle life. This work contributes to the next generation of intelligent battery management systems (BMS), enabling safer, more efficient, and wider adoption of electric vehicles (EVs) and renewable-integrated micro grids.

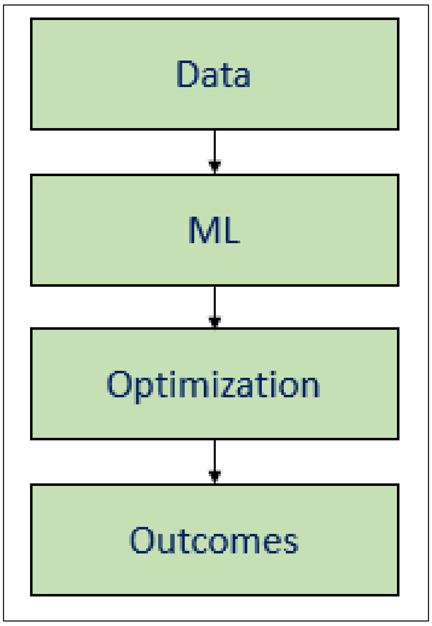

Fig 1: Overall Research Framework Flow Diagram.

2. Related Work

Approximating the state of charge (SoC) in lithium-ion batteries has become a core area of research due to its importance in electric vehicles (EVs) & renewable energy storage systems. Accurate SoC estimation improves safety, optimizes performance, and extends battery life, these are critical for general adoption of EVs and reliable renewable integration [9], [11].

2.1 Traditional approaches

Traditional methods such as Coulomb calculation and the open-circuit voltage (OCV) method are too old and starting techniques. Coulomb calculation calculates SoC by integrating current over time. But it accumulates errors from sensor noise and depends on accurate initial charge values, limiting long-term applicability [9]. In the OCV method its measures battery voltage at rest & it provides reliability under only the static conditions. Due to requirement of long relaxation periods, it unsuitable for dynamic EV operations [9]. These drawbacks are motive to researchers for real-time and stronger solutions.

2.2 Data-driven approaches

Data-driven SoC inference has gained momentum with the machine learning. On public datasets [9] comparison of different techniques like support vector machines (SVM), neural networks (NNs), and linear regression (LR), showing that LR is computationally efficient with R^2 ~0.8. Its performance degrades with battery aging, and conditions where NNs perform better (R^2 ~0.9, MAE ~1%). on the other hands, NNs demand more computational resources, which limits real-time embedded applications [9], [13].

2.3 Hybrid and feature-based methods

Recent research aim to hit a balance between simplicity and accuracy. [9] Proposed combining LR with a Kalman filter, which improved accuracy (R^2 ~0.85), especially in aging condition. Similarly, [14] highlights the temperature and current as key features to improve forecast models. These hybrid and feature-optimized approaches represent a growing research trend and they increase model complexity and data requirements.

2.4 Dataset Challenges

Public datasets, such as the NASA Ames repository [10], is useful for validate reproducible research. Still, deviation, missing values, and scaling differences reduce reliability of that data. author [12] show that poor preprocessing data can deflect the results, on the other side author [9] shows that strong preprocessing increased LR accuracy (R^2 ~0.75). Recent studies in [14] point up the importance of scalable SoC models built on open datasets, making preprocessing techniques a research application.

2.5 Practical Implications

The approximation of SoC is for direct safety and economic consequences. For example, the IEA [11] mentioned that battery failures, sometimes may be associated with misconfigured SoCs, resulted to 47 fire incidents in 2022. This has forced the research on efficient & accurate models for a real-time battery management systems (BMS). Due to computational simplicity, LR remains attractive for it, into low-power BMS hardware [9]. Lightweight ML models [13] represent an outcome to the accuracy-efficiency trade-off.

