EUROPEAN COMMISSION

HORIZON EUROPE PROGRAMME – TOPIC: HORIZON-CL5-2022-D2-01

FASTEST

Fast-track hybrid testing platform for the development of battery systems

Deliverable D3.3: Data-driven model development and validation for virtualization of performance and ageing testing

Primary Author Doniyor Urishov

Organization VTT

Date: [11/05/2026]

Doc.Version: [v1.0]

Co-funded by the European Union

UK Research and Innovation

Co-funded by the European Union under grant agreement N° 101103755 and by UKRI under grant agreement No. 10078013, respectively. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor CINEA can be held responsible for them.

Document Control Information	
Settings	Value
Work package:	WP3
Deliverable:	Data-driven model development and validation for virtualization of performance and ageing testing
Deliverable Type:	Report
Dissemination Level:	Public
Due Date:	31.05.2026 (M36)
Actual Submission Date:	11.05.2026
Pages:	37
Doc. Version:	V1.0
GA Number:	101103755
Project Coordinator:	Bruno Rodrigues | ABEE (bruno.rodrigues@abeegroup.com)

Formal Reviewers		
Name	Organization	Date
Philip Brendel	FHG	15.04.2026
Cinar Batuhan	Surrey	29.04.2026
Mariana Fernandez	SIE	30.04.2026
Maria Grazia Toma	Comau	05.05.2026
Antonio Silvio de Letteriis	Flash Battery	06.05.2026

Document History			
Version	Date	Description	Author
0.1	28.01.2026	Draft version, structure of the document.	Doniyor Urishov
0.2	02.03.2026	Section 4.1	Laura Oca
0.3	31.03.2026	Section 4.2	Akhtar Zeb
0.4	15.04.2026	Section 4.3	Philip Brendel
0.5	29.04.2026	Section 4.4	Cinar Batuhan
0.6	30.04.2026	Quality review	COMAU, Flashbattery, SIE
1.0	11.05.2026	Conclusion	Pankaj Saha

Project Abstract

Current methods to evaluate Li-ion batteries safety, performance, reliability and lifetime represent a remarkable resource consumption for the overall battery R&D process. The time or number of tests required, the expensive equipment and a generalised trial-error approach are determining factors, together with a lack of understanding of the complex multiscale and multi-physics phenomena in the battery system. Besides, testing facilities are operated locally, meaning that data management is handled directly in the facility, and that experimentation is done on one test bench.

The FASTEST project aims develop and validate a fast-track testing platform able to deliver a strategy based on Design of Experiments (DoE) and robust testing results, combining multi-scale and multi-physics virtual and physical testing. This will enable an accelerated battery system R&D and more reliable, safer and long-lasting battery system designs. The project's prototype of a fast-track hybrid testing platform aims for a new holistic and interconnected approach. From a global test facility perspective, additional services like smart DoE algorithms, virtualised benches, and DT data are incorporated into the daily facility operation to reach a new level of efficiency.

During the project, FASTEST consortium aims to develop up to TRL 6 the platform and its components: the optimal DoE strategies according to three different use cases (automotive, stationary, and off-road); two different cell chemistries, 3b and 4 solid-state (oxide polymer electrolyte); the development of a complete set of physic-based and data-driven models able to substitute physical characterisation experiments; and the overarching Digital Twin architecture managing the information flows, and the TRL6 proven and integrated prototype of the hybrid testing platform.

LIST OF ABBREVIATIONS, ACRONYMS AND DEFINITIONS

Acronym	Name
Ah	Ampere-hour
AI	Artificial Intelligence
BMS	Battery Management System
CC	Constant Current
CC-CV	Constant Current – Constant Voltage
CNN	Convolutional Neural Network
DoE	Design of Experiments
DT	Digital Twin
EOL	End of Life
eSOH	Electrode-specific State of Health
FMU	Functional Mock-up Unit
GA	Grant Agreement
HiL	Hardware-in-the-Loop
KPIs	Key Performance Indicators
Li-ion	Lithium-ion
LIMS	Laboratory Information Management System
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
RMSE	Root Mean Squared Error
ROM	Reduced Order Model
RUL	Remaining Useful Life
SiL	Software-in-the-Loop
SoC	State of Charge
SoH	State of Health
TRL	Technology Readiness Level
UUT	Unit Under Test
WLTP	Worldwide Harmonized Light Vehicles Test Procedure

LIST OF TABLES

LIST OF FIGURES

Table of Contents

1. EXECUTIVE SUMMARY

Deliverable D3.3 addresses data-driven model development and validation for virtualization of battery performance and ageing testing within the FASTEST project. The work is conducted under Work Package 3 (Advanced battery ageing and performance modelling) and specifically targets Task 3.3, complementing the physics-based and reduced-order modelling activities addressed in Tasks 3.1 and 3.2.

The main objective of D3.3 is to develop computationally efficient, accurate, and robust data-driven models capable of substituting a significant share of time- and resource-intensive physical battery tests. The work presented in these deliverable focuses on both state estimation (e.g. State of Charge and State of Health) and lifetime prediction (Remaining Useful Life), as well as hybrid physics-informed approaches that link measurable signals to physically interpretable degradation parameters.

Several complementary modelling approaches are developed and validated using data provided by FASTEST industrial partners and physics-based simulation tools. These include LSTM-based SoC estimation models, machine-learning-based SOH and RUL prediction models, and physics-informed deep operator networks enabling fast and interpretable electrode-specific SOH estimation. The models are validated on multiple battery chemistries and operational conditions relevant to FASTEST use cases (automotive, stationary, and off-road).

The final models are implemented in formats compatible with Digital Twin architectures and hybrid testing workflows, including export as Functional Mock-up Units (FMUs) for integration with simulation and Hardware-in-the-Loop environments. The results demonstrate that data-driven and physics-assisted models can achieve high accuracy while reducing computational cost by orders of magnitude compared to conventional numerical solvers.

Overall, D3.3 contributes directly to FASTEST objectives by enabling accelerated battery R&D, reduced physical testing effort, and scalable virtual testing, supporting the advancement of the hybrid testing platform towards TRL 6.

2. OBJECTIVES

The objectives of Deliverable D3.3 are aligned with the overall FASTEST project objectives, the WP3 objectives, and the specific scope of Task 3.3:

- To develop, validate, and demonstrate data-driven and physics-assisted battery models that can reliably virtualize performance, ageing, and lifetime testing across cell, module, and pack levels, contributing to reduced testing time and cost in battery system development.
- Develop data-driven models for key battery states (SoC, SoH, RUL) using experimental and simulated datasets representative of FASTEST use cases.
- Integrate physics-based information into data-driven models, enabling improved robustness, interpretability, and generalisation across operating conditions.
- Validate model accuracy and generalisation across different battery chemistries, temperatures, load profiles, and ageing stages.
- Achieve computational efficiency suitable for real-time or near-real-time virtual testing and hybrid physical-virtual testing workflows.
- Ensure compatibility with FASTEST Digital Twin architecture, including deployment as FMUs for co-simulation, HiL, and SiL environments.
- Contribute to FASTEST KPIs, especially those related to test virtualization ratio, performance accuracy, and reduction of physical testing effort.

The fulfilment of these objectives directly supports FASTEST Objectives O1–O4, particularly the substitution of critical physical characterisation experiments through validated virtual models.

3. INTRODUCTION

The development and qualification of advanced battery systems require extensive testing to evaluate performance, ageing, safety, and reliability across multiple operating conditions and system levels. Conventional battery testing methodologies rely heavily on long-duration physical experiments, resulting in substantial time, cost, and resource consumption during the R&D phase.

Within the FASTEST project, hybrid testing combining physical and virtual testing supported by Digital Twin technologies is proposed as a key enabler for accelerating battery development while maintaining reliability and safety. In this context, data-driven models play a central role, as they offer the potential for fast execution, scalable deployment, and direct integration with digital infrastructures.

While physics-based battery models provide deep physical insight, their computational cost and parametrization effort can limit applicability in large-scale testing campaigns. Data-driven models, on the other hand, can exploit large datasets to learn complex system behaviour but often lack physical interpretability and robustness outside the training domain. FASTEST therefore adopts a hybrid modelling strategy, where data-driven models are informed and constrained by physics-based knowledge.

