POLITECNICO DI TORINO
Repository ISTITUZIONALE

Experimental Evaluation of a Cascaded SoC–SoH Deep Learning Framework for Lithium-Ion Battery Management Systems

Original
Experimental Evaluation of a Cascaded SoC–SoH Deep Learning Framework for Lithium-Ion Battery Management Systems / Randazzo, V., Pasero, E., Bonaccorsi, M., Martínez-Peiró, M.. - In: ELECTRONICS. - ISSN 2079-9292. - ELETTRONICO. - 15:17(2026). [10.3390/electronics15173823]

Availability:
This version is available at: 11583/3015022 since: 2026-08-31T14:27:48Z

Publisher:
MDPI

Published
DOI:10.3390/electronics15173823

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

12 October 2026

electronics

MDPI

Article

Experimental Evaluation of a Cascaded SoC–SoH Deep Learning Framework for Lithium-Ion Battery Management Systems

Vincenzo Randazzo [1], Eros Pasero [1], Maximilian Bonaccorsi [2] and Marcos Martínez-Peiró [3,*]

[1] Department of Electronics and Telecommunications, Politecnico di Torino, 10129 Turin, Italy; vincenzo.randazzo@polito.it (V.R.); eros.pasero@polito.it (E.P.)
[2] Telecommunication School, Universitat Politècnica de València, 46022 Valencia, Spain; maborod@upv.edu.es
[3] The Institute for Instrumentation in Molecular Imaging (I3M), CSIC-Universitat Politècnica de València, 46022 Valencia, Spain
* Correspondence: mpeiro@eln.upv.es

Abstract

Accurate estimation of battery state indicators, particularly State of Charge (SoC) and State of Health (SoH), remains a fundamental requirement for advanced BMS, directly affecting safety, reliability, operational efficiency, and battery lifetime. Recent advances in deep learning have demonstrated promising results for SoC estimation and SoH classification individually; however, the impact of SoC estimation errors on subsequent SoH classification has received limited attention. This work investigates whether accurately estimated SoC information can effectively support data-driven SoH classification and quantifies the propagation of uncertainty across a cascaded battery diagnostic framework. A two-stage deep learning architecture combining LSTM-based SoC estimation and ResNet-based SoH classification is proposed and evaluated using three lithium-ion battery datasets. The SoC estimation stage is assessed using two experimental datasets (LG Chem E66 and Panasonic NCR18650PF), while the SoH classification and uncertainty-propagation analysis are conducted using the GS Yuasa LEV50 dataset generated from a validated battery simulation model. To improve temporal feature extraction and noise robustness, signal preprocessing based on Butterworth and Exponential Moving Average filters was incorporated. Experimental results achieved SoC estimation errors as low as 1.46% and an SoH classification accuracy of 94.23% when reference SoC values were used. Under realistic operating conditions, where SoH classification relied on estimated SoC values, the accuracy decreased to 79.65%. These findings provide a quantitative assessment of uncertainty propagation in integrated SoC-SoH pipelines and demonstrate that highly accurate SoC estimation alone does not necessarily guarantee equivalent SoH classification performance. The study highlights the challenges and limitations of cascaded data-driven battery diagnostics and provides insight for the development of more reliable intelligent BMS solutions.

Keywords: lithium-ion batteries; battery management systems; state of charge estimation; state of health classification; LSTM; ResNet; cascaded SoC–SoH estimation; uncertainty propagation; battery diagnostics; electric vehicles

Check for updates

Academic Editor: Yi-Hua Liu

Received: 31 July 2026
Revised: 20 August 2026
Accepted: 23 August 2026
Published: 26 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The widespread adoption of electric vehicles has driven significant research and development efforts in the field of BMS. A critical function of modern BMSs is the accurate estimation of battery state parameters, particularly State of Charge (SoC) and State of Health (SoH), which directly influence battery safety, reliability, operational efficiency, and

lifetime. Precise knowledge of these metrics is essential for optimizing battery performance, preventing unsafe operating conditions, and extending the useful life of lithium-ion battery packs used in electric vehicles (EVs) [1–5].

Traditional approaches to SoC estimation and SoH classification have mainly relied on electrochemical, equivalent-circuit, and model-based techniques. Although these methods can provide accurate results under controlled conditions, the inherent complexity, ageing mechanisms, parameter variations, and nonlinear behaviour of lithium-ion batteries often limit their effectiveness under the highly dynamic operating conditions encountered in EV applications [6–9]. Recent review studies have highlighted SoC and SoH estimation as two of the most important functions of modern BMSs and have emphasized the growing need for more robust and adaptive estimation methods capable of operating under real-world conditions [10].

Recent advances in artificial intelligence have promoted the adoption of data-driven approaches for battery state estimation, allowing complex nonlinear battery behaviour to be learned directly from operational data without relying exclusively on detailed electrochemical models [11]. Artificial Neural Networks (ANNs), Recurrent Neural Networks (RNNs), Gated Recurrent Units (GRUs), and Long Short-Term Memory (LSTM) networks have emerged as some of the most widely adopted machine-learning techniques for battery state estimation due to their ability to capture nonlinear and temporal dependencies in battery behaviour [11].

In recent years, the application of machine learning, and specifically neural network (NN) techniques, has shown significant promise in addressing challenges related to battery state estimation [12], battery management, energy management, and optimization. These capabilities have motivated their increasing adoption for battery state estimation and health-monitoring applications, where accurate modelling of nonlinear battery behaviour and degradation processes is required [13,14].

Despite the encouraging results reported in the literature, several open challenges remain. First, many existing models are developed and validated using a single battery dataset, limiting their ability to generalize across different battery chemistries, as well as form factors, operating temperatures, and usage conditions. Second, the influence of signal preprocessing strategies on the performance of deep learning-based battery estimators has received comparatively limited attention despite their potential impact on noise reduction and temporal feature extraction. Third, and most importantly, most existing studies evaluate SoC estimation and SoH classification tasks independently, whereas the propagation of uncertainty between both tasks remains largely unexplored.

Recent studies have highlighted the strong coupling between SoC and SoH and suggested that jointly considering both states may improve battery-state estimation performance. Wu et al. [13] proposed a BiLSTM-SA framework for joint SoC-SoH estimation and demonstrated that incorporating health-related information improved charge-state estimation accuracy. Similarly, Wei et al. [14] showed that battery degradation significantly affects SoC estimation performance and advocated integrated SoC-SoH estimation strategies combining data-driven and model-based approaches. Motivated by this reported coupling, the present work investigates the reverse perspective: whether an accurately estimated SoC can provide meaningful information for improving downstream SoH classification performance and how estimation errors propagate through a cascaded SoC-SoH framework.

To address these challenges, this work presents an experimental data-driven framework combining LSTM-based SoC estimation with ResNet-based SoH classification. Rather than proposing a new neural architecture, the objective is to systematically evaluate how signal preprocessing strategies, model selection, and dataset diversity affect the performance and robustness of battery state estimation across different lithium-ion technologies

and operating conditions. Particular emphasis is placed on analyzing uncertainty propagation from SoC estimation to subsequent SoH classification, a scenario that more closely reflects practical BMS deployment conditions.

The main contributions of this work are fourfold:

1. A comparative evaluation of LSTM-based SoC estimation models across heterogeneous lithium-ion battery datasets, covering different cell chemistries, form factors, operating temperatures, and driving conditions.
2. An assessment of the impact of signal preprocessing techniques, including Butterworth filtering, EMA smoothing, and Z-score normalization, on the accuracy and generalization capability of deep learning-based battery state estimators.
3. The development and validation of a ResNet-based framework for SoH classification, capable of extracting discriminative degradation-related features from multivariate battery measurements.
4. A realistic end-to-end SoC–SoH estimation pipeline, in which SoH classification relies on previously estimated rather than ground-truth SoC values, enabling quantitative analysis of uncertainty propagation and practical deployment limitations in real-world BMS implementations.

The remainder of this paper is organized as follows. Section 2 presents the background on lithium-ion batteries, BMS operation, and state estimation methods. Section 3 describes the results, namely the neural network architectures, the preprocessing techniques, the experimental methodology considered in this study, and lists the limitations of this work. Section 4 details the discussion of the main findings obtained for SoC estimation, SoH classification, and the integrated SoC–SoH framework. Finally, Section 5 summarizes the main conclusions and outlines future research directions.

2. Materials and Methods

2.1. Background on Battery State Estimation

State of Charge represents the ratio between the remaining available capacity and the maximum capacity of a battery and is commonly expressed as

$$\mathrm{SoC(t)} = \frac{\mathrm{Q(t)}}{\mathrm{Q_m}} \cdot 100\% \tag{1}$$

where $Q(t)$ is the remaining capacity and Q_m is the maximum available capacity. Equation (2) corresponds to the conventional Coulomb-counting approach, one of the most widely adopted methods for SoC estimation in BMS [15,16].

$$\mathrm{SoC(t)} = \mathrm{SoC(t_0)} + \frac{\eta}{C_n}\int_{\mathrm{t_0}}^{\mathrm{t}} I(t)\mathrm{dt} \tag{2}$$

where C_n denotes rated capacity, $I(t)$ is the instantaneous discharge current of the battery and η is the coulombic efficiency. Accurate SoC estimation is essential for energy management, range prediction, and safe battery operation.

State of Health describes the degradation level of a battery relative to its beginning-of-life condition and is typically associated with capacity fade or resistance increase [17]. The SoH metric based on capacity retention is commonly expressed as Equation (3), commonly used in battery-health estimation studies [18].

$$\mathrm{SoH} = \frac{\mathrm{Q_{max}}}{\mathrm{Q_{BOL}}} \cdot 100\% \tag{3}$$

where Q_{BOL} and Q_{max} denote the beginning-of-life and current maximum capacities, respectively. Unlike SoC, which varies on a short timescale, SoH evolves slowly and reflects long-term battery degradation. Consequently, SoH classification usually requires extracting degradation-related information from extended operational histories.

Battery-state estimation methods are generally divided into model-based and data-driven approaches. While model-based techniques rely on equivalent-circuit or electrochemical models, data-driven methods employ machine learning algorithms to learn the nonlinear relationships between measurable variables such as voltage, current, temperature, SoC, and battery ageing indicators. Recent studies have demonstrated the potential of deep learning approaches for both SoC estimation and SoH assessment, motivating the framework investigated in this work.

Battery-State Estimation Approaches

Estimating the SoC and SoH of lithium-ion batteries remains a challenging task due to the highly nonlinear and time-varying electrochemical processes governing battery behaviour. Existing estimation approaches are commonly categorized into model-based and data-driven methods, as summarized in Figures 1 and 2.

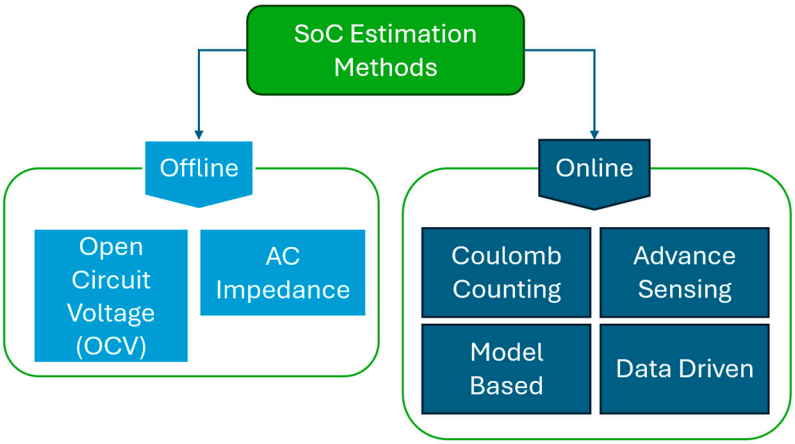


Figure 1. Main SoC estimation methods.