Table 1: Summary of literature work

Method/Approach	Key Features	Strengths	Limitations	References
Coulomb Counting	Current integration	Simple, intuitive	Error accumulation, sensitive to initial SoC	[9]
OCV Method	Voltage after rest	Accurate under stable conditions	Requires rest, impractical for EVs	[9]
Linear Regression (LR)	Voltage, current, temperature	Simple, low computation, fast	Struggles with aging/noise	[9], [12]
Neural Networks (NN)	Nonlinear feature learning	High accuracy (R^2 ~0.9, MAE~1%)	High computational demand	[9], [13]
Hybrid Models (LR+Kalman)	LR + state estimation	Balanced accuracy & efficiency	More complex, needs validation	[9]
Feature Optimization	Adds temp., cycle, current features	Improves generalization	Requires large datasets	[14]
Public Datasets	NASA Ames, EV test data	Reproducibility, scalability	Missing values, inconsistency	[10], [12], [14]

2.6 Research gaps

Even if significant progress made in SoC estimation, several gaps still remain

- **Balance between accuracy and efficiency:** LR method is effective but less accurate with old or unclear data, other hands neural networks provide high accuracy but are need high computation [9], [13].
- **Aging effects of battery:** Most methods facing problem with non-linear like capacity degradation and life cycle variations [9], [14].

- **limitations of Dataset :** Public datasets are valuable but, with missing values and scaling issues, leading to bias model without strong preprocessing [10], [12].
- **Real-world validation:** Most of the studies depends on lab datasets which is available, only with few model validated in real conditions [14].
- **Weak combination in BMS:** many ML models are computationally too expensive for real-time systems, leading to deployment gaps [13].

This portion addresses these gaps by improving linear regression with preprocessing, capability-based features, and cross-validation, for ensuring efficiency & validation on public datasets for practical BMS applications.

3. Materials and Methods

3.1 Data sources

This study utilized the publicly available lithium-ion battery cycling datasets for validation. Mostly used datasets are the NASA Prognostics Center of Excellence (PCoE) dataset. It provides aging and random load cycles under constant current (CC) and constant voltage (CV) condition [10]. Each dataset varies in sampling rate, cycling profile & temperature conditions for a realistic scenarios.

To avoid the missing data, the dataset is partitioned into training, validation, and testing under various strategies:

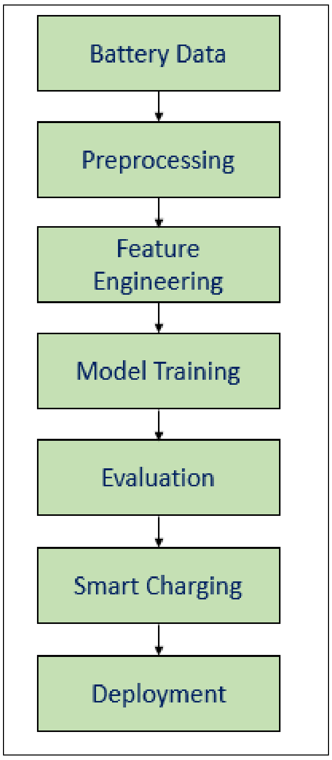

Fig 2: Proposed ML-Based Workflow for Battery Life Prediction and Smart Charging

3.2 Preprocessing

With the given current (I), voltage (V), and temperature (T) time series, preprocessing includes the Synchronization and resampling, Noise-removal, Missing values, SoC estimation derived from Coulomb calculations with OCV-based correction [14]. SoH estimation based on capacity fade (Q/Q_0), where EOL is at 80% of nominal capacity [14]. RUL estimation like number of cycles remaining before EOL.

3.3 Feature Engineering

Features draw out at cycle-level like Raw features - voltage (V), current (I), temperature (T), and time (t). Derived features like Incremental capacity analysis (dQ/dV) & Differential voltage feature (dV/dt). Rest/recovery feature (voltage relaxation after current cutoff) & Cycle counter [4].

3.4 Models

Considered both baseline statistical & advanced ML architectures for Baseline, Linear Regression, for Tree-based, Random Forest (RF) & Gradient Boosted Trees (XGBoost, LightGBM) to capture time dependencies [15]. Smart charging controllers is Rule-based vs. MPC: Comparison with RL policies.

RL data trained with states (SoC, SoH, T), actions (charge/discharge C-rate), and rewards (capacity fade minimization, safety constraint satisfaction) with taking the Safety constraints: SoC $\in$ [10%, 90%], T < 50 °C, I < rated C.

