Original Paper
Eletrônica de Potência
Open Journal of Power Electronics

Received March 12, 2026; Received in revised form June 22, 2026, July 23, 2026; Accepted August 10, 2026; Date of publication August 24, 2026.
The review of this paper was arranged by Associate Editor Janaína G. de Oliveira and Editor-in-Chief Allan F. Cupertino.
Digital Object Identifier http://doi.org/10.18618/REP.e202626

Real-World SoC Estimation for Lithium-Ion Batteries in Electric Buses: Sampling Resolution and Machine Learning Efficiency

Johanna P. Orellana-Iñiguez[1,*], **Waldyr L. R. Gallo**[2], **Madson C. de Almeida**[1]

[1] University of Campinas – UNICAMP, Department of Systems and Energy, Campinas, SP, Brazil.
[2] University of Campinas – UNICAMP, Department of Energy, Campinas, SP, Brazil.

e-mail: j249290@dac.unicamp.br*; gallo@fem.unicamp.br; madsonca@unicamp.br
*Corresponding author.

ABSTRACT Accurate State of Charge (SoC) estimation is essential for safe and efficient lithium-ion battery operation in electric mobility. Although machine learning methods achieve high predictive capability, many studies rely on laboratory cycling data and overlook deployment constraints such as latency and embedded hardware. This paper presents a deployment-oriented evaluation of data-driven SoC estimation using large-scale operational data from a battery electric bus. The dataset includes more than 4.3 million field measurements of current, voltage, temperature, and SoC collected under realistic driving conditions. Five models are compared: Random Forest, LightGBM, XGBoost, Temporal Convolutional Networks (TCN), and Gated Recurrent Unit (GRU) networks. Models are evaluated across sampling intervals from 30 s to 210 s and multiple training-data fractions, considering accuracy, training time, and inference latency. Results show that temporal resolution strongly affects the accuracy-efficiency trade-off. The best configuration, LightGBM at 180 s using 75% of training data, achieved a test MAE of 6.338, RMSE of 8.495, and median inference latency of 1.10 ms. Compared with the Random Forest baseline at 30 s, it reduced MAE by 34.4%, RMSE by 35.6%, and latency by 97.6%, supporting efficient real-time battery management system deployment.

KEYWORDS Battery electric bus, Battery management systems, Embedded systems, Lithium-ion batteries, State of charge estimation.

I. INTRODUCTION

The accelerating transition toward low-carbon energy systems has driven an unprecedented integration of variable renewable energy sources into modern power grids [1]. As centralized and dispatchable generation is progressively replaced by decentralized and weather-dependent resources, power systems face increased uncertainty, operational variability, and new technical demands related to frequency regulation, reserve management, and voltage stability [2]–[4]. In parallel with these changes, electric vehicles (EVs) have become one of the most promising pathways for decarbonizing the transportation sector, supported by ambitious policy initiatives and rapid advances in battery manufacturing [5], [6]. A particularly relevant dimension of this transition occurs in urban environments, where electric mobility has expanded beyond private cars to include buses, taxis, and shared transportation fleets. Public transportation electrification plays a crucial role in reducing local emissions, improving air quality, and lowering operating costs, although it also imposes new challenges related to charging infrastructure, fleet management, and real-time battery monitoring [7], [8]. Within this broader context, energy storage technologies play a pivotal role in mitigating renewable intermittency, enabling energy arbitrage, and supporting the reliable operation of both mobile and stationary applications [9].

Lithium-ion batteries (LIBs) have emerged as the dominant storage technology due to their high gravimetric and volumetric energy density, rapid cost decline, and strong performance in diverse operational environments [10], [11]. Continuous progress in electrode materials, cell design, and manufacturing processes has led to safer chemistries, improved cycle life, and energy densities surpassing 300 Wh/kg. These characteristics have made LIBs indispensable for EVs, renewable energy integration, and grid-support applications.

A cornerstone of safe and efficient battery operation is the accurate estimation of the State of Charge (SoC), which represents the remaining charge relative to nominal capacity. SoC directly influences range prediction, power availability, charging strategies, and safety limits. Inaccurate SoC estimation can accelerate degradation, reduce usable energy, and under extreme conditions lead to critical safety events such as lithium plating or over-potential instabilities [12], [13].

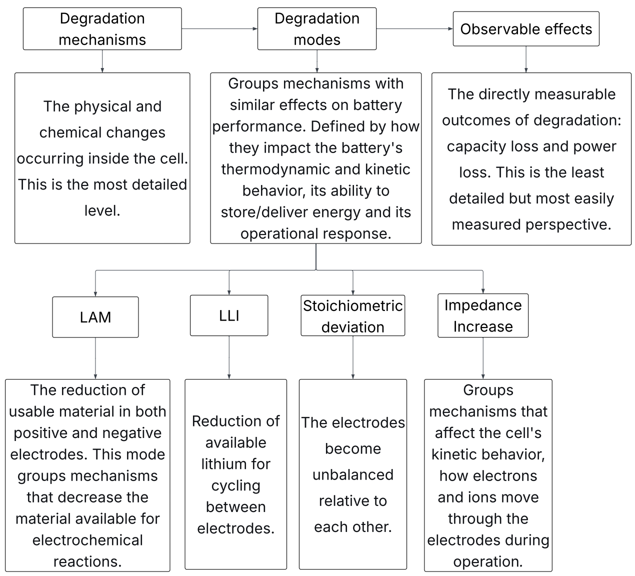

FIGURE 1. Conceptual hierarchy of lithium-ion battery degradation.

Degradation is governed by multiphysics processes (electrochemical reactions, interfacial transport, and mechanical effects). Although SoC is a state variable at a different abstraction level, operating conditions associated with extreme SoC ranges correlate with accelerated aging, as illustrated in Fig. 1. At the fundamental scale, degradation mechanisms (e.g., solid electrolyte interphase (SEI) formation, lithium plating, and structural fatigue) give rise to well-established degradation modes, including loss of lithium inventory (LLI), loss of active material (LAM), and impedance increase.

Ultimately, these modes manifest as observable macroscopic effects such as capacity fade and power loss [14], [15]. Operation at high SoC may promote electrolyte oxidation and SEI growth, whereas prolonged low-SoC operation or deep discharge events may contribute to accelerated degradation depending on cell chemistry, operating history, and voltage limits. For these reasons, precise and continuous SoC tracking plays an important role in enabling degradation-aware Battery Management Systems (BMS) by supporting control strategies that avoid harmful operating conditions and extend the operational lifetime of lithium-ion batteries.

Traditional SoC estimation techniques (ranging from electrochemical models to equivalent-circuit-based observers) encounter limitations when applied to complex, dynamic, and aging-affected environments [16]. This has motivated the rapid adoption of machine learning (ML) methods, which can learn nonlinear mappings between measurable quantities (current, voltage, temperature) and internal states without requiring detailed physicochemical models [17]. Kernel-based regressors such as Support Vector Machine (SVM) models and Gaussian Process Regression (GPR) [18], as well as ensemble models such as Random Forest, XGBoost, and LightGBM, have shown strong predictive capability and robustness to noisy measurements [19]. Optimization frameworks (including evolutionary algorithms and Bayesian hyperparameter search) further enhance these methods' adaptability and performance [12].