Deliverable D3.3 focuses on the development of such data-driven and physics-assisted models for battery performance and ageing virtualization. The work builds upon the physics-based and reduced-order models developed in WP3 and leverages experimental and simulated datasets generated within FASTEST. The presented approaches target practical deployment in Digital Twin-enabled test centres and support the project's ambition to virtualize a substantial fraction of battery testing activities from cell to system level.

4. DESCRIPTION OF WORK

T3.3 – Development of data-driven models from physics-based models focuses on the development of computationally efficient, accurate, and scalable data-driven battery models that complement the physics-based and reduced-order models developed in Tasks T3.1 and T3.2. The work aims to enable virtualisation of battery performance and ageing testing, thereby reducing experimental effort while preserving accuracy required for industrial battery qualification and certification.

The activities under T3.3 are tightly coupled with the FASTEST Digital Twin and hybrid testing concept, ensuring that the developed models can be deployed in simulation, co-simulation, and HiL environments. The work covers the modelling of battery performance, ageing, SoH, and RUL across relevant operating conditions and battery chemistries.

4.1 Physics-based model and virtual cycler for model training

The aim of this section is to describe the link between T3.2 (development and implementation of reduced order models) with the T3.3 (data-driven models). Data-driven models and algorithm development rely heavily on the input data for training. The chosen profiles for training should cover all the voltage, current, and temperature window with fast and low dynamics to be able to characterize properly

the behavior of the cells. That is why; to speed up the training process, a tool based on Simulink was proposed, with the final physics-based model (including electrochemical-thermal-ageing) for different levels (cell, module and pack) coupling a virtual cycler to perform galvanostatic cycles, dynamic profiles and ageing testing. Firstly, a simplified version of the cycler was proposed to speed up the process of this model training task. After, FMU complete version and within FASTEST data structure was embedded with the model for the final hybrid platform (described in more detail in WP6 deliverables). A screenshot of the model and simplified cycler is presented in Figure 1. The same structure has been used for both chemistries (Gen3b and SSB) and (cell, module and pack) levels.

The physics-based model has four different inputs (initial SoC, initial temperature, ambient temperature, and current). The current can be varied using the simplified cycler, according to the specifications of the cells and specific profiles (maximum voltage, minimum voltage, current on charge or on discharge). The model can give as an output the voltage response to a specific profile, the cell average temperature and temperature gradient within the surface of the cell, SoC and SoH of the cell. Therefore, considering those, the necessary data for data-driven model training can be generated using this tool.

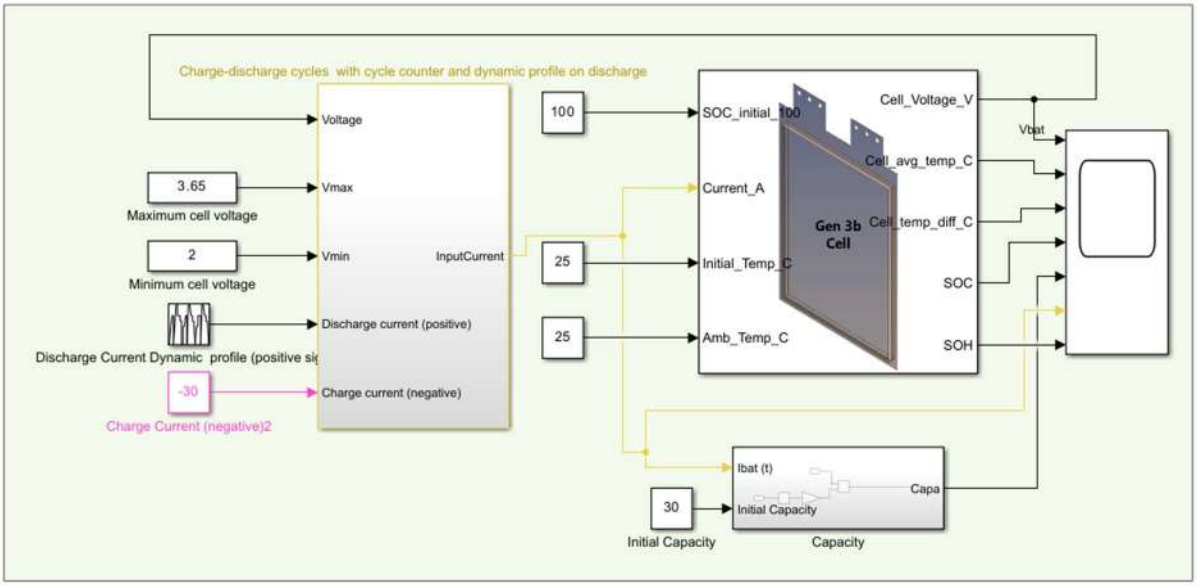

Figure 1: Cell level model of Gen3b cell coupled to simplified virtual cycler for data-driven model training.

4.2 LSTM-based SoC estimation model development

4.2.1 Objective and scope

This section describes the development of data-driven State-of-Charge (SoC) estimation models based on Long Short-Term Memory (LSTM) neural networks for two battery datasets, Generation 3b (Gen-3b) and Generation 4 (Gen-4), provided by ABEE. In both datasets, the model inputs are terminal voltage (V), current (I), and temperature (T), and the model output is the instantaneous SoC of the battery cell. The temporal evolution of the signals is captured using sliding windows with

look-back lengths between 10 and 300 samples, where each sample is recorded at one-second intervals.

4.2.2 Datasets and experimental design

The Gen-3b dataset contains cycling data from LFP cells (liquid electrolyte, ~30 Ah nominal capacity). The Gen-4 dataset consists of cycling data from NMC solid-state cells (nominal capacity ~30 Ah). For both battery generations, several cells were cycled by ABEE under different operating conditions. The cycling included experiments at temperatures of 25ºC and 45ºC and different load profiles such as CC-CV, WLTF-inspired dynamic loads, and stationary profiles. Each cell was cycled ~1000 times, and various battery parameters were recorded, including voltage, current, temperature, capacity throughput, and timestamps. No dataset included ground-truth SoC labels, so SoC had to be reconstructed.

Because supervised machine learning requires known target SoC values, SoC labels were generated using anchored Coulomb counting. The SoC was set to 100% at the end of the CV-charge phase of each full charge cycle and to 0% at the end of discharge when applicable. Between these anchor points, SoC evolves according to:

$$SoC_k = SoC_{k-1} + \frac{I_k \, \Delta t_k}{C_{ref}} \times 100\%$$

where I_k is the measured current (A), Δt_k is the sample interval (h), and C_{ref} is the reference capacity, defined as the median discharge capacity (Ah) extracted from complete CC+CV discharge cycles. This anchoring ensures drift-free and physically meaningful SoC targets for model training.

4.2.3 Problem formulation and windowing

Let $x_t = [V_t, I_t, T_t]$. For each time index t, an input tensor $X_t = \{x_{t-L+1}, \ldots, x_t\}$ was constructed using a look-back window $L \in \{10, 20, \ldots, 300\}$. The model predicts the scalar $\mathrm{y}_t = \mathrm{SoC}_t$. This formulation enables the LSTM to learn the temporal dynamics of battery signals without requiring handcrafted features. Figure 2 illustrates this modelling approach.

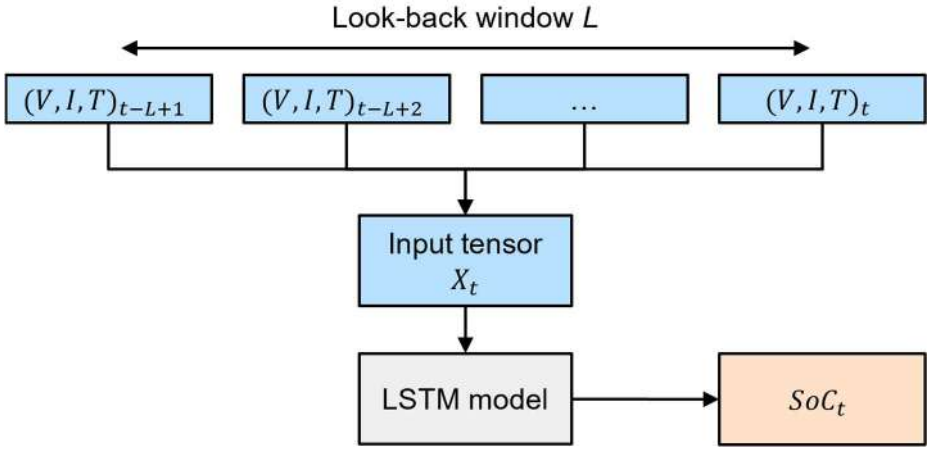

Figure 2: Structure of the LSTM input window. Each prediction uses the past L samples of voltage, current, and temperature to estimate the SoC as time t.