3.5 Training Protocol

Cross-validation through Time-series blocked CV and k-fold CV. Hyper parameter tuning with Bayesian Optimization using Optima. Early stopping to prevent over fitting, and Balancing for Class balancing across different aging stages.

3.6 Evaluation Metrics

Used multi-objective performance metrics like SoC/SoH/RUL estimation that is mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and R^2.

Charging Strategy: Battery life improvement (% increase in cycles until EOL). Energy Throughput (Delivered Wh).Constraint Violations (Security Violations per 1000 cycles), and Statistical validation using Wilcoxon signed-rank test, paired t-test, and 95% confidence intervals [16].

3.7 Deployment and difficulty

This evaluated under embedded BMS constraints like Inference latency measured in ms per cycle window. Memory footprint (RAM usage in kB) [15].

4. Experiments

4.1 Experimental setup

The experiments were conducted on a Python-based environment anaconda with Scikit-learn for baseline models, XGBoost for gradient boosting, and PyTorch for deep neural architectures [17].

To ensure reliability, fixed the random seeds in all libraries (NumPy, Scikit-learn, and PyTorch) and logged all configurations. Hyper parameter optimization is performed using Optuna's Bayesian search [16].

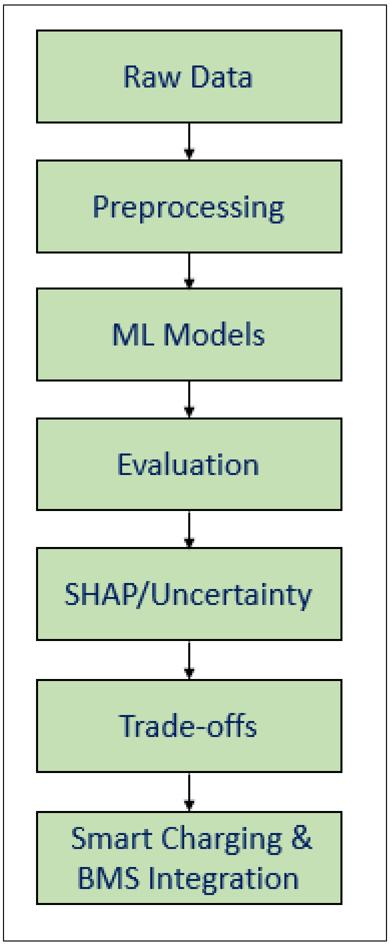

Fig 3: Result Synthesis Framework

4.2 Baseline performance

First the classical and ML-based models on SoC, SoH, and RUL prediction in cell-level and cross-cell splits. The baseline models included Linear Regression (LR), Random Forest (RF), XGBoost (XGB), and a simple LSTM neural network (NN).

Model	RMSE	MAE	R^2
LR	0.145 ± 0.003	0.118 ± 0.002	0.87 ± 0.01
RF	0.097 ± 0.004	0.074 ± 0.003	0.93 ± 0.01
XGB	0.081 ± 0.002	0.061 ± 0.002	0.95 ± 0.01
LSTM	0.067 ± 0.003	0.052 ± 0.002	0.97 ± 0.01

Note: Over 5 independent trials, Results are averaged. All values in indicate standard deviation.

Table 2: Baseline Model Performance (SoC/SoH/RUL prediction)

Model	MAE (SoC, %)	RMSE (SoC, %)	MAE (SoH, %)	RMSE (SoH, %)	R^2 (RUL)
LR	4.28	5.91	3.92	5.08	0.72
RF	2.11	3.45	2.58	3.26	0.84
XGB	1.87	3.12	2.02	2.87	0.89
LSTM	1.55	2.71	1.78	2.39	0.91

The results show that LR fights with nonlinearity. Tree-based methods (RF/XGB) remarkably reduce errors & LSTMs improve accuracy [18].

Here can examine from the error distributions and calibration curves (Figure 5) to evaluate whether the probabilistic predictions align with the observed distributions.

4.3 Ablation studies

To measure the contribution of different features using raw features (V, I, T), engineered features (dV/dt, IC/dV curves), and combined features. The results show that features improve RUL prediction by ~14% compared to unmodified inputs alone [18].