Despite significant advances in ML-based SoC estimation, three persistent gaps limit practical deployment in real electric bus applications [20], [21]. First, most studies rely on laboratory datasets under controlled conditions (predefined charge-discharge cycles and thermal regulation) [19], [22]–[25]. While suitable for model development, these datasets do not capture the stochastic variability, partial SoC operation, and dynamic loads of real-world bus operation. Consequently, the transferability of reported accuracy to field conditions remains uncertain.

Second, among studies using real-world data, the impact of temporal sampling resolution on the accuracy-efficiency trade-off remains underexplored. In practical BMS, the sampling interval directly affects processing load, memory requirements, and communication constraints. However, most existing works adopt fixed or single temporal resolutions without systematically analyzing their impact on both estimation accuracy and computational efficiency [26], [27]. As a result, there is limited quantitative guidance on how to select an appropriate sampling rate for embedded deployment [21].

Third, although temporal modeling techniques such as sliding window approaches and deep learning architectures are capable of capturing sequential dependencies, their computational requirements are rarely analyzed in a systematic manner. In particular, existing studies often focus on prediction accuracy without providing a clear assessment of training cost and inference latency, which are critical factors for real-time embedded BMS applications.

To position the present work within the SoC estimation literature, Table 1 compares representative studies in terms of data source, system scale, reference SoC definition, validation methodology, sampling strategy, and deployment-oriented evaluation criteria. Direct comparison of prediction errors across studies is often difficult because datasets differ substantially in battery chemistry, operating conditions, data acquisition procedures, and validation protocols. Therefore, the objective of this comparison is not to benchmark accuracy values directly, but rather to highlight methodological differences that affect practical deployment and result interpretation.

The comparison reveals that most existing studies rely on laboratory datasets collected under controlled conditions and fixed sampling rates. Furthermore, deployment-related aspects such as inference latency, computational cost, temporal validation, and robustness-oriented metrics are rarely reported. In contrast, the present work focuses on real electric-bus telemetry and explicitly evaluates the trade-off between estimation accuracy and computational efficiency across multiple temporal resolutions.

TABLE 1. Positioning of representative SoC estimation studies with respect to data realism, validation methodology, and deployment-oriented evaluation.

Study	Data source	Scale	Reference SoC	Validation	Sampling analysis	Real-world data	Deployment metrics	Main limitation
How et al. (2020) [28]	CALCE laboratory dataset	Cell	Coulomb counting	Cross-drive-cycle evaluation	Fixed (1 Hz)	No	No	Laboratory conditions and cell-level validation only
Deng et al. (2020) [23]	Experimental battery-pack test bench	Pack	Experimental reference	Train–test split	Fixed (1 Hz)	No	No	Controlled operating conditions and fixed sampling rate
Chandran et al. (2021) [19]	Panasonic 18650PF laboratory dataset	Cell	Dataset-provided SoC	Train–validation–test split	Fixed	No	No	Laboratory dataset with limited deployment considerations
Pranav et al. (2024) [29]	Real EV telemetry	Vehicle	BMS-reported SoC	Random 60/40 split	Fixed	Yes	No	No temporal validation or computational-efficiency assessment
This work	Real electric-bus telemetry	Vehicle/Pack	BMS-reported SoC	Chronological 2021→2022 split	30–210 s	Yes	Training time, inference latency	Single-vehicle study

To address these gaps, this paper presents a systematic, deployment-oriented evaluation of ML-based SoC estimation using a large-scale real-world dataset from a battery electric bus. This work extends the conference version presented in [30] by incorporating additional experimental analyses, a broader evaluation scope, and a more detailed discussion oriented toward real-world deployment constraints. The proposed framework quantifies the accuracy-efficiency trade-off across multiple models, sampling intervals, and data availability scenarios, reporting both prediction performance and inference latency on a central processing unit (CPU)-based platform.

The core novelty of this work does not lie in proposing a new machine learning architecture, but in formulating SoC estimation as a deployment-oriented multi-objective evaluation problem. Specifically, the proposed framework jointly analyzes prediction accuracy, sampling resolution, training-data availability, training cost, and inference latency using large-scale real-world electric-bus telemetry.

Unlike prior studies that commonly evaluate SoC estimators under fixed sampling rates and mainly report prediction error, this work explicitly investigates how temporal resolution and data availability affect both estimation performance and computational feasibility under realistic BMS constraints. By doing so, it provides practical guidelines for selecting model architectures and sampling strategies that balance accuracy and efficiency in embedded battery management systems.

The main contributions of this work are summarized as follows:

- A deployment-oriented evaluation framework that jointly assesses prediction accuracy, temporal sampling resolution, training-data availability, training time, and inference latency under practical BMS constraints, explicitly treating the sampling interval as a design variable in the accuracy-efficiency trade-off.
- A large-scale real-world validation protocol based on more than 4.3 million measurements collected from an electric bus, using a chronological 2021–2022 split to preserve temporal consistency, prevent information leakage, and assess out-of-period generalization under real-world temporal variability.
- A controlled multi-resolution benchmark of five heterogeneous machine learning architectures across sampling intervals from 30 s to 210 s under a unified experimental pipeline, enabling systematic characterization of the relationship between temporal resolution, predictive performance, and computational requirements.
- A data-efficiency and deployment assessment that systematically evaluates multiple training-data fractions (25%, 50%, 75%, and 100%) together with computational metrics, supporting evidence-based selection of model and sampling configurations that balance predictive accuracy and resource constraints in practical BMS deployment scenarios.

II. METHODOLOGY

A. Experimental Platform and Dataset Collection

The SoC estimation framework developed in this study relies on real-world operational data collected from a battery electric bus (BEB) operating at the University of Campinas (UNICAMP) [30]. The vehicle is part of the Sustainable Campus Living Laboratory initiative [31], which promotes research on electric mobility and integrated energy management systems in a real operational environment. This platform enables large-scale battery data collection under realistic driving conditions. The specific BEB unit used for data acquisition is illustrated in Fig. 2.

FIGURE 2. Battery electric bus deployed for field data collection.

Operational data were collected during the vehicle's scheduled service on a predefined circular route within the university campus. The daily operating schedule consists of approximately 20 individual trips, each lasting about 30 minutes and covering an average distance of nearly 10 kilometers. The bus is powered by a lithium iron phosphate ($LiFePO_4$) battery pack and is recharged at a dedicated charging station supplied by photovoltaic generation. An integrated onboard monitoring system continuously records key electrical and thermal variables at a sampling frequency of 10 Hz (corresponding to a 100 ms sampling interval).

The reference SoC used in this study was directly obtained from the onboard BMS of the electric bus. It is important to note that the BMS-reported SoC is itself an estimated quantity and may contain systematic errors. Consequently, the proposed models are effectively trained to reproduce the BMS-reported SoC rather than a directly measurable electrochemical ground truth.

This limitation is inherent to real-world battery datasets, where direct measurement of the true SoC is generally not feasible during normal operation. Similar assumptions are commonly adopted in field-based studies, where BMS estimates serve as the only available reference.

Therefore, the objective of this work is not to recover the exact electrochemical SoC, but rather to develop accurate and computationally efficient estimators consistent with real-world BMS operation. This perspective ensures that the proposed approach remains directly applicable to practical deployment scenarios.