4.2.4 Model architecture and training setup

The model architecture consists of an input layer, two LSTM layers, two fully connected (dense) layers with ReLU activation, and a final linear output layer predicting SoC. The model was trained using the MAE loss function and the Adam optimizer. Model performance was evaluated using MAE and RMSE. All input variables (V, I, T) and the output (Soc) were scaled using a StandardScaler fitted only on the training subset. Hyperparameters such as learning rate, batch size, and look-back window length were tuned based on validation performance.

Python with TensorFlow was used for model development, scikit-learn for preprocessing, and the HyperOpt library for hyperparameter tuning. For deployment, the final models were re-implemented in Simulink and exported as FMUs for integration into co-simulation and system-level digital twin workflows.

4.2.5 Data splitting strategy

To evaluate cross-cell generalization, a cell-disjoint data split was used: one cell for training, one for validation, and one for testing. Multiple split permutations (e.g., train on Cell 2, validate on Cell 3, test on Cell 1) were evaluated to ensure robustness. The same strategy was applied to both Gen-3b and Gen-4 datasets.

4.2.6 Results

Across both datasets, the LSTM-based SoC estimators achieved MAE 0.5-1.6%, RMSE 0.7-2.1%, and maximum deviation < 3% on the held-out test cells. Figure 3 (Gen-4 example) shows time-series overlays of predicted versus reference SoC, demonstrating accurate tracking during charge, discharge, and rest periods, with small residuals at current transients. Applying the same modelling pipeline to the Gen-3b dataset resulted in comparable accuracy, also within ±2% on the test cells. Figure 4 illustrates the FMU of the LSTM model integrated with the cell level simulation model in Simulink.

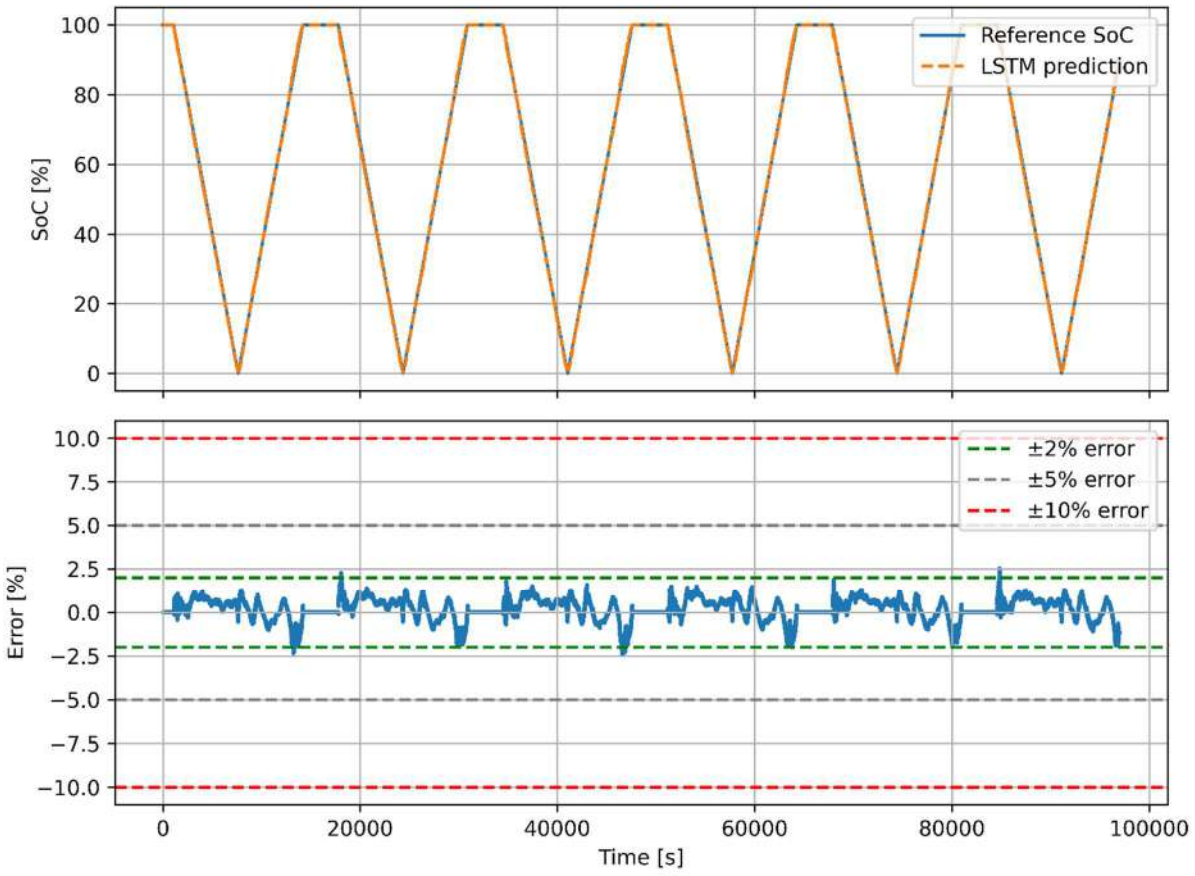

Figure 3: Predicted vs reference SoC for a Gen-4 test cell.

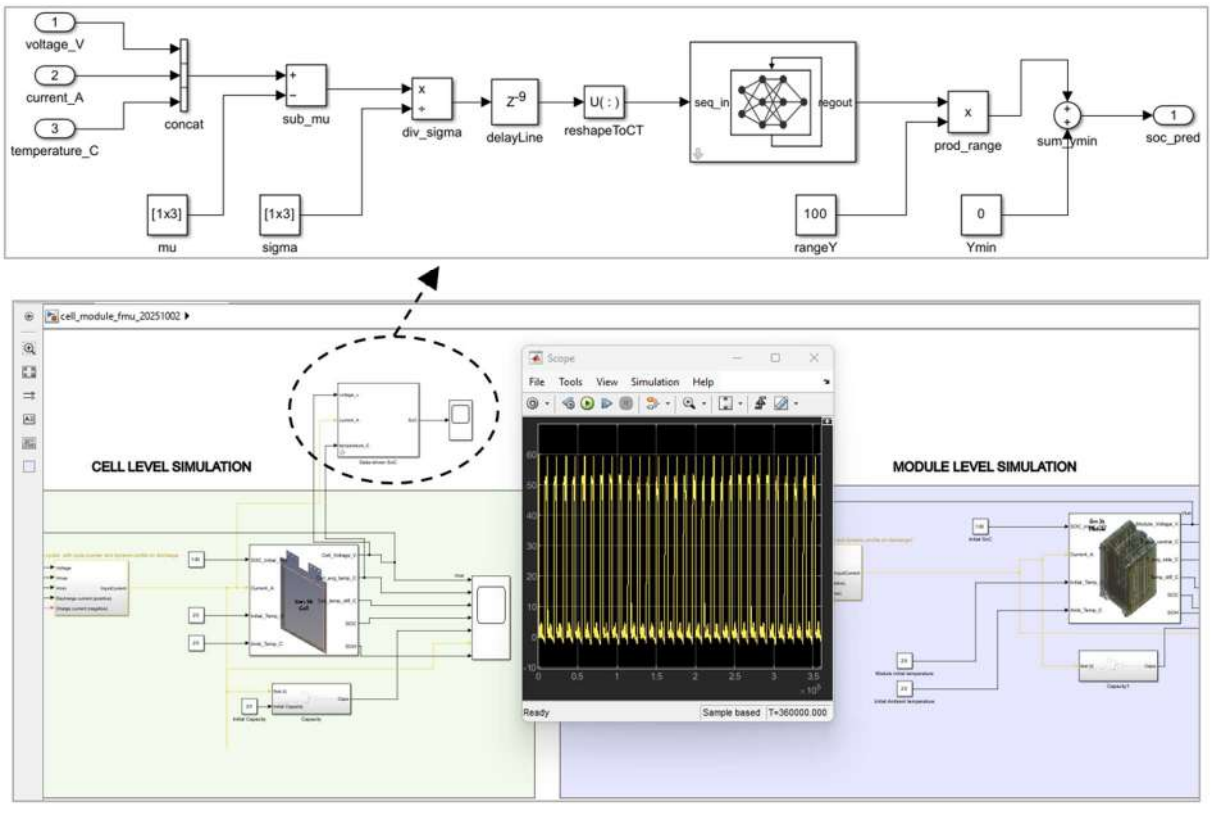

Figure 4: FMU of the LSTM-based Soc estimation integrated into the cell-level simulation model in Simulink.