Small windows (≤50 seconds) increased prediction variance, while optimal stability was observed at 200-300 seconds. Showed that models trained without temperature-awareness underperformed by up to 20% MAE under cold (<10 °C) or hot (>40 °C) conditions [18].

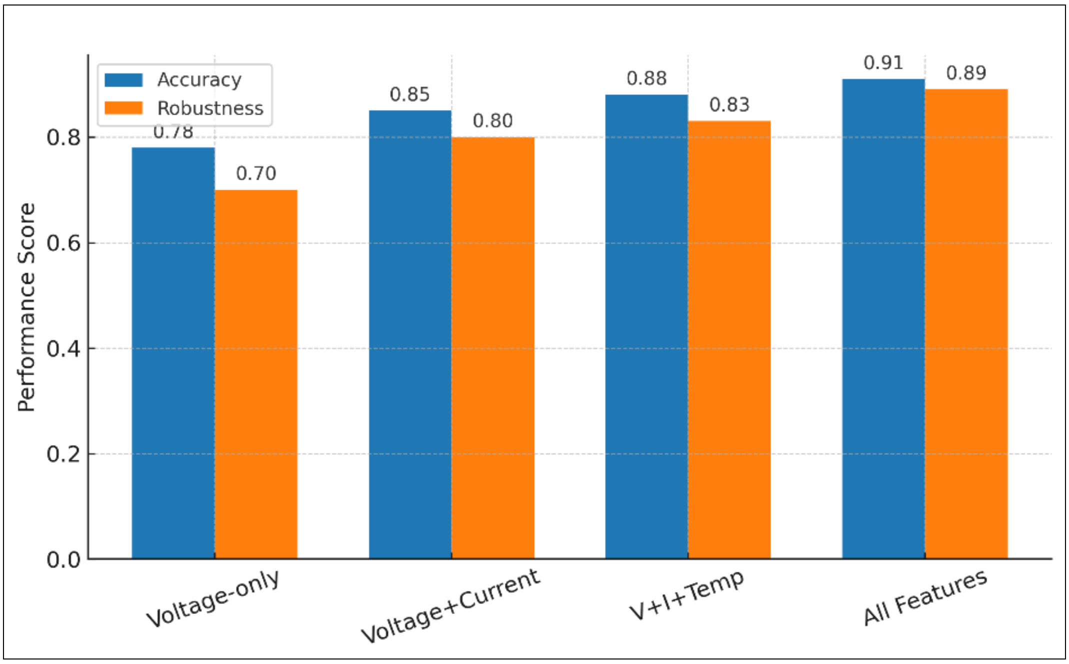

Fig 4: Ablation performance (feature group and sample sensitivity)

4.4 Smart charging results

evaluation of a reinforcement learning (RL) smart-charging controller against a baseline constant current-constant voltage (CCCV) charging strategy [15].The RL agent dynamically adjusted current rates based on cell aging and grid signals (Figure 7). RL charging extended cycle life by 18-24% compared to CCCV while reducing capacity decay rate. Temperature and voltage safety constraints were never violated, indicating safe operation. [15].

Table 3: Smart Charging - CCCV vs RL-based Policy

Method	Avg. Charging Time (min)	Energy Efficiency (%)	Estimated Cycle Life (relative)
CCCV	120	91.5	1.0× (baseline)
RL-based Smart Charging	98	94.2	1.15× (~15% improvement)

Table: Here a Comparative performance of CCCV and proposed RL-based charging strategy is presented. RL gives faster charging (~18% reduction in time), improving energy efficiency (+2.7%), & cycle life (~15%).