The data acquisition period extended from March 2021 to July 2022, as illustrated in Fig. 3, resulting in an initial raw dataset exceeding 33 million measurements. To reflect typical operational conditions, a filtering procedure was applied to remove records corresponding to non-service days (weekends and holidays) and periods outside regular operation (before 07:00 and after 23:00). After preprocessing, the refined dataset used for analysis contained approximately 4.3 million samples, including 2.63 million records from 2021 and 1.69 million from 2022.

It is important to note that the dataset used in this study is derived from a single electric bus operating under specific conditions. Therefore, direct generalization of the results to other vehicles or battery systems cannot be guaranteed.

Nevertheless, the dataset captures a wide range of real-world operating conditions over an extended period. Moreover, the proposed framework is model-agnostic and can be applied to other datasets. Future work will explore validation across different systems to further assess generalizability.

The analytical approach for this study is organized through a sequential framework, graphically depicted in Fig. 4.

B. Data Preparation and Conditioning

A data preparation pipeline was implemented to transform raw operational records into a dataset suitable for computational modeling. First, measurements corresponding to non-operational periods were removed to isolate representative battery usage cycles. To enable multi-resolution analysis,

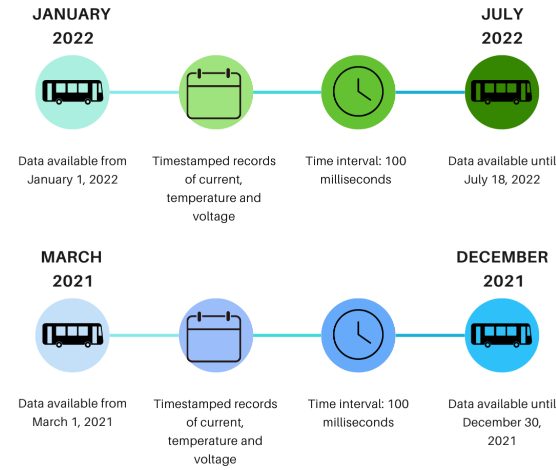

FIGURE 3. Timeline summarizing the availability and characteristics of the real-world battery dataset used in this study.

the dataset was resampled at intervals ranging from 30 s to 210 s using mean aggregation over non-overlapping temporal windows. Missing values in the resampled series were subsequently handled through linear interpolation. This procedure enabled systematic evaluation of how temporal resolution influences model accuracy and computational efficiency.

C. Dataset Partitioning and Feature Engineering

The dataset was partitioned using a temporal split strategy, where records from 2021 were used for model training, while data from 2022 were reserved for validation and testing. This preserves the chronological structure of the measurements and prevents information leakage between training and evaluation phases.

Temporal dependencies were captured using a sliding-window representation. Each input vector F_i contains a sequence of historical sensor measurements, while the corresponding target f_i represents the SoC to be estimated.

Feature selection was guided by exploratory data analysis and domain knowledge. The variables most relevant for SoC estimation were current (I), voltage (V), and battery temperature (T). Variables with a high proportion of missing values or weak correlation with the target variable were excluded to improve generalization and reduce computational complexity.

To assess data efficiency, the training procedure was repeated using four fractions of the available training set (25%, 50%, 75%, and 100%). This allows evaluating how model performance evolves with data availability and identifying the minimum training fraction required to achieve competitive estimation accuracy.

D. Model Development and Hyperparameter Optimization

Five machine learning models were implemented for comparative evaluation: three ensemble-based approaches (Random Forest (RF), XGBoost, and LightGBM) and two temporal deep learning architectures (Temporal Convolutional Net-

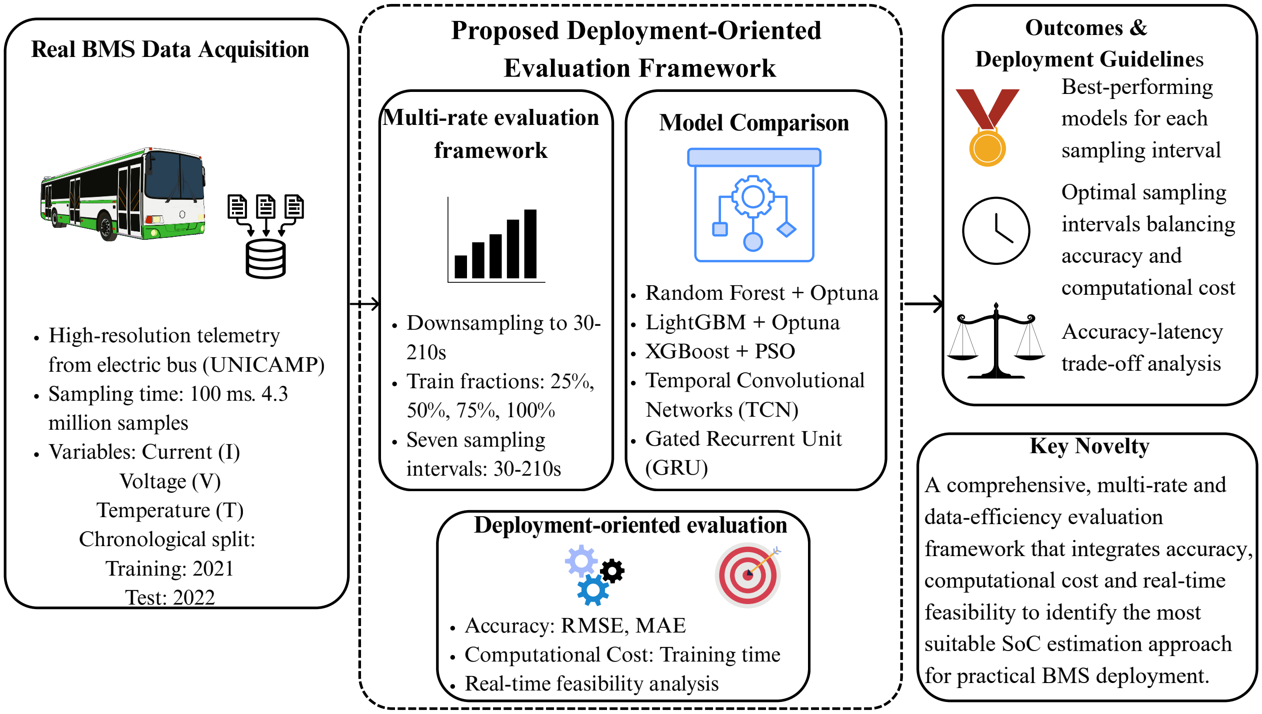

FIGURE 4. Deployment-oriented framework for ML-based SoC estimation using real-world electric bus data and practical BMS constraints.

works and Gated Recurrent Unit networks). These models were selected to capture both static and temporal learning capabilities relevant to battery SoC estimation. All experiments were developed in Python under a unified implementation framework.

Hyperparameter tuning was performed using a validation-based strategy. Optuna was employed to optimize Random Forest and LightGBM configurations, while Particle Swarm Optimization (PSO) was applied to XGBoost. For the deep learning models, key architectural and training parameters were adjusted, and early stopping based on validation loss was used to prevent overfitting. This procedure ensures a fair and consistent comparison across all evaluated models. All experiments were conducted on a CPU-based system (Intel Core i7-1065G7, 8 GB RAM) without graphics processing unit (GPU) acceleration.