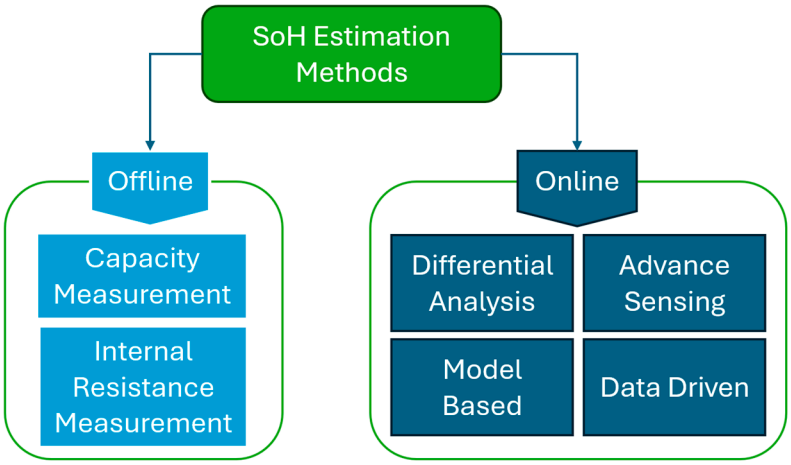


Figure 2. Main SoH estimation methods.

Model-based approaches rely on electrochemical or equivalent-circuit representations of the battery and often require accurate parameter identification. While these methods can provide physically interpretable results, their performance may degrade under varying

operating conditions, ageing mechanisms, and environmental factors. In contrast, data-driven approaches learn the relationship between measurable variables, such as voltage, current, and temperature, and battery state indicators directly from experimental data, avoiding the need for explicit physical modelling [14,15].

Recent advances in machine learning have significantly increased the adoption of data-driven methodologies in battery-state estimation. Techniques ranging from support vector machines and fuzzy inference systems to deep learning architectures have been successfully applied to SoC prediction and SoH classification. Among these approaches, artificial neural networks have demonstrated effectiveness in capturing nonlinear temporal dependencies and degradation patterns from large-scale battery datasets. Consequently, this work focuses on deep learning-based models, specifically Long Short-Term Memory (LSTM) networks for SoC estimation and Residual Neural Networks (ResNets) for SoH classification.

2.2. Deep Learning Architectures

Deep learning models have demonstrated remarkable capabilities for battery state estimation due to their ability to learn nonlinear relationships directly from experimental data. Among the available architectures, recurrent neural networks are particularly suitable for SoC estimation because they can exploit temporal dependencies in sequential measurements, whereas convolutional architectures are effective for extracting degradation-related features associated with SoH.

2.2.1. Long Short-Term Memory Networks

Long Short-Term Memory (LSTM) networks are an advanced form of Recurrent Neural Network (RNN) specifically designed to capture long-term temporal dependencies while mitigating the vanishing-gradient problem encountered in conventional RNNs [19,20]. Through gated memory cells, LSTMs selectively retain and update temporal information, making them well suited to battery applications where the current state depends on previous operating conditions.

Because SoC evolution is inherently sequential and depends on previous current, voltage, and temperature measurements, recurrent neural network architectures have been widely adopted for battery SoC estimation. Among them, Long Short-Term Memory (LSTM) networks are one of the most frequently reported approaches due to their ability to capture long-term temporal dependencies. Although Gated Recurrent Unit (GRU) networks have demonstrated competitive and, in some cases, superior performance under specific conditions [21], particularly when data availability is limited, the literature does not provide a consistent consensus regarding the superiority of either architecture for time-series forecasting tasks [20]. Consequently, considering their extensive use and proven effectiveness in battery-state estimation applications, LSTM networks were selected as the core architecture for SoC prediction in this work. A comparative assessment between LSTM and GRU architectures is considered an interesting direction for future research.

2.2.2. Residual Neural Networks

Residual Neural Networks (ResNets) extend conventional convolutional neural networks by introducing residual connections that facilitate gradient propagation during training [12,22]. Instead of directly learning the target mapping, each residual block learns a residual function $F(x)$, allowing the network output to be expressed as

$$H(x) = F(x) + x \tag{4}$$

where x denotes the input features.

This mechanism improves training stability and enables the extraction of hierarchical features from complex datasets. In battery-state estimation, these characteristics are particularly useful for identifying degradation patterns associated with ageing processes. For this reason, a ResNet-based architecture was adopted in this work for SoH classification.

2.3. Battery Datasets

To evaluate the proposed framework under heterogeneous operating conditions, three lithium-ion battery datasets covering different cell chemistries, form factors, capacities, and test conditions were considered. The LG Chem E66 and Panasonic NCR18650PF datasets were used for SoC estimation, formulated as a regression problem, whereas the GS Yuasa LEV50 dataset was employed for SoH estimation, formulated as a multi-class classification task.

2.3.1. LG Chem E66 Dataset

The experimental dataset consists of measurements collected from twelve fresh LG Chem E66 lithium-ion pouch cells (C1–C12), each with a nominal capacity of 66 Ah. Current, voltage, temperature, and ampere-hour data were acquired during controlled laboratory testing and driving-cycle experiments. Most tests were sampled at 10 Hz, while the Galvanostatic Intermittent Titration Technique (GITT) experiments were recorded at 1 Hz. The dataset includes capacity, HPPC, RCID, and GITT tests, together with Fiorano and Nürburgring driving cycles, providing a broad range of operating conditions for model development and evaluation. These experiments were specifically designed to capture both the slow electrochemical dynamics and the transient behaviour of the cells under realistic operating conditions. The battery dataset comprises several laboratory tests and real driving cycles, summarized in Table 1.

Table 1. Summary of the experimental tests and driving cycles in the LG Chem E66 battery dataset.

Test Type	Objective	Characteristics
Capacity	Characterization of capacity and slow dynamics	Constant-current charge/discharge cycles at different C-rates
HPPC	Dynamic power capability	Charge/discharge current pulses with rest periods
RCID	Internal dynamics and rate capability	Multiple charging/discharging sequences at different C-rates
GITT	Voltage relaxation and diffusion processes	Short pulses followed by relaxation intervals
Fiorano	Real driving profile	High-power dynamic current profile
Nürburgring	Real driving profile	Aggressive racing-oriented current profile

The complete dataset contains measurements from twelve independent cells. Following the protocol adopted in this study, the dataset was partitioned at the cell level. All data from cell C1 were used for training, while data from cells C2–C12 were reserved for model evaluation. This strategy was adopted to assess the ability of the proposed approach to generalize to previously unseen cells of the same battery type. Cell C1 was selected as the training source because it contains the largest number of available operational profiles in the dataset, providing sufficient data for model development. Additionally, two tests from cell C2 (the capacity test at 10 °C and the RCID test at 40 °C) were used exclusively for validation and hyperparameter tuning, while the remaining data were reserved for final testing. Input variables consisted of voltage, current, and temperature, whereas SoC values

were used as ground-truth targets. This dataset was primarily used to evaluate model generalization across different cells and operating temperatures.

2.3.2. Panasonic NCR18650PF Dataset

The second dataset used in this work corresponds to the Panasonic NCR18650PF lithium-ion cell dataset originally published by Kollmeyer et al. [23]. The dataset was acquired from a cylindrical 2.9 Ah lithium-ion cell subjected to a series of standardized driving cycles under controlled laboratory conditions. Measurements were collected at ambient temperatures ranging from -20 °C to 25 °C, enabling the assessment of battery behaviour under a wide range of operating environments.

Testing was performed using high-precision instrumentation inside a temperature-controlled chamber with a stability of ± 0.5 °C. Current, voltage, and temperature signals were acquired at 10 Hz and subsequently downsampled to 1 Hz for data processing. The driving profiles were derived from representative electric vehicle operating conditions and included standard cycles such as US06, HWFET, UDDS, and LA92, together with a custom Neural Network (NN) cycle specifically designed to introduce additional dynamic operating conditions.

The input features consist of voltage, current, temperature, and filtered versions of the voltage and current signals obtained using Butterworth filters. Ground-truth SoC values were generated through coulomb counting and used as target labels. All variables were normalized to the [0, 1] range prior to model training.

Following the partitioning strategy proposed in [23], training data consisted of Cycles 1–4 and the US06 driving cycle, while validation data corresponded to the LA92 cycle at -10 °C. Testing was performed using the completely independent LA92, NN, and UDDS drive cycles, which were not used during model development. This cell-level partitioning strategy prevents data leakage and enables the evaluation of model generalization under previously unseen operating conditions and temperature scenarios.

The Panasonic dataset was employed as an independent benchmark to assess the robustness and architecture reusability of the proposed SoC estimation methodology beyond the LG Chem E66 dataset.

2.3.3. GS Yuasa LEV50 Dataset

The third dataset employed in this work was used for the development and evaluation of the SoH estimation models. Unlike the previous datasets, which focus on SoC estimation, this dataset was specifically designed for battery health assessment and formulates the problem as a multi-class classification task.

The LEV50 dataset is not derived from direct experimental ageing measurements but from a validated Simulink battery model designed to reproduce the dynamic and degradation behaviour, under realistic operating conditions, of the GS Yuasa LEV50 lithium-ion cells (50 Ah, nominal voltage 3.75 V) manufactured by GS Yuasa Corporation, Kyoto, Japan. The simulation repeatedly combines WLTP driving cycles, 2C fast-charging events, and periodic capacity measurements. Battery ageing is represented through Equivalent Full Cycles (EFCs), while measurement uncertainty is emulated by adding controlled noise to voltage and temperature signals. This approach provides a controlled environment for studying uncertainty propagation effects while maintaining consistency across ageing conditions and operating temperatures.

The available variables include voltage, current, temperature, SoC, and SoH. To emulate realistic measurement conditions, uniformly distributed noise was added to the voltage and temperature signals. The dataset includes simulations corresponding to EFC

values of 0, 100, 200, 300, 400, 500, 600, 700, 800, 900, and 1000, covering the battery lifetime from a fresh cell to a significantly aged condition.

The dataset partition was performed according to temperature conditions rather than by random sample allocation. The training subset included temperatures of −20 °C, −10 °C, 0 °C, 10 °C, 20 °C, 25 °C, 35 °C, 40 °C, 50 °C and 60 °C. Validation was conducted exclusively at 55 °C, a temperature deliberately excluded from the training subset to monitor generalization under unseen thermal conditions during hyperparameter tuning. The testing subset comprised additional unseen temperatures (−15 °C, −5 °C, 5 °C, 15 °C, 30 °C and 45 °C), providing an independent evaluation of model performance across operating conditions not used during training.

Each simulation test contains 2421 samples acquired at a sampling frequency of 1 Hz, corresponding to approximately 40 min of battery operation. To make the data suitable for deep learning processing, the input sequences were reorganized into fixed-length windows of 269 samples. Since each sample represents one second of operation, each input window corresponds to approximately 269 s (4.5 min) of battery behaviour.