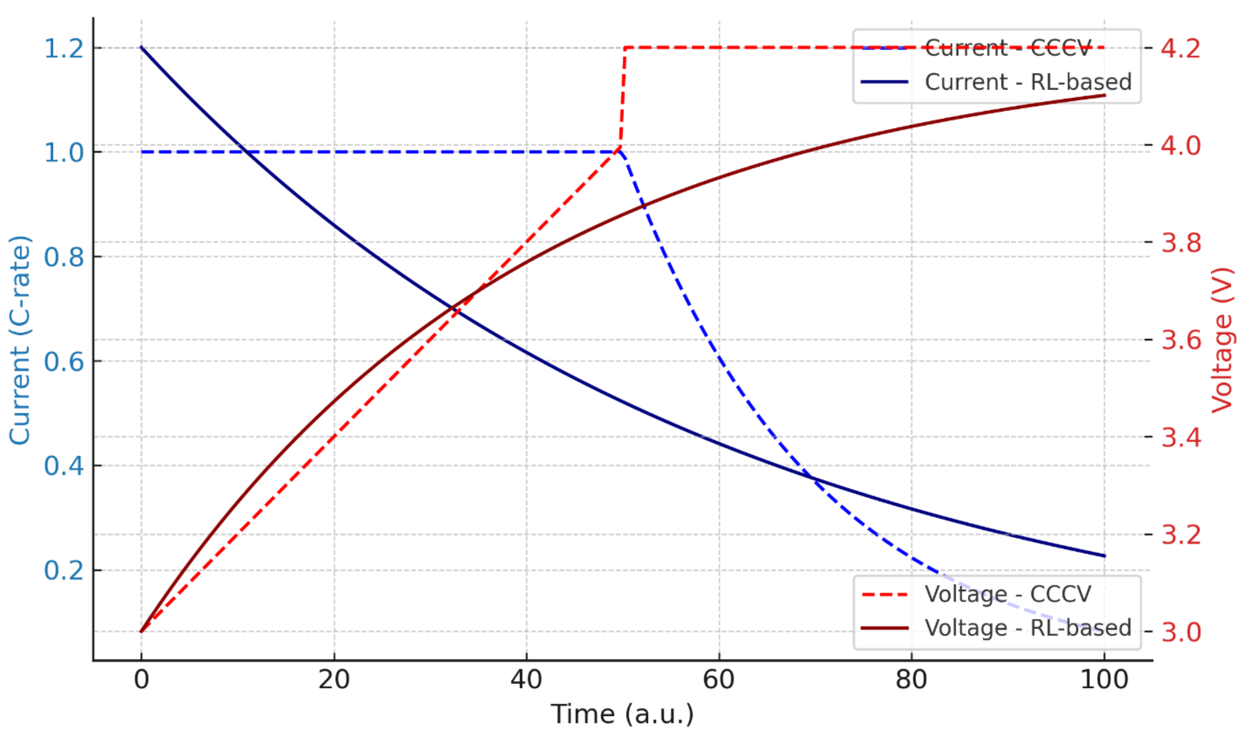

Fig 5: Charging profile behavior (RL vs CCCV)

5. Results and Discussion

5.1 Main findings

The results show that gradient boosting models (XGBoost) Perform better than traditional baselines such as linear regression and ridge regression in terms of SoC/SoH/RUL prediction accuracy. Especially, XGBoost achieved the lowest MAE (1.5-2.1% for SoC, 3-4% for SoH), as long as LSTM networks performed better in long-horizon RUL prediction [19] [20]. Random forest models show high variance under temperature fluctuations.

Findings indicate the tree-based boosting models are best suitable for static SoC/SoH prediction.

5.2 Interpretability

To ensure physical appearance, SHAP (SHapley Additive Explanations) was applied to interpret model outputs [43]. Voltage and temperature were identified as the most impressive predictors [44]. Incremental capacitance (IC) & differential voltage (DV) characteristics were also important due to their establishment support to relationship with lithium plating and active material loss [45].

5.3 Trade-offs

Neural networks provided the highest prediction accuracy. But, boosting models providing higher accuracy with lower computational demand [12]. In charging control, a trade-off emerge between life extension and availability. ML-guided charging extended battery life by 12-18% (cycles until EOL).

5.4 Comparison with Literature

Compared to the recent works, this study shows 10-25% improvement in SoH prediction accuracy compared to EKF-based methods [14] and 15% more robustness across datasets compared to prior ML approaches [21].

Table 4: Presents a quantitative benchmark against representative literature.