E. Machine Learning Model Formulation

This subsection summarizes the mathematical foundations of the machine learning models evaluated in this study. The formulations focus on the core operational mechanisms of each method rather than exhaustive derivations, which is sufficient to support the comparative analysis of SoC estimation accuracy and computational efficiency under realistic operating conditions.

Random Forest (RF). Random Forest is an ensemble learning method that aggregates multiple decision trees trained on bootstrap samples of the dataset while performing random feature selection at each split. This randomness reduces correlation among trees and improves generalization. The final prediction is obtained by averaging the outputs of all trees:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} f_t(\mathbf{x}), \tag{1}$$

where $f_t(\mathbf{x})$ denotes the prediction of the t-th tree for input vector $\mathbf{x}$, and T is the number of trees. Its non-parametric nature and robustness to noise make RF well suited for battery datasets characterized by measurement uncertainty and heterogeneous operating conditions.

Gradient boosting models (XGBoost and LightGBM). Gradient boosting methods build predictive models sequentially, where each new regression tree is trained to correct the residual errors of the current ensemble. The additive formulation can be written as

$$\hat{y}^{(t)} = \hat{y}^{(t-1)} + \eta f_t(\mathbf{x}), \tag{2}$$

where $f_t(\mathbf{x})$ represents the tree added at iteration t and η is the learning rate controlling the contribution of each learner. XGBoost introduces explicit regularization to limit model complexity, whereas LightGBM improves computational efficiency through a leaf-wise tree growth strategy with depth constraints. Both approaches effectively capture nonlinear relationships between electrical variables and SoC while maintaining fast inference suitable for embedded systems.

Temporal Convolutional Network. Temporal Convolutional Networks model sequential data using causal and dilated one-dimensional convolutions. Causality ensures that predictions at time t depend only on present and past inputs, while dilation expands the receptive field without increasing network depth. A simplified convolutional operation is given by

$$\mathbf{h}_t = \sum_{k=0}^{K-1} \mathbf{W}_k \, \mathbf{x}_{t-d\cdot k}, \tag{3}$$

where $\mathbf{h}_t$ denotes the hidden representation at time t, $\mathbf{W}_k$ are convolutional filters, K is the kernel size, and d is the dilation factor. By avoiding recurrent connections, TCNs enable parallel computation across time steps while capturing long-term temporal dependencies in battery time-series data.

Gated Recurrent Unit. GRU networks are recurrent neural architectures that model temporal dynamics through gated state updates. The hidden state evolution is defined as

$$\mathbf{h}_t = (1 - \mathbf{z}_t)\mathbf{h}_{t-1} + \mathbf{z}_t \tilde{\mathbf{h}}_t \tag{4}$$

where $\mathbf{z}_t$ is the update gate and $\tilde{\mathbf{h}}_t$ is the candidate hidden state. This mechanism enables the network to selectively retain or update past information, effectively modeling sequential battery behavior under varying load conditions while maintaining moderate computational complexity.

Overall, ensemble methods emphasize robustness and low inference latency, whereas temporal deep learning models explicitly exploit sequential structure, providing complementary perspectives for evaluating both accuracy and real-time feasibility in battery management systems.

III. RESULTS AND DISCUSSION

Results are analyzed from a deployment-oriented perspective, where predictive accuracy alone is insufficient to determine model suitability. Therefore, both training cost and inference latency are evaluated to identify architectures compatible with real-time BMS deployment.

Table 2 summarizes the predictive accuracy, measured using mean absolute error (MAE) and root mean square error (RMSE), together with computational metrics including training time and median inference latency (50th percentile, p50) across different sampling intervals. Lower MAE and RMSE indicate higher prediction accuracy, while training time and latency represent the computational cost associated with model deployment. All models in Table 2 were trained using the full training dataset (100% of the available training samples).

As it can be observed in Table 2, for the 30 s sampling interval, deep learning models, particularly the TCN, achieve competitive accuracy (MAE = 7.913, RMSE = 10.901). However, this improvement is accompanied by a substantially higher computational burden, with training times exceeding 2500 seconds. In contrast, tree-based models, particularly LightGBM, exhibit a slightly higher prediction error but require considerably shorter training times, highlighting their computational efficiency.

When the sampling interval increases to 60 s, the difference in predictive performance between deep learning and tree-based approaches becomes smaller. Gradient boosting methods show more consistent behavior, with XGBoost+PSO achieving competitive accuracy while maintaining very low training cost (133 s). This indicates that boosting-based models are able to exploit the additional temporal context provided by longer sampling intervals without incurring the computational overhead associated with deep architectures.

At the 90 s sampling interval, XGBoost+PSO achieves the lowest test MAE (7.506) and the shortest training time (72 s), while TCN attains the lowest RMSE (10.031). Although TCN remains competitive in predictive accuracy, its substantially higher training cost limits its practicality for scalable or embedded applications. These results highlight a trade-off between predictive performance and computational cost at this temporal resolution.

TABLE 2. Comparison of prediction accuracy, training cost, and inference latency across different sampling intervals using the full training dataset (100%).

Sampling interval	Model	$\mathrm{MAE}_{\mathrm{test}}$	$\mathrm{RMSE}_{\mathrm{test}}$	Training time [s]	Latency p50 [ms]
30 s	RF	9.667	13.182	2047	46.74
	LGBM	8.619	12.113	920	1.74
	XGB	9.095	12.397	**186**	**0.66**
	TCN	7.913	10.901	2523	3.01
	GRU	9.008	11.868	3898	5.11
60 s	RF	7.881	11.499	895	31.16
	LGBM	7.984	11.236	687	1.82
	XGB	7.958	11.211	**133**	**0.41**
	TCN	7.680	10.542	3270	1.71
	GRU	12.032	14.533	927	2.99
90 s	RF	7.811	11.135	553	31.62
	LGBM	7.733	10.900	439	0.99
	XGB	**7.506**	**10.492**	**72**	**0.38**
	TCN	7.645	10.031	1102	1.92
	GRU	12.276	14.699	481	3.30
120 s	RF	7.006	9.609	538	46.84
	LGBM	7.296	10.014	518	1.41
	XGB	**6.920**	**9.527**	**73**	**0.41**
	TCN	8.326	11.240	1495	2.22
	GRU	7.512	10.726	1646	4.40
150 s	RF	7.086	9.531	251	46.87
	LGBM	**6.438**	8.861	230	0.92
	XGB	6.461	**8.694**	**46**	**0.33**
	TCN	7.156	9.923	896	1.81
	GRU	12.113	14.588	281	2.96
180 s	RF	7.075	9.641	268	33.43
	LGBM	**6.452**	**8.764**	298	1.13
	XGB	6.495	8.872	**44**	**0.34**
	TCN	7.248	10.104	620	2.29
	GRU	12.072	14.563	242	3.64
210 s	RF	6.949	9.395	321	89.87
	LGBM	**6.492**	**8.930**	569	2.95
	XGB	6.959	9.333	**72**	**0.73**
	TCN	6.882	9.974	2767	5.73
	GRU	12.022	14.535	519	8.21

RF: Random Forest optimized with Optuna; LGBM: LightGBM optimized with Optuna; XGB: XGBoost optimized with PSO.

For longer sampling intervals (120 s to 210 s), the differences between models become more nuanced. LightGBM frequently achieves the lowest MAE values (for instance, 6.438 at 150 s and 6.452 at 180 s), while XGBoost+PSO consistently maintains the lowest training times. These results indicate that boosting-based methods provide a robust balance between accuracy and computational efficiency across different temporal resolutions.