Using only V-I-T inputs and LSTM architectures with short to medium look-back windows (10-time steps for Gen-4, 60-time steps for Gen-3b), the models achieved SoC-estimation accuracy within ±2% across held-out cells. The final models were exported as FMUs for integration into project-level simulations and BMS workflows.

4.3 RUL model development

This section describes the development, evaluation, and analysis of a machine-learning-based remaining useful life (RUL) prediction model using cycling data provided by ABEE. The objective was to model the state of health (SoH) evolution of battery cells during cycling and to estimate their end-of-life (EOL) and RUL.

4.3.1 Dataset description and feature extraction

The available dataset contains continuous cycling measurements recorded at 1-second intervals. Each cycle includes a constant-current (CC) discharge phase, from which features were extracted to construct a cycle-wise regression dataset. For each discharge cycle, the CC portion was isolated and used to compute the following features:

- CC discharge duration
- Average temperature (°C)
- Average voltage (V)
- Average current (A)
- Discharge capacity (Ah) – calculated as the difference between the initial and final capacity of the cycle

The target variable is SOH, computed as the discharge capacity normalized by the nominal cell capacity of 30 Ah. Thus, the machine-learning model input consists of cycle-level voltage, current, temperature, and discharge time, while the output is the corresponding SoH value.

4.3.2 ML model and experimental setup

XGBoost regression was selected for this study due to its strong performance on tabular data and its ability to model nonlinear degradation patterns. The dataset was initially divided using a random 80/20 train-test split, resulting in both sets containing a mixture of early-life, mid-life, and late-life cycles. Model performance was evaluated using standard regression metrics: RMSE, MAE, and MAPE.

4.3.3 Results

Figure 5 shows the model's performance using the random split. The predicted SoH trajectory closely matches the observed SoH for both training and test sets, successfully capturing the smooth, nonlinear capacity-fade trend. On the test set, the model achieved an RMSE of 0.45%, an MAE of 0.3%, and an MAPE of 0.005%.

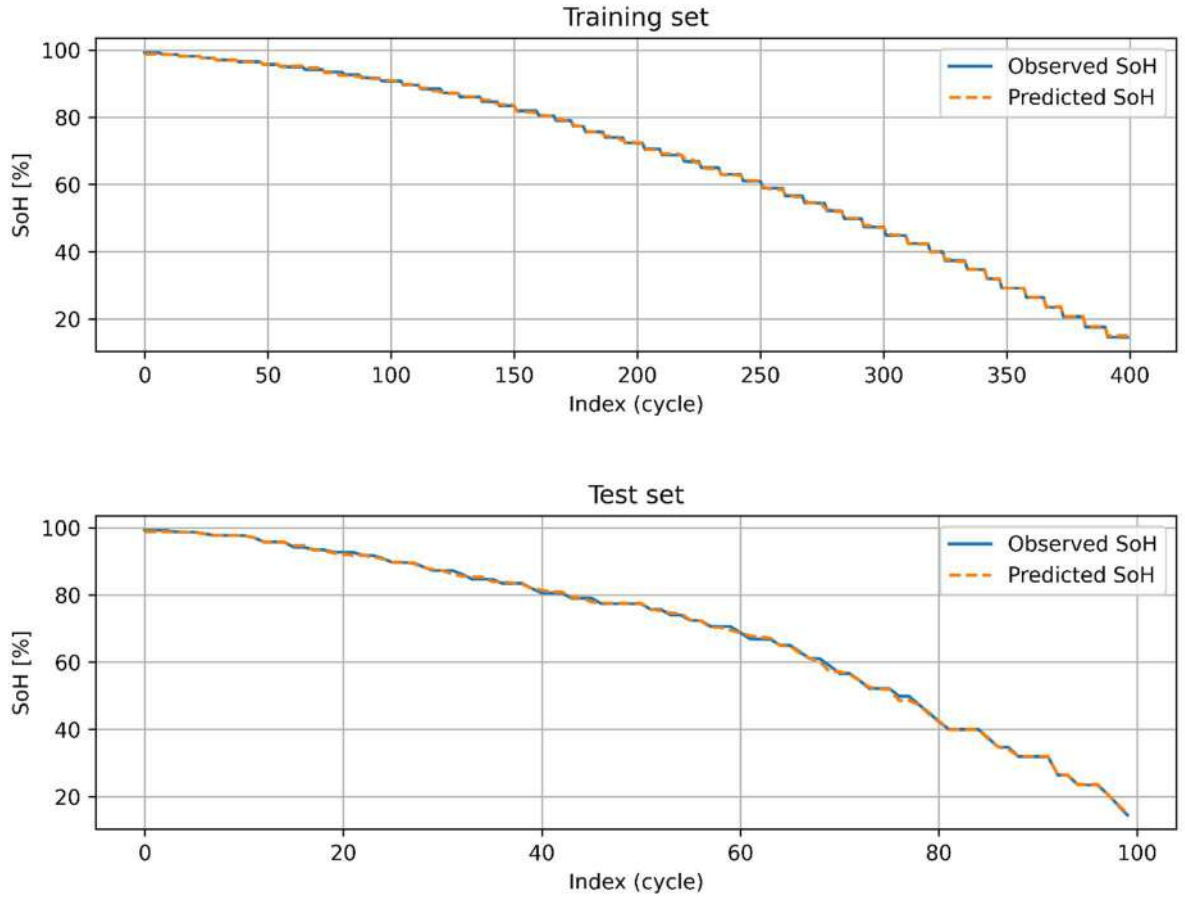

Figure 5: Model performance on randomly split training (top) and testing (bottom) datasets.

The strong performance obtained using the random split is expected. Since early, mid, and late cycles are mixed across training and testing sets, the model effectively has access to the entire degradation trajectory during training. As a result, the late-life degradation region—normally the most difficult to predict—is already known to the model. Consequently, the task becomes an interpolation problem, which tree-based models such as XGBoost handle very effectively.

However, these results do not reflect real-world deployment, where only early-life data would be available at prediction time. A realistic evaluation requires a chronological split, where early cycles are used for training and later cycles for testing. Under such conditions, the model must extrapolate future degradation, a significantly more challenging task, especially when only cycle-averaged features are available.

More advanced methods—such as models that incorporate sequential information (e.g., LSTM, GRU, Temporal Convolutional Networks, Transformers, or physics-informed neural networks)—could provide a more reliable RUL estimate. However, given the current dataset structure, such temporal modelling was not feasible.

RUL estimation can then be performed either by extrapolating the predicted SoH curve or by simulating synthetic future cycles using historical temperature, voltage, and current profiles, followed by ML-based SoH prediction.

The ML model performs exceptionally well under a random train–test split; however, this performance is misleading due to leakage of future information into the training process. When evaluated appropriately using a chronological split, the model struggles to capture late-life degradation because the current feature set lacks temporal depth and physical interpretability. Although the initial results demonstrate the potential of data-driven RUL modelling, further improvements such as incorporating time-series features, richer datasets, or sequential/physics-based models are necessary for accurate real-world RUL prediction.