The selected window length represents a compromise between capturing sufficient temporal information to characterize battery dynamics and maintaining a practical observation period for potential BMS applications. Shorter windows may fail to capture relevant temporal dependencies, whereas significantly longer windows would increase the time required before a reliable prediction can be generated.

As the LEV50 dataset was provided in a pre-processed form, detailed information regarding the original sliding-window implementation, including stride and overlap parameters, was not available. Nevertheless, all experiments were conducted using the same dataset partitions and preprocessing procedure to ensure a consistent and fair comparison between the reference and cascaded SoC–SoH evaluation scenarios.

The SoH values were grouped into four classes: Class 1 ($95\% \leq \text{SoH} \leq 100\%$), Class 2 ($90\% \leq \text{SoH} < 95\%$), Class 3 ($85\% \leq \text{SoH} < 90\%$), and Class 4 ($80\% \leq \text{SoH} < 85\%$). Following the original dataset partitioning strategy, the training set contains data from temperatures between −20 °C and 60 °C, the test set includes intermediate unseen temperatures, and the validation set is composed of data collected at 55 °C. This partitioning strategy enables the evaluation of the proposed methodology under different ageing and temperature conditions.

This dataset was employed to investigate the impact of SoC estimation uncertainty on subsequent SoH classification performance.

2.3.4. Summary of Datasets

Table 2 summarizes the main characteristics of the datasets employed in this work, including cell type, capacity, estimation task, input variables, and data partitioning strategy.

Table 2. Characteristics of Datasets.

Dataset	Cell Type	Capacity	Task	Inputs	Output
LG Chem E66	Pouch	66 Ah	SoC estimation	V, I, T	SoC
Panasonic NCR18650PF	Cylindrical	2.9 Ah	SoC estimation	V, I, T + filtered signals	SoC
GS Yuasa LEV50	Prismatic	50 Ah	SoH Classification	V, I, T, SoC	SoH Class

2.4. SoC Estimation Framework

Accurate SoC estimation constitutes the first stage of the proposed methodology and provides the input information required for the subsequent SoH classification framework. Therefore, the objective was not only to achieve high SoC estimation tolerance-based

accuracy, but also to obtain a robust estimator capable of maintaining its performance under different operating conditions and battery datasets.

Several neural network architectures were initially evaluated, including Feedforward Neural Networks (FNNs), Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) networks. Different preprocessing techniques, activation functions, optimization algorithms, and hyperparameter configurations were also investigated. Among all evaluated alternatives, LSTM-based architectures consistently achieved the best trade-off between estimation tolerance-based accuracy, training stability, and generalization capability. Consequently, the final SoC estimation framework was developed using an LSTM network combined with signal filtering and feature engineering techniques.

The framework was first optimized using the LG Chem E66 dataset and subsequently evaluated on the independent Panasonic NCR18650PF dataset. This dual-dataset evaluation strategy enables the assessment of both estimation of tolerance-based accuracy and applicability across independent datasets of different battery technologies, operating conditions, and driving profiles.

2.4.1. Data Preprocessing

To improve the quality of the input signals and enhance the extraction of temporal battery dynamics, voltage and current measurements were processed using first-order Butterworth low-pass filters. These filters attenuate high-frequency noise while preserving the low-frequency information associated with the electrochemical behaviour of the battery. These filters were implemented using a causal first-order recursive formulation. Consequently, the filter can be deployed in real-time battery-state estimation with a limited phase delay associated with the selected cut-off frequencies.

The filter transfer function is given by

$$H(f) = \frac{1}{\sqrt{1 + \varepsilon^2 \left(\frac{f}{f_c}\right)^{2n}}} \tag{5}$$

where f is the signal frequency, f_c is the cut-off frequency, ε is the passband ripple coefficient, and n is the filter order. In this work, a first-order Butterworth filter ($n = 1$) was adopted, and two cut-off frequencies were evaluated to capture battery dynamics at different time scales (f_1 = 0.5 mHz, f_2 = 3 mHz). The effect of the cut-off frequencies can be summarized as follows:

- Lower cut-off frequencies: Leads to stronger attenuation of high-frequency components, resulting in a smoother signal that captures long-term trends.
- Higher cut-off frequencies: Let more of the high-frequency components to pass through, making the signal more responsive to rapid changes but potentially introducing more noise.

Following the filtering stage, all input variables were standardized using Z-score normalization:

$$x_{norm} = \frac{x - \bar{x}}{\sigma_x}, \tag{6}$$

where x is the original variable, $\bar{x}$ is its mean value, and σ_x is its standard deviation.

2.4.2. Input Feature Selection

Different combinations of raw and filtered measurements were evaluated to determine the most informative input representation. Table 3 summarizes the feature configurations investigated during the experimental phase.

Table 3. Different input configurations.

Input Configuration	Inputs $\varnothing_k$
Original	$\varnothing_k = [V_k, I_k, T_k]$
Conf. 1	$\varnothing_k = [V_k, I_k, T_k, V_{EMA,k}, I_{EMA,k}, V_{f1,k}, I_{f1,k}, V_{f2,k}, I_{f2,k}]$
Conf. 2	$\varnothing_k = [T_k, V_{EMA,k}, I_{EMA,k}, V_{f1,k}, I_{f1,k}, V_{f2,k}, I_{f2,k}]$
Conf. 3	$\varnothing_k = [T_k, V_{f1,k}, I_{f1,k}, V_{f2,k}, I_{f2,k}]$
Conf. 4	$\varnothing_k = [V_k, I_k, T_k, V_{EMA,k}, I_{EMA,k}]$
Conf. 5	$\varnothing_k = [V_k, I_k, T_k, V_{f1,k}, I_{f1,k}, V_{f2,k}, I_{f2,k}]$

Among the evaluated alternatives, Configuration 5 achieved the highest validation performance and was therefore selected for the final model. This configuration combines the original voltage, current, and temperature measurements with two sets of Butterworth-filtered voltage and current signals:

$$\varnothing_k = [V_k,\ I_k, T_k, V_{f1,k}, I_{f1,k}, V_{f2,k}, I_{f2,k}], \tag{7}$$

The inclusion of filtered signals allows the network to simultaneously exploit instantaneous measurements and slower battery dynamics contained in the low-frequency components.

2.4.3. LSTM Architecture and Training Procedure

The final SoC estimator consists of a single hidden LSTM layer containing 22 memory units, followed by a fully connected output layer that generates the SoC prediction.

The network was trained using the Adam optimizer with a learning rate of 0.01 and Mean Squared Error (MSE) as loss function:

$$\text{MSE} = \frac{1}{\text{N}}\sum_{\text{i}=1}^{\text{N}} (\text{y}_\text{i} - \hat{\text{y}}_\text{i})^2, \tag{8}$$

The validation MSE was used as the optimization objective during the exploratory Optuna tuning process. Input sequences of 15 timesteps were employed, enabling the model to exploit short-term temporal dependencies in battery behaviour. Training was performed for up to 50 epochs with a batch size of 64, using the Adam optimizer, with a learning rate set to 0.01.

To improve convergence and reduce overfitting, Early Stopping with a patience of 5 epochs and an adaptive learning-rate strategy were implemented. The learning rate was reduced automatically by a factor of 0.1 when the validation loss reached a predefined threshold of 6.5×10^{-5} when the model reaches a plateau in terms of improvement, allowing finer weight adjustments during the final training stages (Figure 3).

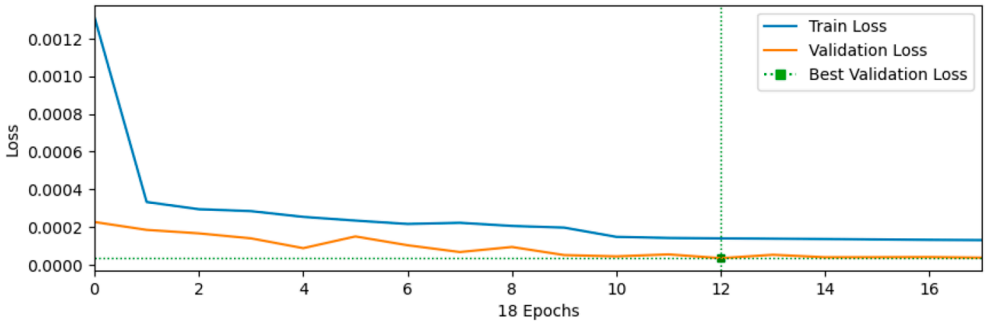


Figure 3. Training and validation loss plot for the LG dataset. Best validation loss performance of 6.5×10^{-5} was achieved at epoch 12.

Class-dependent weights were incorporated into the categorical cross-entropy loss function to mitigate the moderate class imbalance observed in the training dataset. The

adopted weights were selected empirically to improve the representation of minority classes during training and should not be interpreted as theoretically optimal coefficients derived from Bayesian risk-minimization principles.

The final architecture and hyperparameter configuration are summarized in Table 4.

Table 4. Final neural network architecture and hyperparameters for the LG Dataset.

LSTM Neural Network	
Input configuration (Conf. 5)	$\varnothing_k = [V_k, I_k, T_k, V_{f1,k}, I_{f1,k}, V_{f2,k}, I_{f2,k}]$
Hidden layers	1
Units in the hidden layer	22
Timesteps	15
Loss function	MSE
Loss function optimizer	ADAM
Learning Rate	0.01
Training epochs	50
Butterworth filter	
Cut-off frequency 1	0.5 mHz
Cut-off frequency 2	3 mHz

Optuna v3.6.0 was employed as an auxiliary tool to explore the influence of selected hyperparameters and to identify promising parameter ranges during the model development stage. However, the objective of this work was not to conduct an exhaustive hyperparameter optimization study. Due to the large search space and computational constraints, only a subset of parameters was explored automatically, while the final model configurations were selected based on the overall experimental performance and engineering judgement. Therefore, the reported results should be interpreted as representative configurations used to evaluate the proposed methodology rather than globally optimal solutions.

2.4.4. Performance Evaluation

Model performance was primarily assessed using a tolerance-based accuracy indicator, while MSE was employed during model training. Although regression metrics such as MAE, RMSE, and MSE are widely used for SoC estimation, several BMS-oriented studies also report tolerance-based accuracy metrics that quantify the proportion of estimates lying within a predefined error band. Such metrics provide a practical interpretation of estimation performance because many battery-state estimation functions require SoC estimates to remain within a specific accuracy threshold rather than minimizing the average error alone. Therefore, the tolerance-based accuracy metric adopted in this work complements conventional regression measures by evaluating compliance with a $\pm 3\%$ SoC tolerance requirement. Although MSE was used during training, the primary evaluation metric was defined as Within-Band Accuracy (WBA) as the percentage of predictions lying within a $\pm 3\%$ error band around the reference SoC value:

$$WBA = \frac{N \cdot (|SoC_{true} - SoC_{pred}| < 3)}{N_{total}} \times 100 \quad (9)$$

The $\pm 3\%$ tolerance threshold was selected because SoC estimation errors below this level are commonly regarded as acceptable for practical BMS applications and are frequently reported in the literature as a representative performance target.

Accordingly, the Within-Band Accuracy (WBA) metric is defined as the proportion of predictions falling within the ±3% error band around the reference SoC value. This metric complements conventional regression measures by providing a practical indicator of compliance with a predefined estimation-accuracy requirement rather than reflecting the magnitude of the prediction error itself.