Overall, the results reveal a clear accuracy-complexity trade-off. Sampling intervals between 150 s and 180 s provide the most favorable compromise for real-time SoC estimation, combining low prediction error with reduced computational burden. A plausible statistical explanation is related to the mean-based temporal aggregation used during resampling. At finer resolutions, real-world bus telemetry retains substantial short-term variability from traction loads, regenerative braking, voltage transients, and measurement noise, while consecutive observations may remain strongly correlated. Aggregation over longer windows attenuates rapid fluctuations and reduces temporal redundancy, improving the statistical regularity of the input–target relationship while preserving the slower trends relevant to SoC evolution. Conversely, excessively coarse intervals may suppress informative dynamics and reduce the number of available training observations. The slight performance deterioration from 180 s to 210 s is consistent with this trade-off. Therefore, the 150–180 s range should be interpreted as a dataset- and model-dependent optimum rather than a universal sampling interval for lithium-ion batteries.

To evaluate data efficiency, an ablation study was conducted using four training fractions (25%, 50%, 75%, and 100%) of the available training set. The results are summarized in Fig. 5, which identifies the configuration achieving the lowest test MAE for each sampling interval across all evaluated models and training fractions. Table 3 reports the corresponding best-performing configurations, including the model, training fraction, test MAE, and inference latency.

This ablation analysis allows evaluating how model performance evolves with the amount of available training data and identifying configurations that maintain competitive accuracy with reduced training data.

TABLE 3. Best-performing configuration identified for each sampling interval

Sampling interval [s]	Model	Training fraction	$\mathrm{MAE}_{\mathrm{test}}$	Latency p50 [ms]
30	TCN	100%	7.9130	3.0100
60	TCN	100%	7.6800	1.7100
90	XGB	75%	7.3943	0.3693
120	TCN	75%	6.5313	1.9726
150	LGBM	100%	6.4380	0.9200
180	LGBM	75%	6.3382	1.1005
210	LGBM	75%	6.3868	1.2669

We acknowledge that the proposed framework was developed using a large-scale dataset (4.3 million measurements). However, as demonstrated by the ablation study shown in (Fig. 5), competitive accuracy can be achieved with reduced training data, with several configurations reaching near-optimal performance using 75% of the dataset.

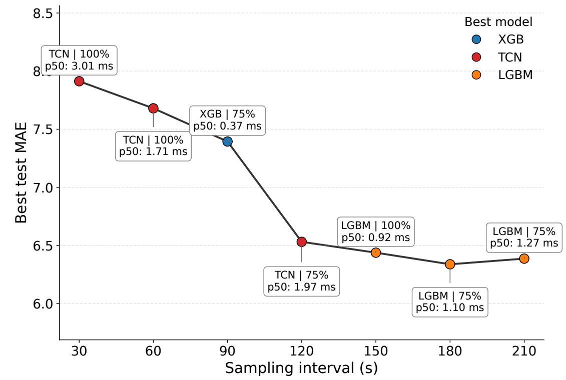

FIGURE 5. Best-performing configuration at each sampling interval. Models were trained using four training-data fractions (25%, 50%, 75%, and 100%). Each point indicates the lowest test MAE, with annotations showing the model, training fraction, and p50 inference latency.

This indicates that the proposed approach exhibits a certain degree of data efficiency and can be applied in scenarios with moderate data availability. For more limited cases, approaches such as model transfer across similar systems or the use of synthetic data may be considered as potential extensions. Nevertheless, the availability of representative operational data remains an important factor for reliable deployment.

Fig. 5 shows that deep learning models, particularly TCN, dominate the finest temporal resolutions (30–90 s). However, as the sampling interval increases, gradient boosting methods such as XGBoost and LightGBM become more competitive and achieve the best performance at coarser resolutions (150–210 s).

Another important observation is that the optimal configuration does not always require the full training dataset. In several cases, near-optimal accuracy is already achieved with 75% of the training data, indicating that boosting-based models maintain strong predictive performance even with reduced training data.

As illustrated in Fig. 6, the radar charts summarize the trade-off between prediction accuracy, training cost, and the resulting performance index across the evaluated models. The normalized performance index combines normalized MAE and training cost to provide a compact indicator of the balance between predictive accuracy and computational efficiency.

XGBoost+PSO consistently exhibits large radar areas due to its combination of competitive accuracy and extremely low training cost, whereas deep learning models achieve strong accuracy but at the expense of significantly higher computational overhead.

Beyond training cost, real-time deployment also depends on inference latency (p50) during online operation. XGBoost optimized via PSO consistently achieves the lowest p50 inference latency, maintaining sub-millisecond response times across all sampling intervals. This behavior is explained by the efficient evaluation of shallow decision trees. LightGBM