4.4 Physics-Informed Neural Networks for electrode-specific SOH estimation

4.4.1 Objective and scope

As part of T3.3, a model-based framework for on-board capable State-of-Health (SOH) estimation was developed based on physics-informed machine learning. The objective was to enable accurate and computationally efficient estimation of degradation-related parameters in lithium-ion batteries using only measurable signals such as current and voltage.

To this end, the Single Particle Model (SPM) was selected as the underlying electrochemical model due to its favorable balance between physical interpretability and computational complexity. However, it is well known that direct parameter estimation in such models is hindered by limited parameter identifiability and high computational cost when classical numerical solvers are used.

To overcome these limitations, parametrized Physics-Informed Deep Operator Networks (PI-DeepONets) were developed as fast surrogate models of the SPM. These models were designed to generalize across varying operating conditions (e.g., current profiles) and parameter sets, thereby enabling both efficient simulation and gradient-based analysis. In addition, a workflow combining global design of experiments (DoE), parameter estimation, and identifiability analysis was implemented to systematically improve estimation robustness.

The presented developments were demonstrated as proof-of-concept in two open-source articles [1], [2] in which they are showcased for a publicly available parameter set obtained from measurement data [3].

4.4.2 SOH estimation via electrode-specific SOH (eSOH) parameters

For the purpose of physically meaningful SOH estimation, the concept of electrode-specific State-of-Health (eSOH) parameters was adopted following [4]. Instead of relying on aggregate health indicators, SOH was described through a set of parameters directly linked to degradation mechanisms at electrode level.

In particular, parameters such as stoichiometric limits at 0% and 100% State-of-Charge, active material volume fractions, and Lithium diffusivities were considered. These quantities are known to be closely related to dominant

degradation modes such as loss of lithium inventory (LLI) and loss of active material (LAM).

Within the developed framework, these eSOH parameters were treated as explicit inputs to the model and estimated from voltage responses. This enables a more interpretable representation of battery health, as changes in estimated parameters can be directly linked to physical degradation processes.

In addition to eSOH parameters, transport-related parameters such as solid-phase diffusivities were included, as they influence both dynamic behavior and identifiability. The resulting parameter space was therefore high-dimensional and required dedicated strategies for efficient exploration and estimation.

4.4.3 Physics-Informed Deep Operator Networks

To enable efficient evaluation of the SPM across this high-dimensional parameter space, parametrized Physics-Informed Deep Operator Networks (PI-DeepONets) were developed.

These networks approximate the mapping from:

- input functions (current profiles), and
- scalar parameters (including eSOH parameters)

to model outputs such as lithium concentration profiles and voltage responses. The architecture, cf. Figure 6, consists of two main components:

- a branch network encoding the input current profile, and
- a trunk network encoding spatio-temporal coordinates and model parameters.

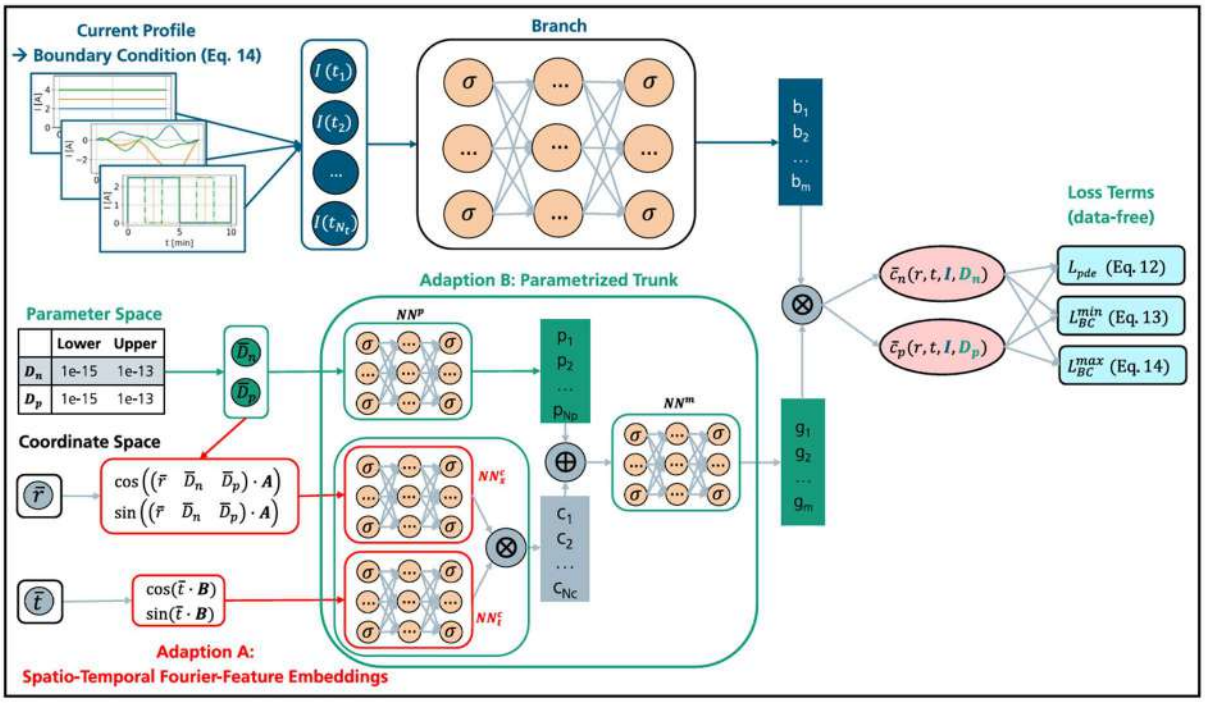

Figure 6: Parametrized PI-DeepONet. Image taken from [1].

A key feature of the approach is that the networks were trained in a purely physics-informed manner, i.e., without relying on measurement or simulation data. Instead, training was driven by minimizing the residuals of the governing partial differential equations of the SPM and associated boundary conditions.

Several architectural enhancements were implemented to improve performance and generalization, including:

- parametrization of the network with respect to electrochemical parameters,
- Fourier feature embeddings to capture multi-scale temporal and spatial dynamics, and
- hard constraints to enforce initial and boundary conditions directly within the network structure.

After training, the PI-DeepONet acts as a fast surrogate model, providing predictions of internal states and voltage responses within milliseconds. Furthermore, due to its differentiable structure, the model enables efficient computation of sensitivities with respect to parameters, which is essential for both parameter estimation and experimental design.

Building on this capability, Fisher Information Matrices (FIMs) were computed to quantify parameter identifiability. These were used within a framework for global optimal experimental design, where input profiles were optimized to maximize information content with respect to the parameters of interest.

4.4.4 Results

The developed PI-DeepONet models demonstrated high accuracy in approximating the SPM across a wide range of operating conditions and parameter values, cf., e.g., **¡Error! No se encuentra el origen de la referencia.¡Error! No se encuentra el origen de la referencia.**. Voltage prediction errors were found to be in the order of millivolts, indicating that the surrogate model is suitable for use in SOH estimation tasks.

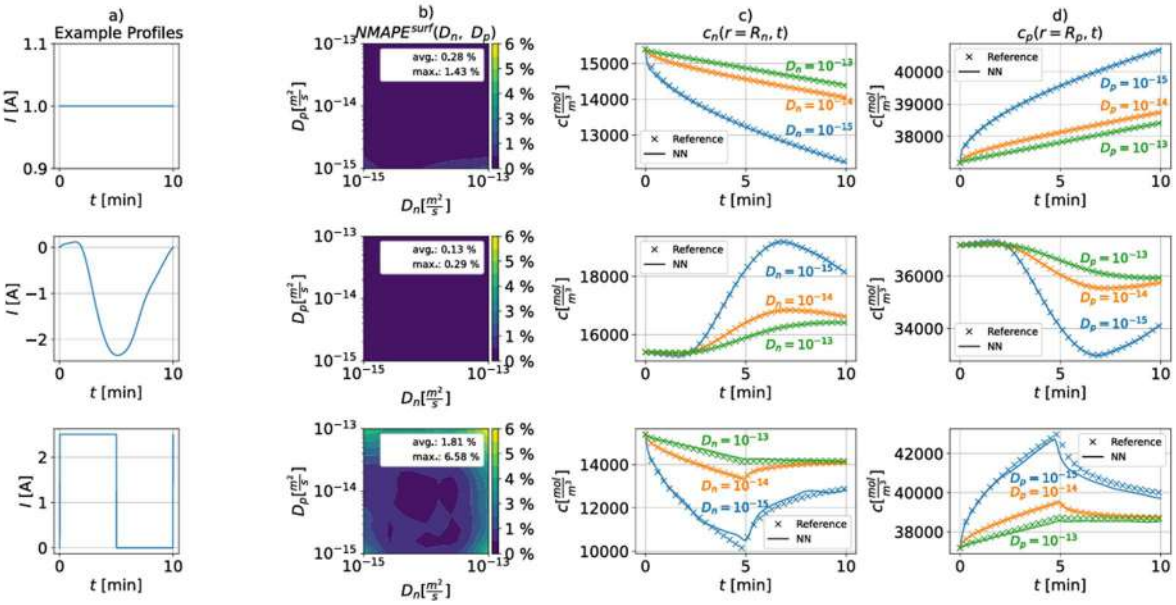

Figure 7: Accuracy of PI-DeepONet for exemplary current profiles and varying diffusivity parameters. Image taken from [1].