3. Results

3.1. LG-CHEM 66 Results

The performance of the proposed framework was evaluated across different temperature ranges. The analysis first considered cells operating under positive temperatures and was subsequently extended to negative-temperature conditions to assess the model's generalization capability beyond the training domain. The final LSTM model based on Configuration 5 was compared with several alternative input configurations as well as with a more sophisticated architecture combining CNN, bidirectional LSTM, and an attention mechanism inspired by [24,25]. The average estimation accuracies obtained across all test conditions are summarized in Table 5.

Table 5. Evaluation of NN architectures. Average test tolerance-based accuracy (%) across different Temp. ranges.

Average Test Accuracy (%)						
LSTM model, original inputs	LSTM model, Config. 1	LSTM model, Config. 2	LSTM model, Config. 3	LSTM model, Config. 4	LSTM model, Config. 5	CNN1D + BiLSTM + Attention Mechanism
Positive Temperatures						
69.65	79.82	78.08	76.33	73.40	84.48	52.14
Negative Temperatures						
20.77	28.35	55.35	24.57	26.47	40.56	15.99
Total						
56.08	65.52	71.76	61.96	60.37	72.28	42.10

Table 5 shows that Configuration 5 achieved the highest overall performance, reaching an average tolerance-based accuracy of 72.28% across all test conditions. For positive-temperature tests, the model attained 84.48% tolerance accuracy, outperforming both the original-input LSTM model and the CNN1D + BiLSTM + Attention architecture. The experimental results suggest that, for the architectures and datasets considered in this work, signal preprocessing may have a greater influence on performance than further increases in network complexity.

The results demonstrate that the proposed methodology retains a measurable predictive capability when evaluated under previously unseen operating conditions. However, a substantial performance degradation is observed at negative temperatures, where the within-band accuracy decreases from 84.48% to 40.56%. This behaviour suggests that the model remains sensitive to conditions outside the temperature range represented during training and highlights the importance of temperature diversity in battery datasets.

Overall, these findings demonstrate that the proposed LSTM-based framework, combined with the selected input configuration, provides stable performance across a wide range of operating conditions. Moreover, the ability to maintain an overall tolerance-based accuracy above 72% when evaluated on both seen and unseen temperature ranges supports the robustness of the proposed approach.

Representative prediction examples obtained with Configuration 5 at different temperatures are presented in Figures 4 and 5.

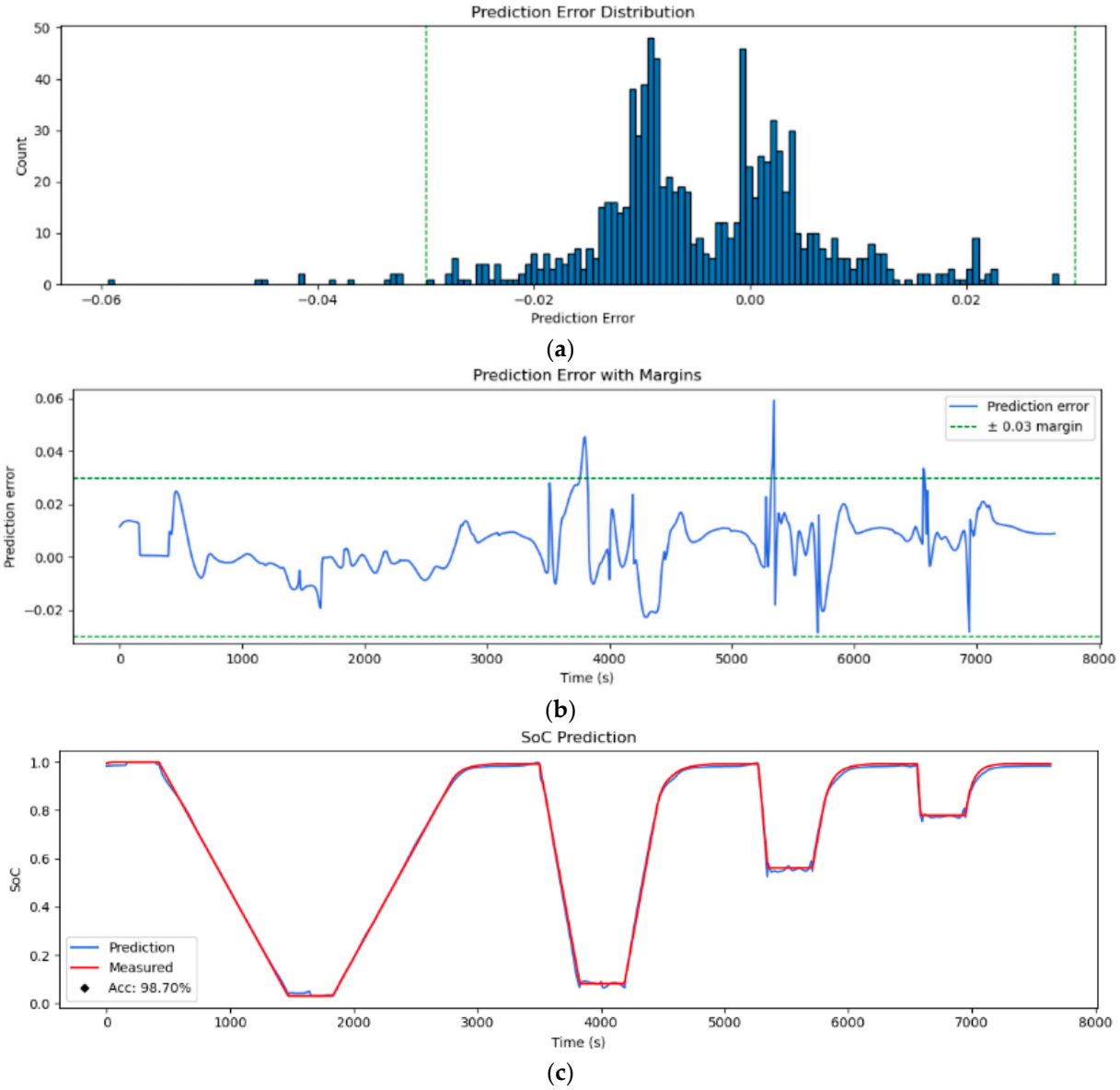


Figure 4. LG battery SoC prediction on 'Capacity' test for C2 at 10 °C: (**a**) error distribution; (**b**) prediction error; (**c**) SoC prediction. The green dashed lines in (**a**,**b**) indicate the ±0.03 error margin. The diamond marker in (**c**) denotes the 98.70% prediction accuracy achieved in this test.

3.2. PANASONIC 18650 Dataset

Building upon the insights obtained from the LG-CHEM dataset, the proposed methodology was further evaluated using the Panasonic 18650 battery dataset. This second dataset was employed as an independent validation benchmark to assess the architecture reusability of the proposed framework and its applicability beyond the original training domain.

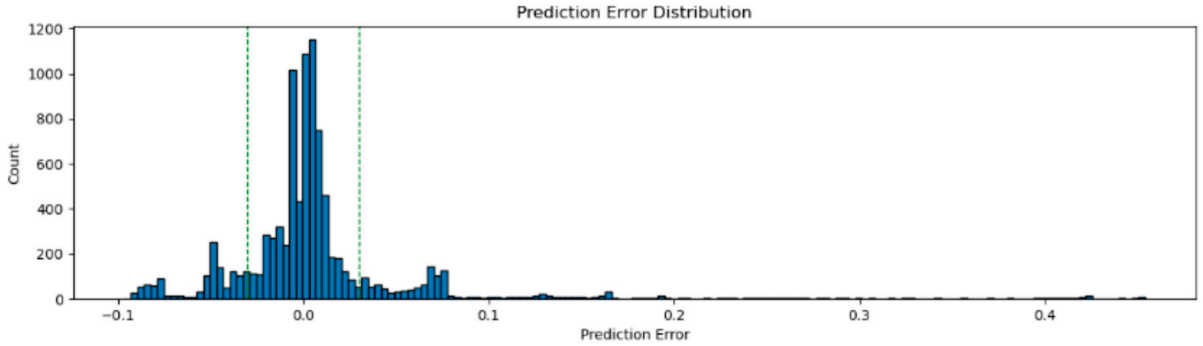
The neural network architecture selected from the previous study was retained with minimal modifications. Specifically, the final model consisted of a single LSTM hidden layer with 22 units followed by a fully connected output layer. In contrast to the LG-CHEM dataset, the Panasonic dataset was provided with pre-processed signals, including Butterworth low-pass filtering at two cutoff frequencies. Figure 6 illustrates the differences between the raw and filtered voltage and current measurements.

Although the original data were normalized within the range [0, 1], Z-score standardization was applied before training, following the findings obtained in the LG-CHEM analysis regarding improved convergence behaviour and training stability. The selected

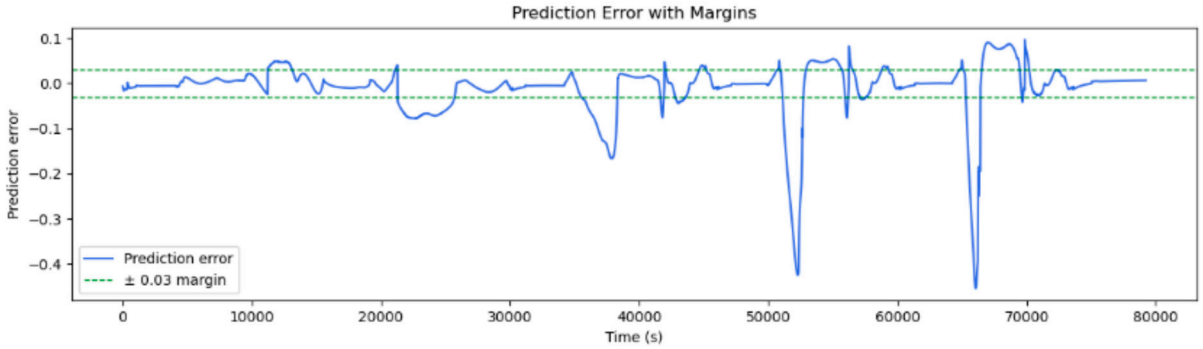
input configuration (Configuration 3), Equation (10), follows the methodology proposed in the original Panasonic study and incorporates temperature measurements together with the two filtered versions of voltage and current.

$$\varnothing_k = [T_k, V_{f1,k}, I_{f1,k}, V_{f2,k}, I_{f2,k}], \tag{10}$$

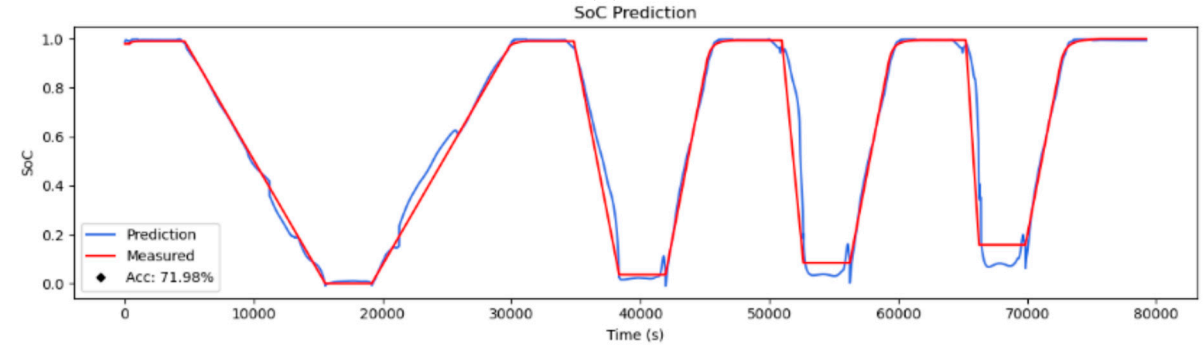
The training procedure remained consistent with that employed for the LG-CHEM dataset. Early stopping with a patience of five epochs was implemented to mitigate overfitting, and an adaptive learning-rate reduction strategy was applied whenever the validation loss reached a plateau. This ensured methodological consistency and facilitated a meaningful comparison between both datasets.