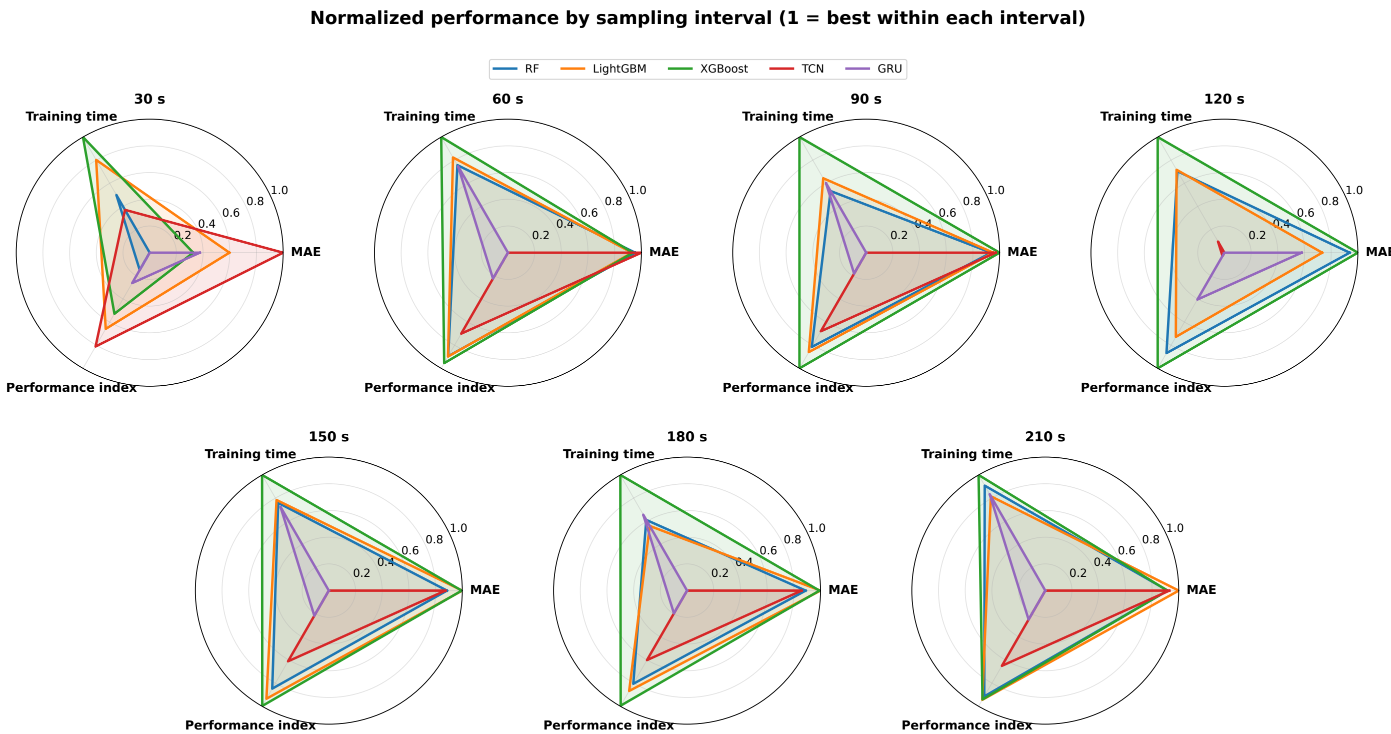

FIGURE 6. Normalized performance comparison of the evaluated models across different sampling intervals (30 s to 210 s). All metrics are normalized such that higher values indicate better performance. The radar charts jointly illustrate prediction accuracy (MAE), training cost, and the resulting performance index, highlighting the trade-off between estimation accuracy and computational efficiency across models.

also exhibits low-latency performance, while Random Forest requires higher inference time due to the sequential evaluation of multiple estimators. Deep learning models (TCN and GRU) operate within a few milliseconds, confirming real-time feasibility but with higher computational overhead.

The similarity of latency values across sampling intervals indicates that inference cost is primarily determined by model architecture rather than the sampling interval. When jointly considering predictive accuracy and inference latency, XGBoost+PSO provides the most balanced solution, combining low estimation error with sub-millisecond inference latency. This balance makes boosting-based models particularly suitable for real-time BMS applications where both estimation accuracy and computational efficiency are critical.

To further evaluate the dynamic behavior of the proposed framework, Fig. 7 presents representative time-series comparisons between the reference and estimated SoC signals for the best-performing model associated with each sampling interval. The selected window includes real-world charging and discharging events, allowing qualitative assessment of tracking capability, transient response, and possible estimation lag under realistic operating conditions. Overall, the results show that the evaluated configurations are able to capture the main SoC dynamics, with particularly consistent tracking observed at coarser sampling intervals, where LightGBM achieved the best overall performance.

To provide additional insight into model behavior under different operating conditions, the best-performing overall configuration, corresponding to LightGBM + Optuna with a 180 s sampling interval and 75% training fraction, was further evaluated across different SoC regions. Table 4 summarizes the estimation performance for four SoC intervals.

TABLE 4. Performance of the best-performing configuration across different SoC ranges.

SoC range [%]	Samples	$\mathrm{MAE}_{\mathrm{test}}$	$\mathrm{RMSE}_{\mathrm{test}}$
20–40	1094	14.855	17.084
40–60	24138	4.473	6.297
60–80	14438	7.311	8.484
80–100	7461	9.238	12.058

Results correspond to the LightGBM + Optuna model trained with 75% of the 2021 dataset and evaluated on the 2022 test set using a 180 s sampling interval.

The results show that the lowest prediction error is obtained in the 40–60% region, which also contains the largest number of samples. Higher errors are observed at high SoC values and particularly below 40% SoC. However, the 20–40% interval contains only a limited number of samples, corresponding to approximately 2.3% of the test set, suggesting that data imbalance contributes to the reduced performance in this region.

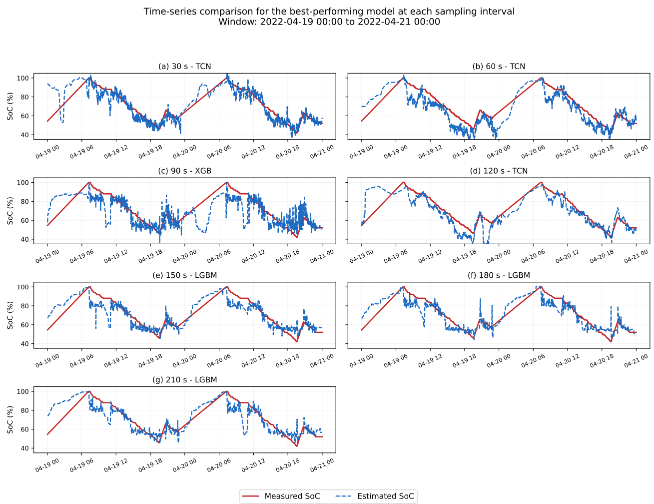

FIGURE 7. Time-series comparison between measured and estimated SoC for the best-performing model associated with each sampling interval. The selected window includes representative charging and discharging events, allowing qualitative assessment of tracking capability, transient behavior, and estimation lag under realistic operating conditions.

Overall, this analysis shows that model accuracy is not uniform across the SoC range, while remaining satisfactory in the dominant operating regions of the electric bus.

From a practical deployment perspective, the measured inference latency demonstrates that all evaluated models are capable of generating predictions within milliseconds, which is several orders of magnitude faster than the sampling intervals considered (30–210 s). In particular, XGBoost+PSO consistently operates in the sub-millisecond range, ensuring negligible computational overhead during real-time operation. These results indicate that the proposed approach is compatible with embedded BMS implementations, where estimation updates must be computed well within the available sampling period.

Battery temperature is included as an input variable due to its well-known influence on electrochemical behavior and SoC estimation accuracy. The dataset inherently captures natural temperature variations under real operating conditions, including ambient changes and load-dependent thermal dynamics.

Therefore, the reported results reflect model performance under varying thermal conditions. A dedicated sensitivity analysis with respect to temperature is not included in this study, as the primary focus is on evaluating the trade-off between estimation accuracy and computational efficiency under realistic operating scenarios.

A. Quantitative Performance Comparison

To further illustrate the trade-off between accuracy and computational efficiency, Table 5 compares a baseline model with the best-performing configuration identified in this study. The baseline corresponds to a commonly adopted setup in SoC estimation studies, consisting of a RF model trained with a high-resolution sampling interval (30 s) using 100% of the available training data.

It is important to note that this comparison reflects a deployment-oriented scenario, where sampling resolution, model architecture, and data availability are jointly optimized rather than evaluated in isolation.

TABLE 5. Relative improvement of the best configuration against a baseline. The baseline corresponds to RF (30 s, 100%), and the best configuration corresponds to LGBM (180 s, 75%).

Metric	Baseline	Best config.	Improvement
MAE test	9.667	6.338	**34.4%** ↓
RMSE test	13.182	8.495	**35.6%** ↓
Inference latency p50 [ms]	46.74	1.10	**97.6%** ↓

The results demonstrate that the appropriate joint selection of sampling resolution, model architecture, and training-data fraction can yield substantial improvements in both estimation accuracy and online computational efficiency. In particular, the optimized configuration achieves a 34.4% reduction in MAE, a 35.6% reduction in RMSE, and a 97.6% reduction in inference latency compared with the RF baseline at 30 s.

These results indicate that the proposed framework effectively identifies configurations that maintain high predictive accuracy while significantly reducing online computational burden. This balance is critical for real-time embedded BMS applications, where both estimation performance and resource constraints must be considered.