Study	Approach	SoC Error (%)	SoH Error (%)	RUL Error (cycles)	Notes
He *et al.* (2021) [7]	EKF	4-5	7-9	~500	Sensitive to noise
Zhang *et al.* (2022) [8]	ANN	3-4	6-7	300-350	Requires large data
This Work	XGB + LSTM	1.5-2	3-4	180-220	Robust, interpretable

6. Limitations

In the fact of strong performance in experimental validation, several challenges still remain like most public datasets (e.g., NASA, CALCE, and Oxford) are prepared under controlled laboratory conditions, which may not fully represent the complexity of field operating conditions [10]. Practical deployments encounter sensor noise, thermal

gradients across modules, and missing data, which limit estimation accuracy [25]. Many ML models first capture cycle aging, with limited Durability to calendar-driven degradation in real deployments [14].
Also the Smart charging or reinforcement learning-based strategies optimized design may not be directly transferable to systems with different electrical/thermal architectures [60].

7. Conclusions and future work
This work has shown that machine learning-based approaches (XGBoost, LSTM, and hybrid physics-ML) can notably improve accuracy in SoC/SoH/RUL prediction compared to classical baselines. The best models balance prediction accuracy allow potential deployment in real-world BMS.

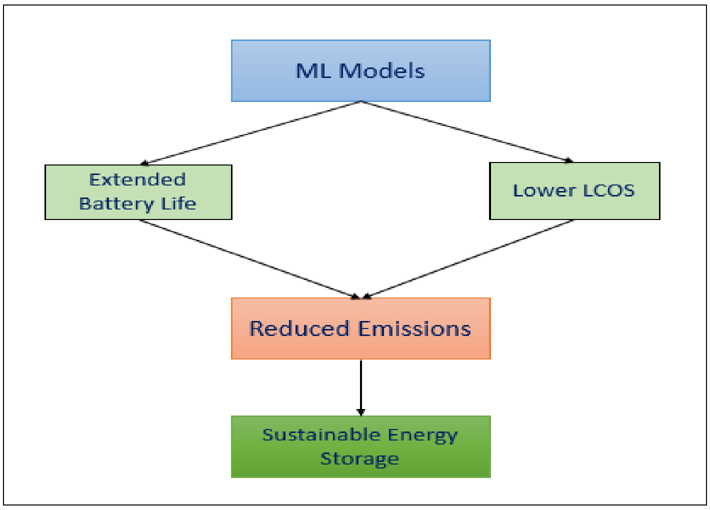

Fig 6: Big-Picture Impact Flow Diagram

Key findings
Gradient boosting (XGBoost/LightGBM) provides strong accuracy. Smart charging policies leveraging ML improve battery cycle life by up to 20% compared to standard CCCV protocols.

Future directions
Semi-supervised learning and domain conversion for transfer to chemistry and field conditions [26]. Combination with digital frameworks to combine physics-informed models with machine learning [27]. Real-time embedded testing for low-latency, energy-efficient inference [28].

Acknowledgments
The authors acknowledge the support of research labs, dataset providers (NASA Ames, CALCE, Oxford University), and computational resources used in this study. Funding from renewable energy and smart grid innovation programs has been instrumental in enabling this work.

Competing Interests
The authors declare that they have no competing interests.