(**a**)

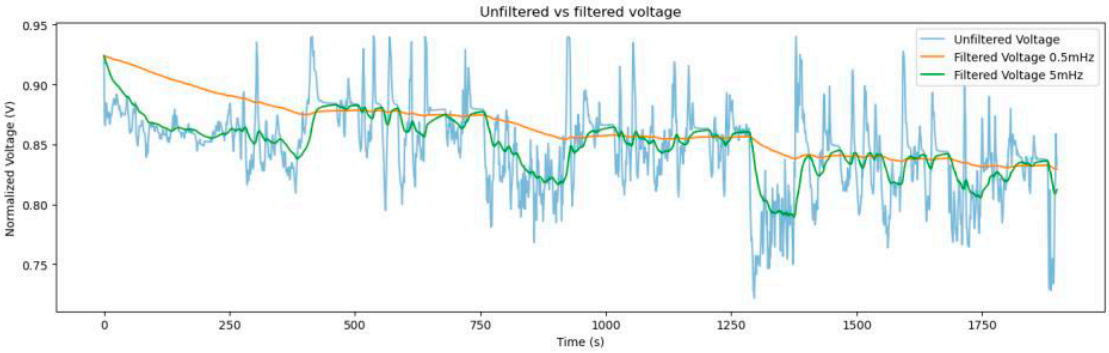


(**b**)

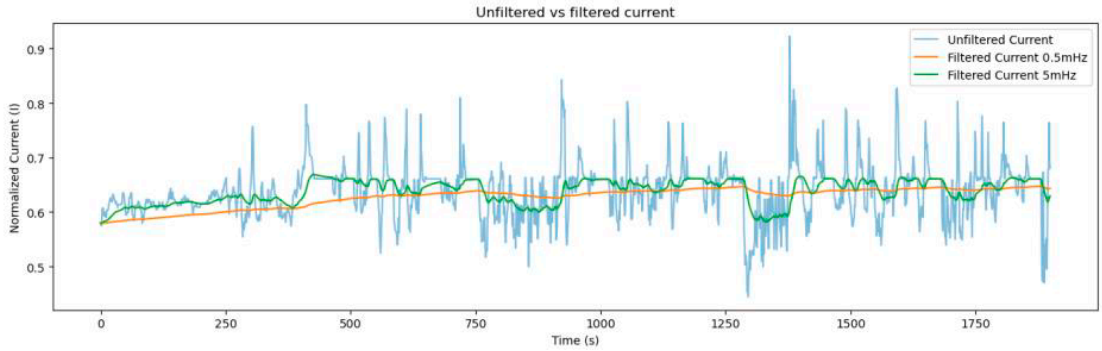


(**c**)

Figure 5. LG battery SoC prediction on 'Capacity' test for C3 at 60 °C: (**a**) error distribution; (**b**) prediction error; (**c**) SoC prediction. The green dashed lines in (**a**,**b**) indicate the ±0.03 error margin. The diamond marker in (**c**) denotes the 71.98% prediction accuracy achieved in this test.



(**a**)



(**b**)

Figure 6. Comparison between filtered and unfiltered voltage and current signals using Butterworth filters with cutoff frequencies of 0.5 mHz and 5 mHz: (**a**) normalized voltage; (**b**) normalized current.

3.3. PANASONIC 18650 Results

The proposed model achieved an overall estimation tolerance-model accuracy of 98.54% on the Panasonic test dataset, as shown in Figure 7. This result confirms the suitability of the selected LSTM architecture for SoC estimation and demonstrates that the methodology can be successfully reusable to an independent dataset.

Unlike the LG-CHEM dataset, where separate operating conditions enabled a detailed performance analysis across different temperature ranges and test scenarios, the Panasonic dataset was provided as a single consolidated test set. Consequently, no breakdown of performance according to specific operating conditions could be performed. Nevertheless, the reported tolerance-based accuracy provides a global assessment of the model's predictive capability.

The prediction errors were concentrated around zero, with most estimates remaining within a $\pm 3\%$ interval of the reference values. This narrow error distribution indicates stable model behaviour and a high degree of consistency in the estimation results.

It should be noted, however, that direct comparison with the LG-CHEM results should be conducted with caution. The Panasonic dataset benefited from extensive preprocessing, including pre-filtered input signals, and exhibits a more curated structure than the LG-CHEM dataset. These characteristics likely contributed to the higher tolerance-based accuracy achieved. Therefore, the outcomes should not be interpreted solely as evidence of superior model performance, but also as an indication of the influence that dataset characteristics and signal quality can have on the final SoC estimation tolerance accuracy.

Overall, the successful deployment of the same LSTM architecture on two independent battery datasets supports the robustness and methodological generality of the proposed methodology. Together with the LG-CHEM results, these findings indicate that data quality, preprocessing strategy, and training-data representativeness may have a greater influence on battery SoC estimation performance than further increases in network complexity.

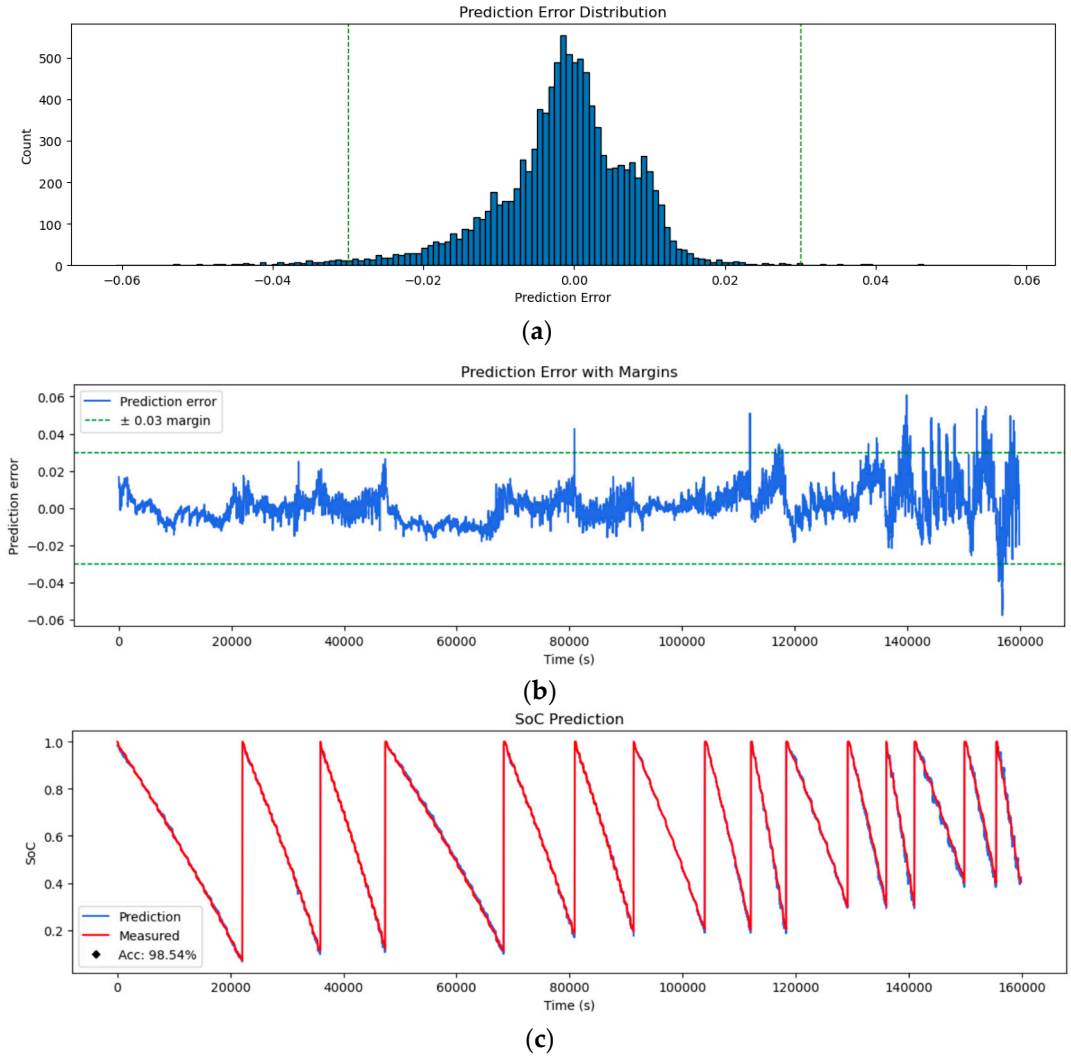


Figure 7. Evaluation results obtained on the Panasonic 18650 Li-On battery: (**a**) prediction error distribution; (**b**) prediction error; (**c**) SoC prediction. The green dashed lines in (**a**,**b**) indicate the ±0.03 error margin. The diamond marker in (**c**) denotes the 98.54% prediction accuracy achieved in this test.

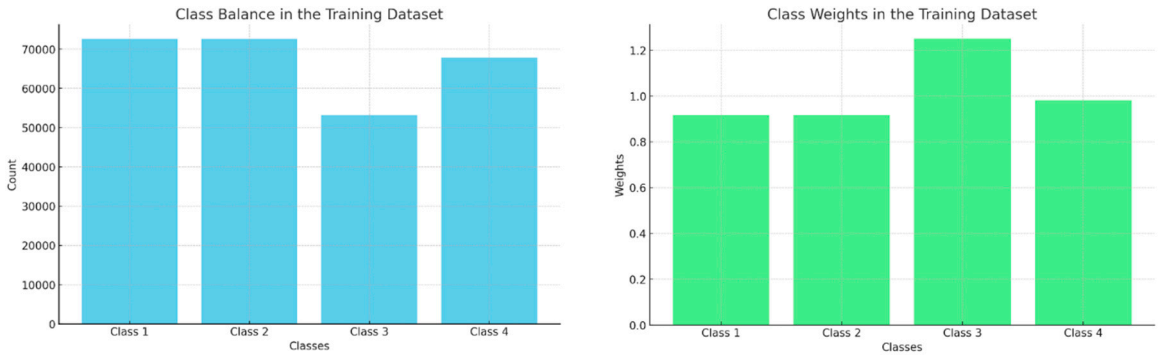
3.4. SOH Estimation and Impact of SoC Prediction Errors

3.4.1. LEV50 Dataset and Data Preprocessing

The LEV50 dataset was employed to evaluate the proposed battery SoH classification framework and to analyze the influence of SoC estimation errors on the final SoH classification performance.