IV. CONCLUSION

This study presented a comprehensive, deployment-oriented evaluation of machine-learning-based State of Charge estimation using a large-scale real-world dataset comprising more than 4.3 million measurements collected from an electric bus operating under realistic conditions. By moving beyond controlled laboratory cycling data, the analysis captures practical load variability and operating dynamics, providing a more representative basis for assessing algorithm performance in electric vehicle applications.

Five machine learning architectures, including ensemble methods and temporal neural networks, were systematically evaluated across multiple sampling intervals and training-data fractions. The results show that sampling interval is a key design variable, as moderate temporal aggregation can preserve or even improve estimation accuracy while reducing computational cost. In this context, the most favorable trade-offs were generally observed for sampling intervals between 150 s and 180 s.

The comparative analysis also showed that model superiority is not fixed across all operating conditions. Temporal Convolutional Networks achieved the best accuracy at finer temporal resolutions, whereas gradient boosting methods became increasingly competitive at coarser sampling intervals. In particular, XGBoost optimized via PSO consistently provided low training times and sub-millisecond inference latency, while LightGBM achieved the lowest MAE in several coarse-resolution settings. These results indicate that boosting-based methods offer the most attractive balance between predictive performance and computational efficiency for real-time BMS applications.

By explicitly incorporating training cost, inference latency, and data-efficiency considerations into the evaluation framework, this work frames SoC estimation as a deployment-oriented multi-objective problem rather than a purely accuracy-driven task. Overall, the results highlight that both temporal resolution and model architecture strongly influence the practical feasibility of data-driven SoC estimators in electric mobility systems.

From a practical perspective, all evaluated models operate within the millisecond range, with XGBoost+PSO achieving sub-millisecond inference latency on a CPU-based platform. These inference times are several orders of magnitude smaller than the sampling intervals considered (30–210 s), ensuring that real-time implementation in embedded BMS is feasible with negligible computational overhead.

Furthermore, compared to conventional approaches that primarily focus on prediction accuracy, the proposed framework provides a more realistic assessment by explicitly incorporating computational constraints and deployment considerations. This enables informed model selection and system design for real-world electric mobility applications.

Overall, the results demonstrate that data-driven SoC estimation can be effectively deployed in practical BMS, provided that model architecture and sampling strategy are jointly optimized.

ACKNOWLEDGMENT

This study received financial support from FAPESP (grant #2024/23796-0) and from Fundação de Desenvolvimento da Pesquisa – FUNDEP MOVER, Linha VI (grant 2927109). The authors also acknowledge the support from the Centro Paulista de Estudos da Transição Energética (CPTEn).

AUTHOR'S CONTRIBUTIONS

J.P.ORELLANA-IÑIGUEZ: Conceptualization, Data Curation, Formal Analysis, Methodology, Visualization, Writing – Original Draft, Writing – Review & Editing. **W.L.R.GALLO:** Writing – Review & Editing. **M.C.ALMEIDA:** Conceptualization, Formal Analysis, Supervision, Writing – Review & Editing.

PLAGIARISM POLICY

This article was submitted to the similarity system provided by Crossref and powered by iThenticate – Similarity Check.

DATA AVAILABILITY

The data used in this research is available in the body of the document.

REFERENCES

[1] R. Farheen, K. Fatima, "Integration of Renewable Energy Sources into Modern Power Grids: Challenges and Solutions", *Journal of Electrical Power System Engineering and Technology*, vol. 10, no. 2, 2025.

[2] M. A. Mendieta Sarmiento, J. P. Orellana Iñiguez, H. E. Hernandez-Figueroa, O. Probst, L. I. Minchala Ávila, "Reducing Variability in Wind Power Generation by Combining Slope Control with SOC Feedback", *in 2024 IEEE PES Generation, Transmission and Distribution Latin America Conference and Industrial Exposition (GTDLA)*, pp. 1–6, 2024, doi:10.1109/GTDLA61236.2024.10913718.

[3] C. Gallego-Castillo, A. Cuerva-Tejero, O. Lopez-Garcia, "A review on the recent history of wind power ramp forecasting", *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1148–1157, 2015, doi:10.1016/j.rser.2015.07.154.

[4] R. C. de Barros, W. C. S. Amorim, W. do Couto Boaventura, A. F. Cupertino, V. F. Mendes, H. A. Pereira, "Methodology for BESS Design Assisted by Choice Matrix Approach", *Eletrônica de Potência*, vol. 29, pp. e202412–e202412, 2024, doi:10.18618/REP.2005.1.019027.

[5] International Energy Agency, "Global Critical Minerals Outlook 2024", Licence: CC BY 4.0, 2024, URL: https://www.iea.org/reports/global-critical-minerals-outlook-2024.

[6] A. L. Rodrigues, S. R. Seixas, "Battery-electric buses and their implementation barriers: Analysis and prospects for sustainability", *Sustainable Energy Technologies and Assessments*, vol. 51, p. 101896, 2022, doi:10.1016/j.seta.2021.101896.

[7] R. Wang, Y. Wu, W. Ke, S. Zhang, B. Zhou, J. Hao, "Can propulsion and fuel diversity for the bus fleet achieve the win–win strategy of energy conservation and environmental protection?", *Applied Energy*, vol. 147, pp. 92–103, 2015, doi:10.1016/j.apenergy.2015.01.107.

[8] J.-M. Clairand, P. Guerra-Terán, X. Serrano-Guerrero, M. González-Rodríguez, G. Escrivá-Escrivá, "Electric vehicles for public transportation in power systems: A review of methodologies", *Energies*, vol. 12, no. 16, p. 3114, 2019, doi:10.3390/en12163114.

[9] M. Y. Suberu, M. W. Mustafa, N. Bashir, "Energy storage systems for renewable energy power sector integration and mitigation of intermittency", *Renewable and Sustainable Energy Reviews*, vol. 35, pp. 499–514, 2014, doi:10.1016/j.rser.2014.04.009.

[10] J. W. Choi, D. Aurbach, "Promise and reality of post-lithium-ion batteries with high energy densities", *Nature reviews materials*, vol. 1, no. 4, pp. 1–16, 2016, doi:10.1038/natrevmats.2016.13.

[11] M. Kabir, D. E. Demirocak, "Degradation mechanisms in Li-ion batteries: a state-of-the-art review", *International Journal of Energy Research*, vol. 41, no. 14, pp. 1963–1986, 2017, doi:10.1002/er.3762.

[12] N. Ghaeminezhad, Q. Ouyang, J. Wei, Y. Xue, Z. Wang, "Review on state of charge estimation techniques of lithium-ion batteries: A control-oriented approach", *Journal of Energy Storage*, vol. 72, p. 108707, 2023, doi:10.1016/j.est.2023.108707.

[13] J. Jaguemont, F. Bardé, "A critical review of lithium-ion battery safety testing and standards", *Applied Thermal Engineering*, vol. 231, p. 121014, 2023, doi:10.1016/j.applthermaleng.2023.121014.

[14] C. A. Rufino Júnior, E. Riva Sanseverino, P. Gallo, D. Koch, S. Diel, G. Walter, L. Trilla, V. J. Ferreira, G. B. Pérez, Y. Kotak, *et al.*, "Towards to battery digital passport: reviewing regulations and standards for second-life batteries", *Batteries*, vol. 10, no. 4, p. 115, 2024, doi:10.3390/batteries10040115.