A key outcome of the work is the significant reduction in computational effort. Compared to conventional numerical solvers, speed-up factors of several orders of magnitude were achieved, enabling:

- rapid evaluation of large parameter spaces,
- efficient computation of sensitivities and FIMs, and
- practical implementation of iterative estimation and DoE workflows.

The proposed parameter estimation framework, combining:

1. global experimental design,
2. parameter estimation via population-based optimization, and
3. local identifiability analysis,

was shown to recover degradation-related parameters with good accuracy in a model-based setting. At the same time, the framework provided insights into parameter correlations and limitations in identifiability under different operating conditions.

Overall, the results demonstrate that the developed approach provides a unified methodology for:

- fast physics-based modelling,
- electrode-specific SOH estimation via eSOH parameters, and
- systematic design of informative experiments for improved SOH estimation.

This dual-use capability is particularly relevant for future battery management systems, where both accurate state estimation and efficient data evaluation strategies are required. While the present work was conducted as a proof-of-concept based on literature parameters [3], the methodology is directly transferable to application-specific systems given appropriate parameterization and validation.

4.5 US SoH cell/module/pack

This section extends the RUL model-development work of section 4.3 to cover three system hierarchy levels explicitly: individual Gen-3b LFP cell, 16s1p module, and a two-module parallel pack (16s2p). The available continuous cycling measurements are at cell level; module- and pack-level SOH trajectories are derived by applying the scaling relationships of the supplied hardware specifications. The same set of supervised regressors is trained and evaluated at each level, with results reported in both normalised SOH units and the corresponding absolute capacity (Ah) and energy (kWh) values to support physical interpretation at each hierarchy.

4.5.1 Hierarchy Scaling Relationships

Table 1 summarises how the key electrical quantities scale from cell level to module level and to pack level. Because SOH is defined as the ratio of measured discharge capacity to the nominal capacity at each level, and because the normalisation factor at each level is consistent with the topology (cells series/parallel), the SOH value is numerically identical at all three levels under the representative-cell assumption. The consequence is that a single SOH prediction model trained on cell data applies directly to module and pack levels without retraining. Only the physical interpretation of a given SOH prediction error changes: the corresponding absolute capacity shortfall scales with the nominal capacity of each level.

Table 1: Electrical quantity scaling from cell to 16s1p module to 16s2p pack.

Quantity	Cell (×1)	Module (16s1p)	Pack (16s2p, 2 modules)
Nominal capacity (Ah)	30	30	60
Nominal voltage (V)	≈3.25 (avg)	52.00	52.00
Gross energy (kWh)	≈0.098	1.752	3.504
Discharge current	I_cell	I_cell	2 × I_cell
SOH formula	Q_cell / 30 Ah	Q_module / 30 Ah	Q_pack / 60 Ah

SOH value (numerically)	SOH_cell	= SOH_cell	= SOH_cell
Capacity error for ΔSOH	ΔSOH × 30 Ah	ΔSOH × 30 Ah	ΔSOH × 60 Ah
EOL threshold	SOH < 0.80	SOH < 0.80	SOH < 0.80
EOL capacity (Ah)	24	24	48

4.5.2 Dataset and Preprocessing

The preprocessing discovered 30 candidate cell folders and produced 3380 valid cycle-level records from 4 usable cell datasets provided by ABEE. These cover three operating conditions: 25 °C full-swing CC-CV (Cells 1 and 2 from Run 1), 15 °C full-swing CC-CV (Cell 1 from Run 13), and 45 °C full-swing CC-CV (Cell 1 from Run 2). The usable cell datasets and their SOH ranges are listed in Table 1.

Table 2: Processed cell datasets with SOH ranges at cell, module, and pack level.

Cell dataset	Cycles	Cycle range	Min SOH (cell)	Min SOH (module)	Min SOH (pack)
Run 1 - 25 Celsius CC-CV 0-100 SoC / Cell 1	1000	1–1000	0.925	0.925	0.925
Run 1 - 25 Celsius CC-CV 0-100 SoC / Cell 2	940	1–940	0.197	0.197	0.197
Run 13 - 15 Celsius CC-CV 0-100 SoC / Cell 1	940	1–940	0.196	0.196	0.196
Run 2 - 45 Celsius CC-CV 0-100 SoC / Cell 1	500	1–500	0.982	0.982	0.982

Cell SOH was computed as the CC-discharge capacity divided by the nominal cell capacity of 30 Ah, as shown in Figure 8.

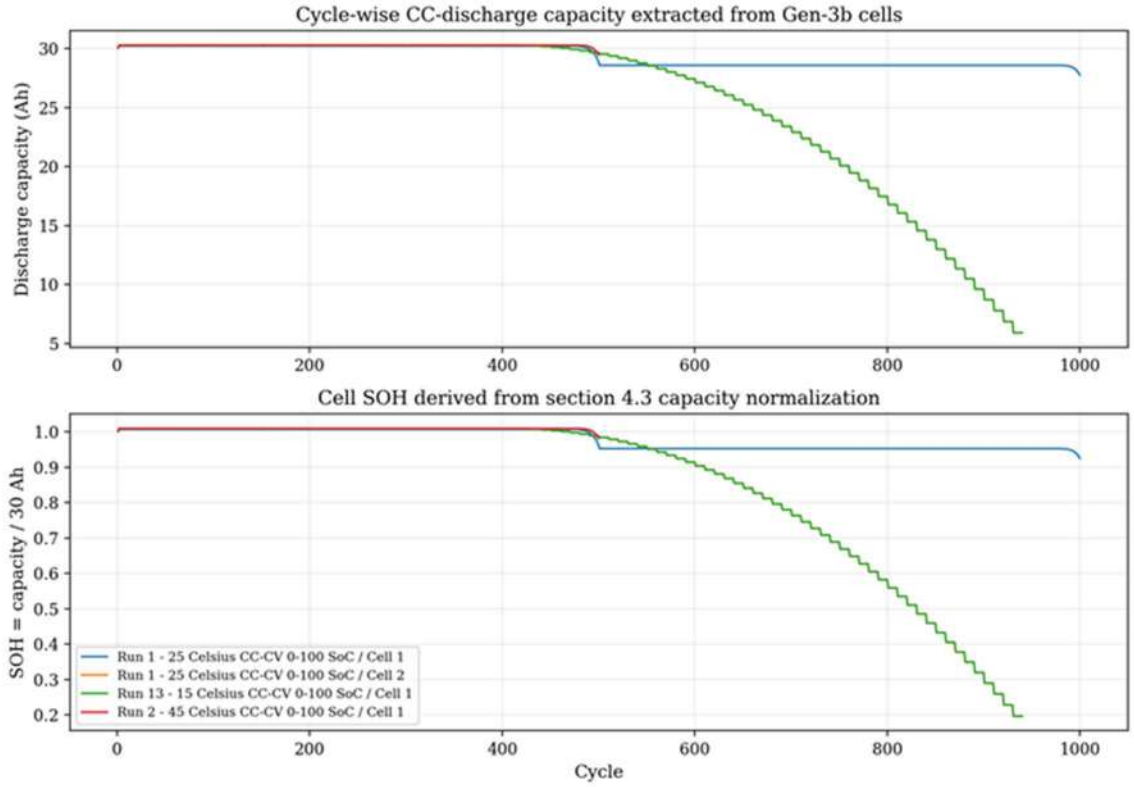

Figure 8: Cell-level CC-discharge capacity and SOH trajectories extracted from the Gen-3b cycling data.