The LEV50 dataset provides both SoC and SoH values generated independently by the battery simulation model. During training, the SoH classifier used the reference SoC obtained directly from the simulator. To evaluate uncertainty propagation in a cascaded architecture, the reference SoC was replaced during the testing phase by the SoC estimated by the LSTM network.

Prior to training, a preprocessing stage was applied to improve model robustness and reduce potential sources of bias. As shown in Figure 8a, the dataset exhibited a moderate class imbalance, with certain SoH classes being more represented than others. To compensate for this effect, class-dependent weights were assigned during training, increasing the penalty associated with errors in underrepresented classes (Figure 8b). This strategy encouraged the model to learn more representative features across all health categories.



(**a**) (**b**)

Figure 8. (**a**) Class distribution of the LEV50 dataset (**b**) Assignment of class weights used during training.

Similarly to the LG-CHEM and Panasonic datasets, Z-score standardization was applied to all input variables. Although the original measurements were already normalized to the range [0, 1], previous experiments showed that standardization improved convergence stability and facilitated gradient-based optimization.

3.4.2. ResNet-Based SoH Classification Framework

To perform SoH classification, a one-dimensional residual neural network inspired by the ResNet architecture was adopted. Residual connections were selected because they facilitate feature propagation through deeper networks while mitigating optimization difficulties typically associated with conventional convolutional architectures.

The proposed model consists of three main stages, illustrated in Figure 9.

- Input Block. The input stage comprises a one-dimensional convolutional layer with 16 filters, followed by layer normalization and a max-pooling layer. This block extracts low-level temporal features while reducing the dimensionality of the input sequence.
- Residual Blocks. The feature-extraction stage is composed of two residual blocks, each containing a main branch and a shortcut branch connected through residual summation. In the first residual block, the main branch applies two consecutive one-dimensional convolutional layers with 16 and 32 filters, respectively, each followed by layer normalization. The shortcut branch includes a 32-filter convolution and layer normalization to match feature dimensions prior to summation. The second residual block follows the same structure while increasing the number of filters from 32 to 64, enabling the extraction of more discriminative feature representations.
- Classification Block. The final stage consists of a global average pooling layer followed by a fully connected classification layer. A softmax activation function converts the output scores into class probabilities corresponding to the different SoH categories.

Input sequences were organized using windows of 269 samples, corresponding to approximately 4.5 min of battery operation. This duration was experimentally selected as a compromise between capturing meaningful temporal dynamics and maintaining computational efficiency.

3.4.3. Training Procedure and Evaluation Metrics

The model was trained using categorical cross-entropy as the loss function and the Adam optimizer with a batch size of 64. Training was performed for 900 epochs. Unlike the SoC estimation models presented previously, no early-stopping or adaptive learning-rate scheduling strategies were applied.

To evaluate model performance, Accuracy (Equation (11)), Precision (Equation (12)), Recall (Equation (13)) and F1-score (Equation (14)) were considered:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{13}$$

$$\text{F}_{1\ \text{score}} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

These metrics were computed from the confusion matrix, where TP, TN, FP and FN denote true positives, true negatives, false positives and false negatives, respectively.

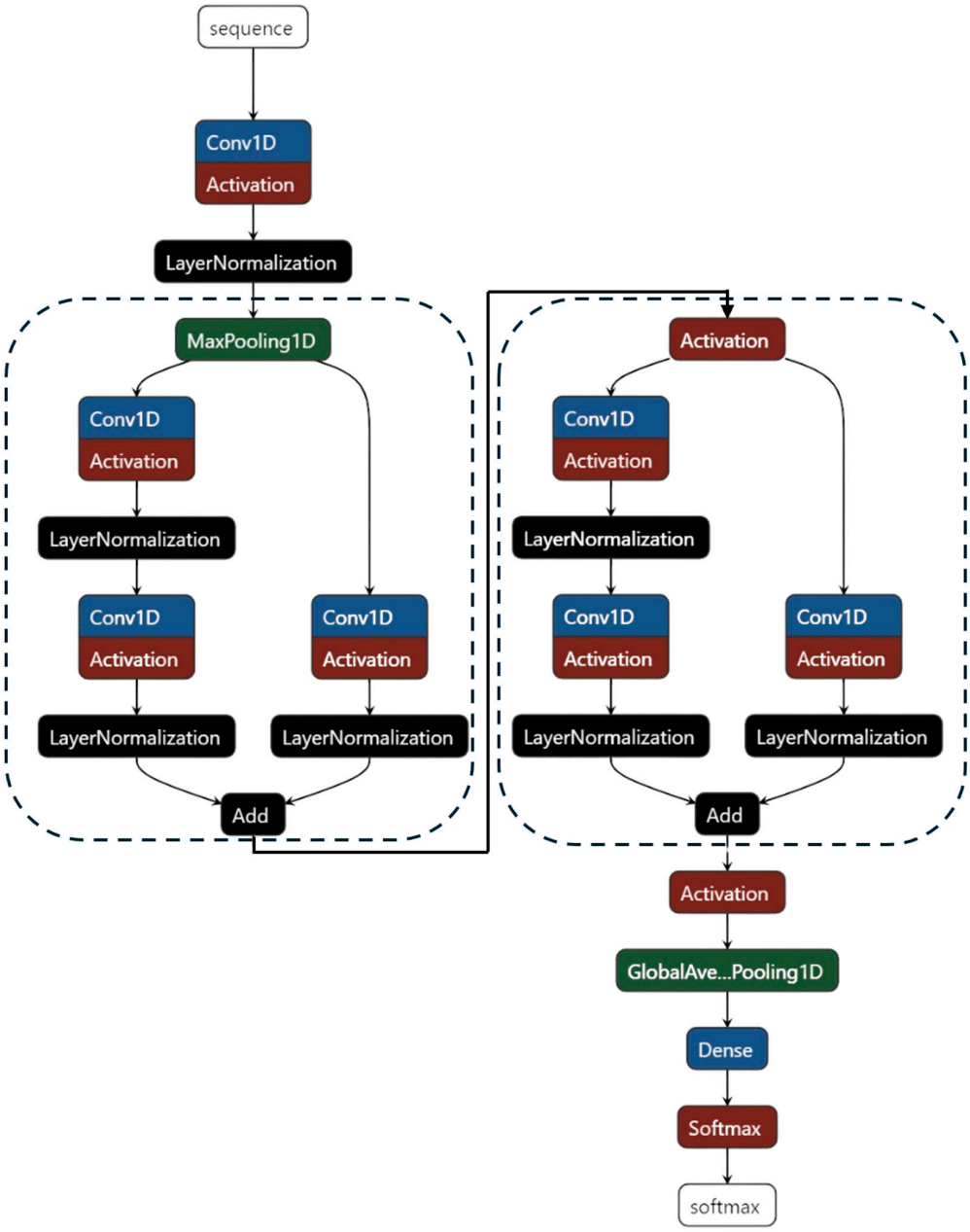


Figure 9. Resnet proposed architecture to perform SoH classification.

Although all metrics are reported for completeness, overall classification accuracy was considered the primary performance indicator throughout this study.

3.4.4. SoH Classification Using Reference SoC Values

As a first experiment, the ResNet classifier was evaluated under ideal conditions using the reference SoC values provided by the dataset. The input feature vector was therefore defined as:

$$\varnothing_k = [V_k,\ I_k, T_k, SoC_k] \tag{15}$$

The resulting confusion matrix is shown in Figure 10, while the corresponding classification report is summarized in Table 6.

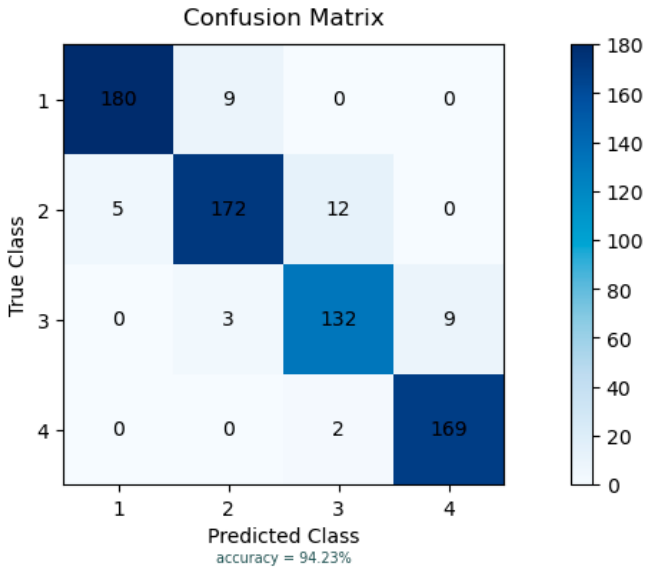


Figure 10. Confusion matrix obtained using reference SoC values.

The proposed model achieved an overall classification accuracy of 94.23%, with F1-scores above 0.90 for all health categories. The best performance was obtained for Classes 1 and 4, which achieved F1-scores of 0.96 and 0.97, respectively.

These results demonstrate that the ResNet architecture can effectively exploit the information contained in the SoC signal when accurate SoC measurements are available.

However, this scenario represents an upper performance bound rather than a realistic deployment condition, since direct SoC measurements are typically unavailable in practical BMS.

Table 6. Classification report for the LEV 50 test dataset.

Class	Precision	Recall	F1-Score
1 (95% $\leq$ SoH $\leq$ 100%)	0.97	0.95	0.96
2 (90% $\leq$ SoH < 95%)	0.93	0.91	0.92
3 (85% $\leq$ SoH < 90%)	0.90	0.92	0.91
4 (80% $\leq$ SoH < 85%)	0.95	0.99	0.97
Accuracy			0.94

3.4.5. SoC Estimation for the LEV50 Dataset

In practical applications, SoC must first be estimated from available measurements such as voltage, current and temperature before being used for higher-level diagnostic tasks.

Therefore, an independent SoC estimation stage was implemented using the LSTM-based framework previously validated on the LG-CHEM and Panasonic datasets.

After evaluating several input configurations, the best results were obtained using Configuration 2 (Table 7), which combines temperature measurements with Butterworth-filtered and EMA-filtered voltage and current signals, as illustrated in Figure 11. The selected architecture achieved an overall SoC prediction accuracy of 99.71% on the LEV50 test dataset (Figure 12). Figure 13 depicts the confusion matrix for SoC estimation.