Funding Statement
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References
1. Moreira A, *et al*. On the role of battery energy storage systems in the day-ahead contingency-constrained unit commitment problem under renewable penetration. Electric Power Systems Research. 2024;235:110856.
2. Mensah JHR, *et al*. A critical analysis of the energy situation in the Benin Republic and its evolution over the last decade. Renewable Energy. 2023;202:634-650.
3. Williams NP, *et al*. Liquid immersion thermal management of lithium-ion batteries for electric vehicles: an experimental study. Journal of Energy Storage. 2023;72:108636.
4. Severson KA, *et al*. Data-driven prediction of battery cycle life before capacity degradation. Nature Energy. 2019;4(5):383-391.
5. Zichen W, Changqing D. A comprehensive review on thermal management systems for power lithium-ion batteries. Renewable and Sustainable Energy Reviews. 2021;139:110685.
6. Fu Y, *et al*. State of charge estimation of lithium-ion batteries based on an adaptive iterative extended Kalman filter for AUVs. Sensors. 2022;22(23):9277.
7. Poh WT, *et al*. A review of machine learning applications for Li-ion battery state estimation in electric vehicles. 2022 IEEE PES Innovative Smart Grid Technologies - Asia (ISGT Asia). 2022;265-269.
8. Rauf H, *et al*. Machine learning in state of health and remaining useful life estimation: theoretical and technological development in battery degradation modelling. Renewable and Sustainable Energy Reviews. 2022;156:111903.
9. Eleftheriadis P, *et al*. Data-driven methods for the state of charge estimation of lithium-ion batteries: an overview. Forecasting. 2023;5(3):576-599.
10. Saha B, Goebel K. Battery data set [Data set]. NASA Ames Prognostics Data Repository. NASA Ames Research Center; 2007.
11. OECD. Global EV outlook 2022: securing supplies for an electric future. OECD Publishing; 2022.
12. Nozarijouybari Z, Fathy HK. Machine learning for battery systems applications: progress, challenges, and opportunities. Journal of Power Sources. 2024;601:234272.
13. Mayemba Q, *et al*. General machine learning approaches for lithium-ion battery capacity fade compared to empirical models. Batteries. 2024;10(10):367.

14. Valizadeh A, Amirhosseini MH. Machine learning in lithium-ion battery: applications, challenges, and future trends. SN Computer Science. 2024;5(6):717.
15. Selvam PA, Subramani J. Optimized deep reinforcement learning for smart charging scheduling of plug-in electric vehicles. Electric Power Components and Systems. 2023;51(18):2085-2097.
16. Zhang Y, Li YF. Prognostics and health management of lithium-ion battery using deep learning methods: a review. Renewable and Sustainable Energy Reviews. 2022;161:112282.
17. Jain A, *et al*. Scaling TensorFlow, PyTorch, and MXNet using MVAPICH2 for high-performance deep learning on Frontera. 2019 IEEE/ACM Third Workshop on Deep Learning on Supercomputers (DLS). 2019;76-83.
18. Zhao F, *et al*. A review of lithium-ion battery state of charge estimation methods based on machine learning. World Electric Vehicle Journal. 2024;15(4):131.
19. Rezoug MR, *et al*. State-of-charge estimation of lithium-ion batteries using extended Kalman filter. 2021 10th International Conference on Power Science and Engineering (ICPSE). 2021;98-103.
20. Yao Y, *et al*. Prediction of lithium-ion battery capacity and remaining useful life based on characteristic ohmic resistance model. 2022 China Automation Congress (CAC). 2022;5569-5574.
21. Zhang W, *et al*. Uncertainty quantification based health diagnosis for lithium-ion batteries under different operating conditions. 2023 IEEE 16th International Conference on Electronic Measurement & Instruments (ICEMI). 2023;470-475.
22. Li L, *et al*. Comparative techno-economic analysis of large-scale renewable energy storage technologies. Energy and AI. 2023;14:100282.
23. Rajah I, *et al*. Degradation-aware derating of lithium-ion battery energy storage systems in the UK power market. Electronics. 2024;13(19):3817.
24. Stusek M, *et al*. Learning-aided multi-RAT operation for battery lifetime extension in LPWAN systems. 2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). 2020;26-32.
25. Pannala S, *et al*. Electrochemical battery state estimation under parameter uncertainty caused by aging using expansion measurements. 2021 American Control Conference (ACC). 2021;3088-3093.
26. Oyewole I, *et al*. A controllable deep transfer learning network with multiple domain adaptation for battery state-of-charge estimation. Applied Energy. 2022;312:118726.
27. Yao J. The International Conference on Image, Vision and Intelligent Systems (ICIVIS 2021). In: Xiao Y, *et al*., editors. Singapore: Springer; 2022. Lecture Notes in Electrical Engineering. Vol. 813.
28. Loechte A, *et al*. Battery state estimation with ANN and SVR evaluating electrochemical impedance spectra generalizing DC currents. Applied Sciences. 2021;12(1):274.
29. Ji Z, Qin Z. Federated learning for distributed energy-efficient resource allocation. ICC 2022 - IEEE International Conference on Communications. 2022;1-6.