Module SOH was derived under the 16s1p assumption: in a series string every cell carries the same discharge current, so the module discharge capacity (Ah) equals the cell discharge capacity (Ah) and the normalised module SOH equals the cell SOH. This can be seen in Figure 9.

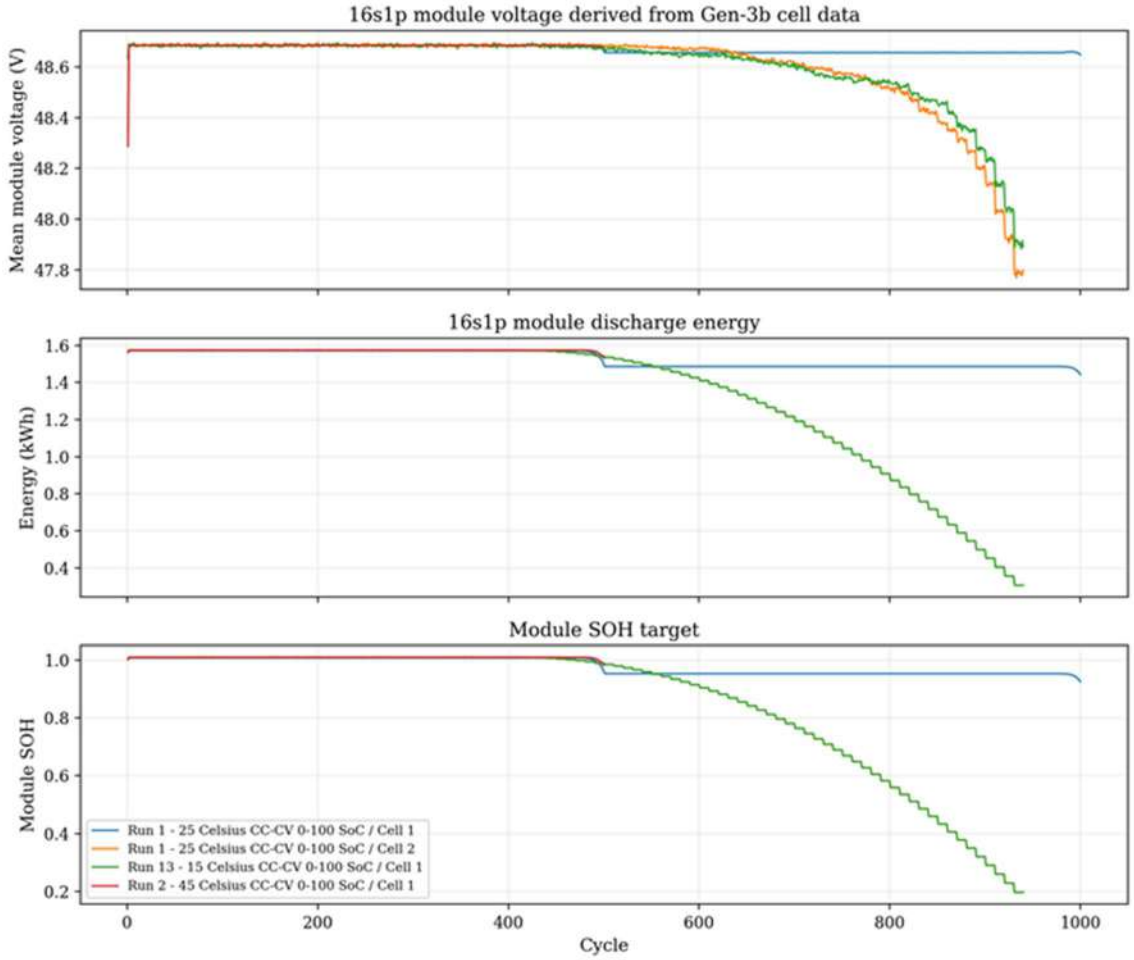

Figure 9: Module-level (16s1p) voltage, energy, and SOH trajectories derived from the cell measurements.

Pack SOH was computed from the two-module parallel configuration: two 16s1p modules in parallel double the current and capacity while keeping the voltage at 52 V, giving a nominal pack capacity of 60 Ah; dividing the measured pack capacity by this nominal value again returns the same normalised SOH which is visualised in Figure 10.

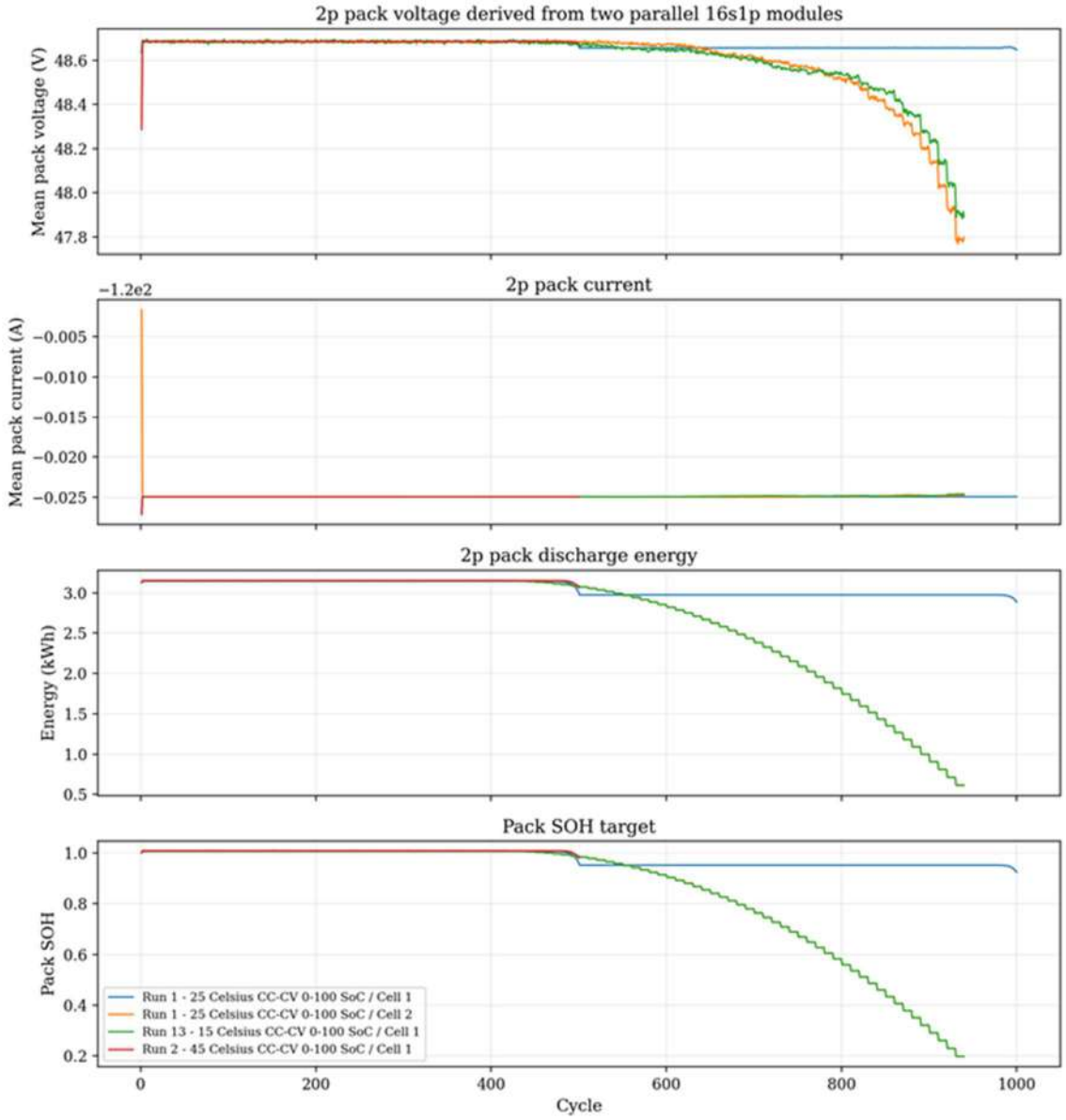

Figure 10: Pack-level (two 16s1p modules in parallel, 16s2p equivalent) voltage, current, energy, and SOH trajectories.