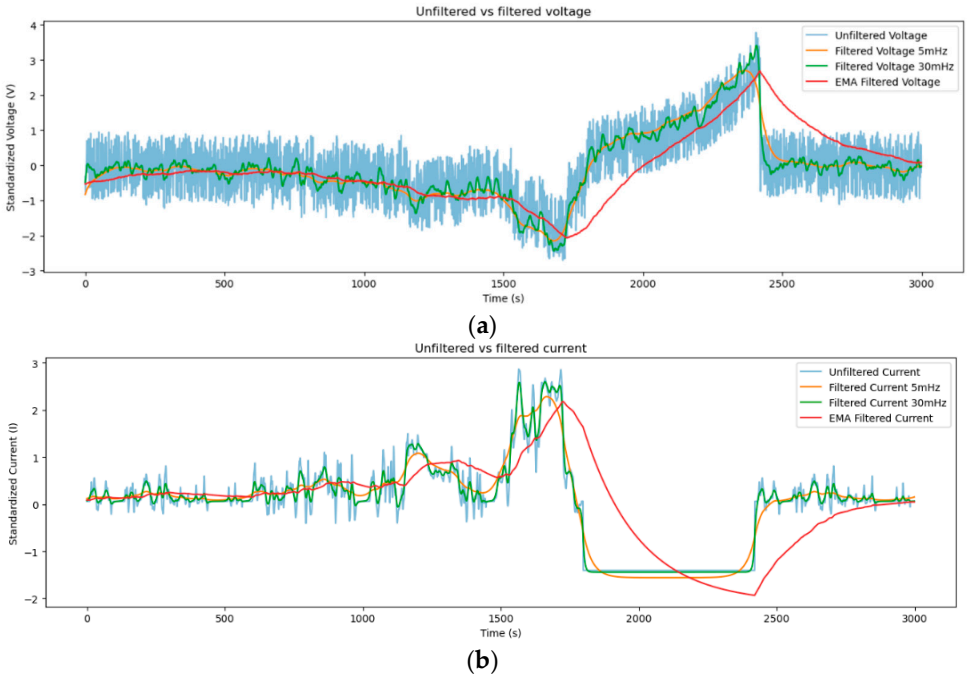


Figure 11. Comparison between raw and filtered signals: (**a**) voltage; (**b**) current.

Table 7. Final neural network architecture and hyperparameters for the LEV50 Dataset.

LSTM Neural Network	
Input configuration (Conf. 2)	$\varnothing_k = [T_k, V_{EMA,k}, I_{EMA,k}, V_{f1,k}, I_{f1,k}, V_{f2,k}, I_{f2,k}]$
Hidden layers	1
Units in the hidden layer	22
Timesteps	3
Loss function	MSE
Loss function optimizer	ADAM
Learning Rate	0.01
Training epochs	50
Butterworth filter	
Cut-off frequency 1	5 mHz
Cut-off frequency 2	30 mHz
EMA filter	
τ	200

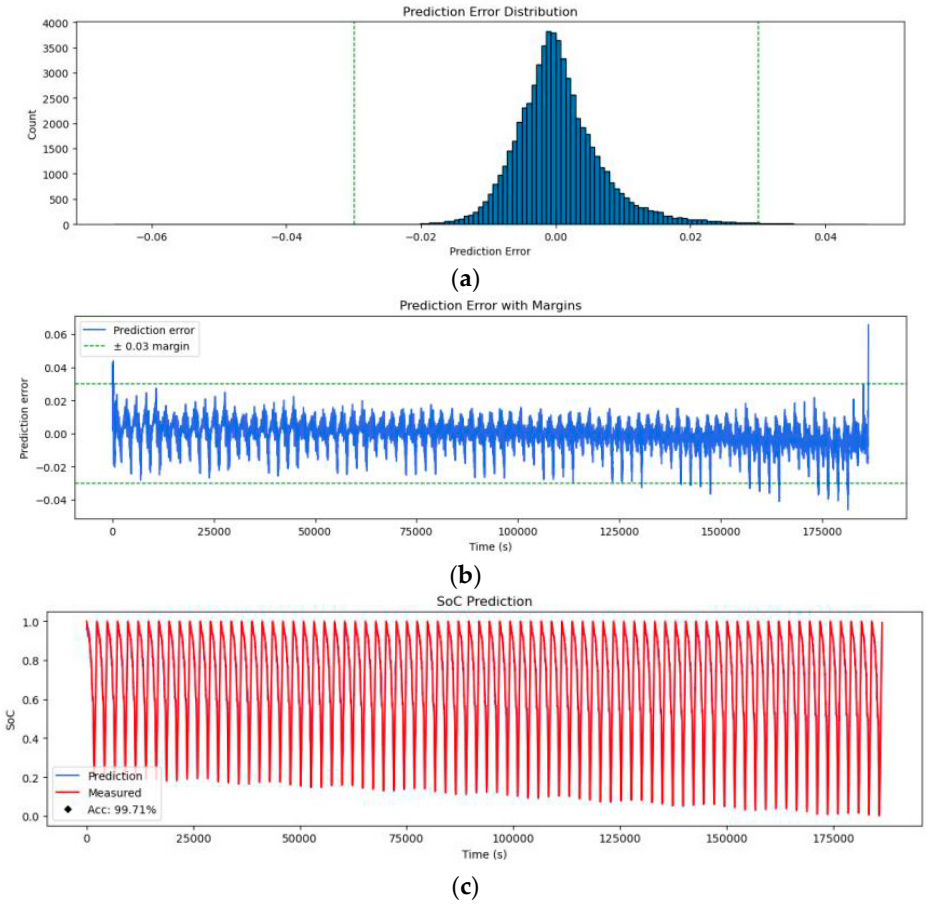


Figure 12. SoC prediction results on the LEV50 test dataset: (**a**) prediction error distribution; (**b**) prediction error with margins; (**c**) SoC prediction. The green dashed lines in (**a**,**b**) indicate the ± 0.03 error margin. The diamond marker in (**c**) denotes the 99.71% prediction accuracy achieved in this test.

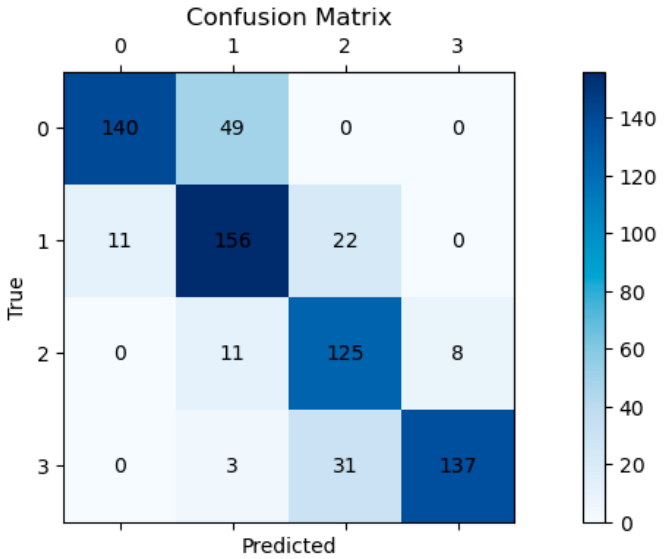


Figure 13. Confusion matrix obtained for the SoC estimation task.

3.4.6. SoH Classification Using Estimated SoC Values

To reproduce a realistic BMS deployment scenario, the reference SoC values were replaced by the outputs generated by the LSTM-based SoC estimator. The resulting feature vector becomes:

$$\Sigma\varnothing_k{}' = \left[V_{k'},\ I_k, T_k, \widehat{SoC_k}\right], \tag{16}$$

where $\widehat{SoC_k}$ denotes the estimated SoC value.

The classification results obtained under these conditions are summarized in Table 8.

Table 8. Classification report for the LEV50 dataset using estimated SoC values.

Class	Precision	Recall	F1-Score
1 (95% $\leq$ SoH $\leq$ 100%)	0.98	0.73	0.84
2 (90% $\leq$ SoH < 95%)	0.73	0.90	0.81
3 (85% $\leq$ SoH < 90%)	0.64	0.80	0.74
4 (80% $\leq$ SoH < 85%)	0.95	0.68	0.75
Accuracy			0.79

The overall SoH classification accuracy decreased from 94.23% to 79.65% when estimated SoC values were used instead of the reference measurements.

A detailed analysis reveals that Classes 1 and 4 preserve very high precision but experience a substantial reduction in recall, indicating a more conservative classification behaviour. In contrast, Classes 2 and 3 exhibit higher recall values coupled with lower precision, suggesting an increased tendency to overpredict intermediate battery states.

Notably, this performance degradation occurs despite the SoC estimator achieving a prediction accuracy of 99.71%. This observation indicates that even relatively small errors in SoC estimation can propagate through the subsequent SoH classification stage and significantly affect the final decision-making process.

3.5. Limitations

It should be noted that the SoH classifier was trained using the reference SoC generated by the simulation model, whereas the cascaded evaluation employed LSTM-estimated SoC values. Therefore, the reported reduction in classification performance reflects the practical impact of replacing ideal SoC information with estimated SoC values, although part of the observed degradation may also be associated with an input distribution shift between training and deployment conditions.

The present work focuses on uncertainty propagation in a cascaded SoC–SoH estimation framework rather than on feature selection for SoH classification. Therefore, a baseline classifier excluding SoC was not considered. While such a comparison would provide additional insight into the contribution of SoC information, it would not directly address the main objective of the study, namely quantifying the effect of replacing ideal SoC values with estimated SoC values on downstream SoH classification performance. Nevertheless, this constitutes a limitation of the current work and will be investigated in future research through feature-ablation analyses.

Furthermore, the adopted architecture follows a cascaded SoC–SoH paradigm and therefore does not exploit the potential advantages of multitask-learning approaches, where shared feature representations and joint optimization enable bidirectional interactions between both tasks. Consequently, the present study does not provide a quantitative comparison between cascaded and joint-learning frameworks, and the reported results should not be interpreted as evidence that cascaded architectures outperform multitask-learning alternatives.

An important limitation of the present study is that the uncertainty-propagation analysis was performed using the LEV50 dataset, which was generated from a validated battery simulation model rather than from experimental ageing measurements. Although the simulation framework incorporates realistic driving cycles, charging procedures, temperature variations, ageing progression, and measurement noise, real batteries may exhibit additional degradation mechanisms and variability not fully captured by the model. Consequently, the observed reduction in SoH classification accuracy should be interpreted as evidence of uncertainty-propagation effects within the evaluated simulation framework and should be further validated using experimental long-term ageing datasets.

Hyperparameter selection was guided through an exploratory tuning process supported by Optuna. However, the objective of this work is not to provide a comprehensive hyperparameter optimization study or to identify globally optimal configurations. Instead, the focus is placed on evaluating the impact of preprocessing strategies and SoC estimation uncertainty within a cascaded SoC-SoH framework. Consequently, the selected hyperparameters should be interpreted as representative configurations that enabled the experimental assessment reported in this study rather than as optimal solutions.

Likewise, the weighting strategy adopted for class-imbalance mitigation was selected empirically and was not derived from a formal Bayesian risk-minimization framework. Although this approach proved effective for the purposes of the present study, future work may investigate cost-sensitive learning strategies and theoretically grounded weighting schemes.

4. Discussion

The experiments performed on the LEV50 dataset provide two main findings. First, the proposed ResNet architecture achieves excellent SoH classification performance when accurate SoC information is available, reaching an overall accuracy of 94.23%. This result indicates that SoC contains information relevant to battery degradation and can effectively support battery-health assessment when reliable charge-state information is provided.

Second, replacing measured SoC values with estimated SoC values reduces the SoH classification accuracy to 79.65%, despite the SoC estimator achieving a tolerance-based accuracy of 99.71%. This result indicates that even small deviations in an intermediate state estimate can have a disproportionate effect on a downstream classification task.

From a system-level perspective, these results suggest that the performance of a cascaded diagnostic framework cannot be assessed solely from the accuracy of its individual components. The uncertainty characteristics of an intermediate estimation stage can directly influence the reliability of subsequent diagnostic decisions. Consequently, improvements in SoH classification performance may depend not only on the design of the SoH classifier itself but also on the robustness and uncertainty characteristics of the preceding SoC estimation process. This observation is consistent with recent reviews highlighting that battery-state estimation accuracy is strongly influenced by data quality, model uncertainty, operating conditions, and battery-ageing effects [10,11].