[15] J. S. Edge, S. O'Kane, R. Prosser, N. D. Kirkaldy, A. N. Patel, A. Hales, A. Ghosh, W. Ai, J. Chen, J. Yang, *et al.*, "Lithium ion battery degradation: what you need to know", *Physical Chemistry Chemical Physics*, vol. 23, no. 14, pp. 8200–8221, 2021, doi:10.1039/D1CP00359C.

[16] P. Dini, A. Colicelli, S. Saponara, "Review on modeling and soc/soh estimation of batteries for automotive applications", *Batteries*, vol. 10, no. 1, p. 34, 2024, doi:10.3390/batteries10010034.

[17] W. Waag, C. Fleischer, D. U. Sauer, "Critical review of the methods for monitoring of lithium-ion batteries in electric and hybrid vehicles", *Journal of Power Sources*, vol. 258, pp. 321–339, 2014, doi:10.1016/j.jpowsour.2014.02.064.

[18] A. Haraz, K. Abualsaud, A. Massoud, "State-of-health and state-of-charge estimation in electric vehicles batteries: a survey on machine learning approaches", *IEEE Access*, 2024, doi:10.1109/ACCESS.2024.3486989.

[19] V. Chandran, C. K. Patil, A. Karthick, D. Ganeshaperumal, R. Rahim, A. Ghosh, "State of charge estimation of lithium-ion battery for electric

vehicles using machine learning algorithms”, *World Electric Vehicle Journal*, vol. 12, no. 1, p. 38, 2021, doi:10.3390/wevj12010038.
[20] J. P. Rivera-Barrera, N. Muñoz-Galeano, H. O. Sarmiento-Maldonado, “SoC estimation for lithium-ion batteries: Review and future challenges”, *Electronics*, vol. 6, no. 4, p. 102, 2017, doi:10.3390/electronics6040102.
[21] J. Urquizo, P. Singh, “A review of health estimation methods for Lithium-ion batteries in Electric Vehicles and their relevance for Battery Energy Storage Systems”, *Journal of Energy Storage*, vol. 73, p. 109194, 2023, doi:10.1016/j.est.2023.109194.
[22] C. Vidal, P. Malysz, P. Kollmeyer, A. Emadi, “Machine learning applied to electrified vehicle battery state of charge and state of health estimation: State-of-the-art”, *Ieee Access*, vol. 8, pp. 52796–52814, 2020, doi:10.1109/ACCESS.2020.2980961.
[23] Z. Deng, X. Hu, X. Lin, Y. Che, L. Xu, W. Guo, “Data-driven state of charge estimation for lithium-ion battery packs based on Gaussian process regression”, *Energy*, vol. 205, p. 118000, 2020, doi:10.1016/j.energy.2020.118000.
[24] E. Chemali, P. J. Kollmeyer, M. Preindl, R. Ahmed, A. Emadi, “Long short-term memory networks for accurate state-of-charge estimation of Li-ion batteries”, *IEEE Transactions on Industrial Electronics*, vol. 65, no. 8, pp. 6730–6739, 2017, doi:10.1109/TIE.2017.2787586.
[25] P. Khumprom, N. Yodo, “A data-driven predictive prognostic model for lithium-ion batteries based on a deep learning algorithm”, *Energies*, vol. 12, no. 4, p. 660, 2019, doi:10.3390/en12040660.
[26] P. Reshma, V. J. Manohar, “Collaborative evaluation of SoC, SoP and SoH of lithium-ion battery in an electric bus through improved remora optimization algorithm and dual adaptive Kalman filtering algorithm”, *Journal of Energy Storage*, vol. 68, p. 107573, 2023, doi:10.1016/j.est.2023.107573.
[27] A. Di Martino, S. Volturno, M. Longo, W. Yaici, D. Zaninelli, “Machine Learning-Based SoC and RUL Estimation Applied to an Electric Bus Fleet”, *in 2025 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, pp. 329–334, 2025, doi:10.1109/ICVES65691.2025.11376570.
[28] D. N. How, M. A. Hannan, M. S. H. Lipu, K. S. Sahari, P. J. Ker, K. M. Muttaqi, “State-of-charge estimation of li-ion battery in electric vehicles: A deep neural network approach”, *IEEE Transactions on Industry Applications*, vol. 56, no. 5, pp. 5565–5574, 2020, doi:10.1109/TIA.2020.3004294.
[29] P. S. Babu, I. V, A. B, V. S, K. C, “Enhanced SOC estimation of lithium ion batteries with RealTime data using machine learning algorithms”, *Scientific Reports*, vol. 14, no. 1, p. 16036, 2024, doi:10.1038/s41598-024-66997-9.
[30] J. P. Orellana-Iñiguez, W. L. Gallo, M. C. De Almeida, “Real-World Evaluation of Machine Learning Techniques for State of Charge Estimation in Electric Bus Batteries”, *in 2025 Brazilian Power Electronics Conference (COBEP)*, pp. 1–6, 2025, doi:10.1109/COBEP66423.2025.11231456.
[31] L. F. Ugarte, D. N. Sarmiento, F. T. Mariotto, E. Lacusta, M. C. de Almeida, “Living lab for electric mobility in the public transportation system of the university of campinas”, *in 2019 IEEE PES Innovative Smart Grid Technologies Conference-Latin America (ISGT Latin America)*, pp. 1–6, 2019, doi:10.1109/ISGT-LA.2019.8895442.

BIOGRAPHIES

Johanna Patricia Orellana-Iñiguez was born in Cuenca, Ecuador (1993). She received the B.Sc. degree in Chemical Engineering from the Universidad de Cuenca and a professional Master’s in Industrial Process Improvement from ESPOL (2023). In 2025, she obtained a Master’s degree in Energy Systems Planning from UNICAMP. She is currently pursuing the Ph.D. degree in Electrical Engineering at the School of Electrical and Computer Engineering (FEEC), UNICAMP. Her research focuses on lithium-ion battery degradation, State-of-Charge estimation using machine learning, and data-driven diagnostics for electric mobility, with interests in battery energy storage systems, electrochemical data analysis, power electronics, and sustainable mobility.

Waldyr Luiz Ribeiro Gallo is a Full Professor in the Energy Department of the School of Mechanical Engineering at the University of Campinas. He holds both M.Sc. and Ph.D. degrees in Mechanical Engineering. His research focuses on energy systems planning, vehicle propulsion, biofuels, renewable energy integration, and advanced energy storage technologies. His work also includes thermodynamics, with emphasis on internal combustion engines, gas turbines, cogeneration, and energy efficiency in industrial and transportation systems. He collaborates on multidisciplinary projects aimed at developing cleaner and more sustainable energy solutions.

Madson Cortes de Almeida holds a Bachelor’s degree in Electrical Engineering from the Federal University of Minas Gerais (UFMG), a Ph.D. in Electrical Engineering from the University of Campinas (UNICAMP), and completed a postdoctoral fellowship at the University of Manchester, U.K. His research focuses on power and energy systems, including distribution system analysis, state estimation, distributed generation, renewable energy sources, fault location, and electric mobility. He served as Vice-Coordinator of the Technical Committee on Power Systems of the Brazilian Society of Automation (2015–2016). He is currently an Associate Professor in the Department of Systems and Energy (DSE) at the School of Electrical and Computer Engineering, University of Campinas (FEEC/UNICAMP), where he coordinates the Living Laboratory for Electric Mobility in Public Transportation.