The inclusion of SoC information is motivated by the reported coupling between SoC and SoH and its widespread use in practical BMSs. Previous studies have explored this relationship through integrated estimation frameworks [13,14]. Wu et al. [13] reported improved battery-state prediction through joint estimation, while Wei et al. [14] showed that the battery ageing can significantly affect SoC estimation. In contrast, the present study considers a cascaded architecture and explicitly evaluates the effect of SoC estimation uncertainty on downstream SoH classification.

An important implication of these findings is that highly accurate SoC estimation alone does not necessarily guarantee equivalent SoH classification performance. These results

suggest that preserving degradation-relevant information throughout the estimation chain may be as important as minimizing the SoC estimation error itself. Therefore, the future battery diagnostic systems should consider not only estimation accuracy metrics but also the robustness of information transfer between successive battery-state estimation stages.

The uncertainty-propagation analysis was conducted using the validated GS Yuasa LEV50 simulation dataset, whereas the SoC estimation framework was evaluated using experimental LG Chem E66 and Panasonic NCR18650PF datasets.

5. Conclusions and Future Work

This work investigated the impact of SoC estimation uncertainty on downstream SoH classification within a cascaded battery-management framework. The results show that accurate SoC estimation does not necessarily translate into equivalent SoH classification performance when the estimated state is subsequently used by a downstream diagnostic model.

The comparison between reference and estimated SoC inputs demonstrates that the reliability of a cascaded battery-state estimation system depends not only on the accuracy of individual models but also on the robustness of the information transferred between successive stages. Within the architectures and datasets evaluated in this study, signal preprocessing and input representation also showed a substantial influence on estimation performance.

From a practical BMS perspective, these findings highlight the importance of evaluating uncertainty propagation when designing cascaded data-driven diagnostic systems. In particular, intermediate estimation errors should be considered when assessing the reliability of subsequent battery-health decisions.

The main limitation of this study is that the SoH uncertainty-propagation analysis was conducted using a validated simulation dataset rather than experimental ageing data. Future work should therefore evaluate the proposed cascaded framework using real battery-ageing datasets and investigate approaches that explicitly account for intermediate-state uncertainty, including joint SoC–SoH learning strategies.

Author Contributions: Conceptualization, M.M.-P. and V.R.; methodology, M.M.-P., V.R. and E.P.; software, M.B.; validation, M.B. and M.M.-P.; formal analysis, M.B. and M.M.-P.; investigation, M.B.; resources, V.R. and E.P.; data curation, M.B., V.R. and E.P.; writing—original draft preparation, M.B.; writing—review and editing, M.M.-P. and V.R.; visualization, M.B.; supervision, M.M.-P. and V.R.; project administration, M.M.-P.; funding acquisition, V.R. and E.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study were obtained from a combination of publicly available sources, cited in the reference list, and proprietary datasets provided by the laboratories of the Politecnico di Torino. Due to data ownership and institutional restrictions, the datasets generated and/or analyzed during the current study are not publicly available. Requests for additional information should be directed to the corresponding author and may require approval from the data owners.

Acknowledgments: During the preparation of this study, the authors used MS Copilot Agent to improve the readability and linguistic quality of the manuscript and Optuna v3.6.0 to automate hyperparameter optimization of the machine learning models. These tools were used solely as assistance tools and did not determine the scientific methodology, analysis, interpretation of results,

or conclusions. All scientific decisions and the final manuscript content were reviewed and approved by the authors, who take full responsibility for the published work.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ANN	Artificial Neural Network
BOL	Beginning of Life
BiLSTM	Bidirectional Long Short-Term Memory
BMS	Battery Management System
CNN	Convolutional Neural Network
CNN1D	One-Dimensional Convolutional Neural Network
EFC	Equivalent Full Cycles
EMA	Exponential Moving Average
EV	Electric Vehicle
FN	False Negative
FNN	Feedforward Neural Network
FP	False Positive
GITT	Galvanostatic Intermittent Titration Technique
GRU	Gated Recurrent Unit
HPPC	Hybrid Pulse Power Characterization
HWFET	Highway Fuel Economy Test
LA92	LA92 Unified Dynamometer Driving Schedule
LSTM	Long Short-Term Memory
MSE	Mean Squared Error
NN	Neural Network
RCID	Rate Capacity Incremental Discharge
ResNet	Residual Neural Network
RNN	Recurrent Neural Network
SoC	State of Charge
SoE	State of Energy
SoH	State of Health
TN	True Negative
TP	True Positive
UDDS	Urban Dynamometer Driving Schedule
US06	US06 Supplemental Federal Test Procedure
WLTP	Worldwide Harmonized Light Vehicles Test Procedure

References

1. Nath, A.; Gupta, R.; Mehta, R.; Bahga, S.S.; Gupta, A.; Bhasin, S. Attractive Ellipsoid Sliding Mode Observer Design for State of Charge Estimation of Lithium-Ion Cells. *IEEE Trans. Veh. Technol.* **2020**, *69*, 14701–14712. [CrossRef]
2. Huang, C.-S. An Adaptive Moving Window Total Least Square Method Considering Load Current Condition Aimed at Online Lithium-Ion Battery State-of-Health Estimation. *IEEE Trans. Veh. Technol.* **2024**, *73*, 18539–18547. [CrossRef]
3. Kim, W.; Lee, P.-Y.; Kim, J.; Kim, K.-S. A Robust State of Charge Estimation Approach Based on Nonlinear Battery Cell Model for Lithium-Ion Batteries in Electric Vehicles. *IEEE Trans. Veh. Technol.* **2021**, *70*, 5638–5647. [CrossRef]
4. Obeid, H.; Petrone, R.; Chaoui, H.; Gualous, H. Higher Order Sliding-Mode Observers for State-of-Charge and State-of-Health Estimation of Lithium-Ion Batteries. *IEEE Trans. Veh. Technol.* **2023**, *72*, 4482–4492. [CrossRef]
5. Xue, Q.; Shen, S.; Li, G.; Zhang, Y.; Chen, Z.; Liu, Y. Remaining Useful Life Prediction for Lithium-Ion Batteries Based on Capacity Estimation and Box-Cox Transformation. *IEEE Trans. Veh. Technol.* **2020**, *69*, 14765–14779. [CrossRef]
6. Guo, R.; Shen, W. A Model Fusion Method for Online State of Charge and State of Power Co-Estimation of Lithium-Ion Batteries in Electric Vehicles. *IEEE Trans. Veh. Technol.* **2022**, *71*, 11515–11525. [CrossRef]

7. He, Z.; Yang, Z.; Cui, X.; Li, E. A Method of State-of-Charge Estimation for EV Power Lithium-Ion Battery Using a Novel Adaptive Extended Kalman Filter. *IEEE Trans. Veh. Technol.* **2020**, *69*, 14618–14630. [CrossRef]
8. An, F.; Jiang, J.; Zhang, W.; Zhang, C.; Fan, X. State of Energy Estimation for Lithium-Ion Battery Pack via Prediction in Electric Vehicle Applications. *IEEE Trans. Veh. Technol.* **2022**, *71*, 184–195. [CrossRef]
9. Li, N.; He, F.; Ma, W.; Wang, R.; Jiang, L.; Zhang, X. An Indirect State-of-Health Estimation Method Based on Improved Genetic and Back Propagation for Online Lithium-Ion Battery Used in Electric Vehicles. *IEEE Trans. Veh. Technol.* **2022**, *71*, 12682–12690. [CrossRef]
10. Roy, P.K.; Shahjalal, M.; Shams, T.; Fly, A.; Stoyanov, S.; Ahsan, M.; Haider, J. A Critical Review on Battery Aging and State Estimation Technologies of Lithium-Ion Batteries: Prospects and Issues. *Electronics* **2023**, *12*, 4105. [CrossRef]
11. Miraftabzadeh, S.M.; Longo, M.; Di Martino, A.; Saldarini, A.; Faranda, R.S. Exploring the Synergy of Artificial Intelligence in Energy Storage Systems for Electric Vehicles. *Electronics* **2024**, *13*, 1973. [CrossRef]
12. Xue, Q.; Li, J.; Chen, Z.; Zhang, Y.; Liu, Y.; Shen, J. Online Capacity Estimation of Lithium-Ion Batteries Based on Deep Convolutional Time Memory Network and Partial Charging Profiles. *IEEE Trans. Veh. Technol.* **2023**, *72*, 444–457. [CrossRef]
13. Wu, L.; Chen, C.; Li, Z.; Chen, Z.; Li, H. The Joint Estimation of SOC-SOH for Lithium-Ion Batteries Based on BiLSTM-SA. *Electronics* **2025**, *14*, 97. [CrossRef]
14. Wei, Z.; Sun, X.; Li, Y.; Liu, W.; Liu, C.; Lu, H. A Joint Estimation Method for the SOC and SOH of Lithium-Ion Batteries Based on AR-ECM and Data-Driven Model Fusion. *Electronics* **2025**, *14*, 1290. [CrossRef]
15. How, D.N.T.; Hannan, M.A.; Hossain Lipu, M.S.; Ker, P.J. State of Charge Estimation for Lithium-Ion Batteries Using Model-Based and Data-Driven Methods: A Review. *IEEE Access* **2019**, *7*, 136116–136136. [CrossRef]
16. Lucaferri, V.; Quercio, M.; Laudani, A.; Riganti Fulginei, F. A Review on Battery Model-Based and Data-Driven Methods for Battery Management Systems. *Energies* **2023**, *16*, 7807. [CrossRef]
17. Huang, Q.; Wang, S.; Chen, Z.; Xiong, R.; Fernandez, C.; Stroe, D.-I. *Long-Term Health State Estimation of Energy Storage Lithium-Ion Battery Packs*; Springer Nature: Singapore, 2023. [CrossRef]
18. Jiang, N.; Pang, H. Study on Co-Estimation of SoC and SoH for Second-Use Lithium-Ion Power Batteries. *Electronics* **2022**, *11*, 1789. [CrossRef]
19. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
20. López, V.; Jove, E.; Gato, F.Z.; Pinto-Santos, F.; Piñón-Pazos, A.J.; Casteleiro-Roca, J.-L.; Quintian, H.; Calvo-Rolle, J.L. Intelligent Model for Power Cells State of Charge Forecasting in EV. *Processes* **2022**, *10*, 1406. [CrossRef]
21. Gruber, N.; Jockisch, A. Are GRU Cells More Specific and LSTM Cells More Sensitive in Motive Classification of Text? *Front. Artif. Intell.* **2020**, *3*, 40. [CrossRef] [PubMed] [PubMed Central]
22. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2016. [CrossRef]
23. Kollmeyer, P.J. Panasonic 18650PF Li-ion Battery Data. In *Mendeley Data*; Elsevier: Amsterdam, The Netherlands, 2018. [CrossRef]
24. Vidal, C.; Kollmeyer, P.J. *Panaso*nic 18650PF Li-ion Battery Data and Example FNN and LSTM Neural Network SOC Estimator Training Script. In *Mendeley Data*; Elsevier: Amsterdam, The Netherlands, 2021. [CrossRef]
25. Sherkatghanad, Z.; Ghazanfari, A.; Makarenkov, V. A self-attention-based CNN-Bi-LSTM model for accurate state-of-charge estimation of lithium-ion batteries. *J. Energy Storage* **2024**, *88*, 111524. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.