4.5.3 Model Setup

Six supervised regressors were evaluated: XGBoost, random forest, gradient boosting, support vector regression (SVR), LSTM, and one-dimensional CNN. Because SOH is numerically identical at cell, module, and pack level under the representative-cell assumption (Table 3), a single model trained on cell-derived cycle features predicts SOH at all three levels simultaneously. The absolute capacity error at each level is recovered by multiplying the SOH prediction error by the relevant nominal capacity (30 Ah for cell and module, 60 Ah for pack).

The tabular models (XGBoost, random forest, gradient boosting, SVR) consume the four-scalar cycle-level features directly. The LSTM and 1D CNN use a sliding window of 32 consecutive cycles so that each prediction conditions on a short degradation history. Both sequence models were trained with a hidden size of 128

(LSTM: 2 layers, dropout 0.1; CNN: strided convolutions followed by global pooling), a learning rate of 0.001, and early stopping with patience 20.

Two evaluation protocols were used. The primary protocol is a chronological 70/30 split that reserves the latest degradation cycles for testing; this is the only setting representative of real-world RUL deployment, where the model must extrapolate from early-life data into late-life behaviour (Figure 11). A random 80/20 split is also reported as a section 4.3-style baseline to isolate the magnitude of the interpolation-leakage effect.

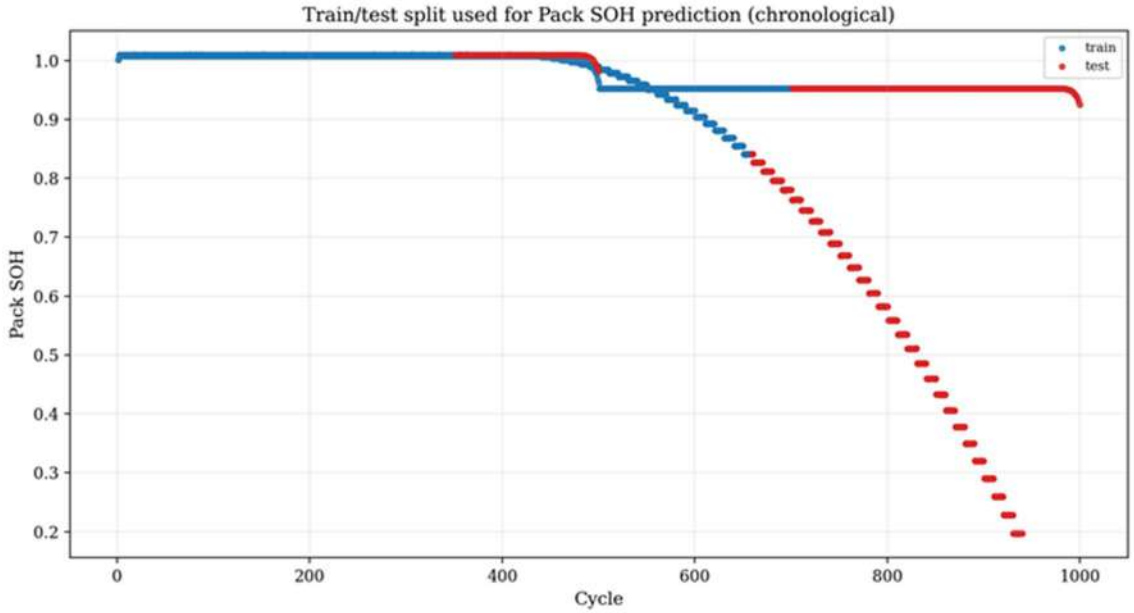

Figure 11: Chronological 70/30 train-test split used for the primary SOH/RUL evaluation across all three levels.

4.5.4 Cell-levle SOH and RUL Prediction Results

At cell level the prediction target is SOH_cell = Q_cell / 30 Ah. Under the chronological split the best cell SOH estimator was LSTM, achieving RMSE = 0.1606 SOH (≈ 4.82 Ah), MAE = 0.0911 SOH (≈ 2.73 Ah), MAPE = 26.81 %, and R² = 0.596. Full per-model results for the chronological split are given in Table 3 and the random-split baseline in

Table 4.

Table 3: Cell-level SOH prediction metrics – chronological split (30 Ah cell basis).

Model	RMSE (SOH)	RMSE (30 Ah basis)	MAE (SOH)	MAPE (%)	R²
LSTM	0.1606	4.82 Ah	0.0911	26.81	0.596
Gradient boosting	0.2584	7.75 Ah	0.1611	44.16	-0.045
Random forest	0.2585	7.75 Ah	0.1611	44.18	-0.046
XGBoost	0.2605	7.82 Ah	0.1646	44.73	-0.063
SVR	0.3331	9.99 Ah	0.2194	57.27	-0.738
1D CNN	1.1065	33.19 Ah	0.6827	191.81	-18.171

Table 4: Cell-level SOH prediction metrics – random split baseline (30 Ah cell basis).

Model	RMSE (SOH)	RMSE (30 Ah basis)	MAE (SOH)	MAPE (%)	R^2
Gradient boosting	0.0002	0.01 Ah	0.0001	0.02	1.000
Random forest	0.0005	0.01 Ah	0.0002	0.03	1.000
XGBoost	0.0022	0.06 Ah	0.0007	0.10	1.000
SVR	0.0149	0.45 Ah	0.0017	0.19	0.992
LSTM	0.0199	0.60 Ah	0.0188	2.30	0.986
1D CNN	0.0274	0.82 Ah	0.0221	2.67	0.974

Figure 12 compares all six models visually on the chronological split. The chronological errors are substantially larger than the random-split values because the model must extrapolate into the steep late-life capacity fade not observed during training.

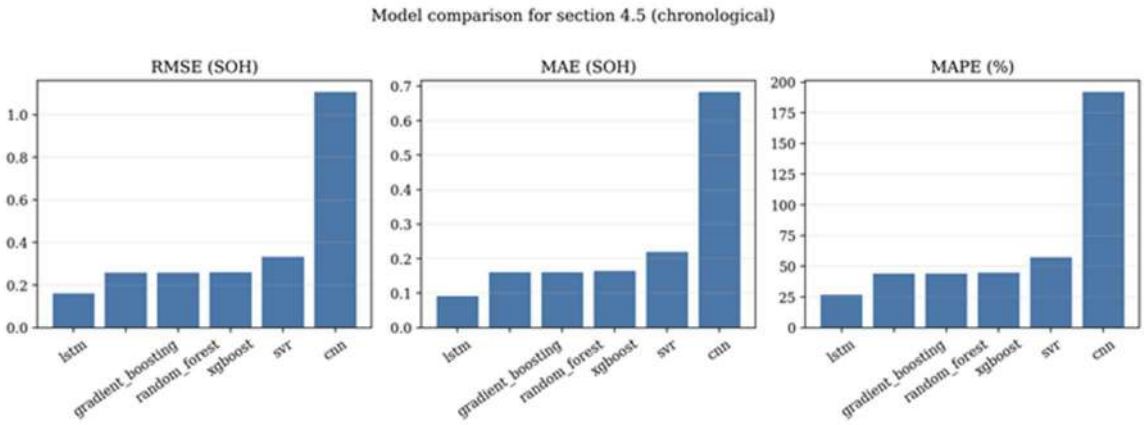

Figure 12: All-model SOH prediction error comparison under the chronological split (RMSE, MAE, MAPE). LSTM achieves the lowest chronological RMSE; 1D CNN diverges under chronological extrapolation.

Under the random split the best cell SOH estimator was Gradient boosting with RMSE = 0.00024 SOH (≈ 0.0072 Ah) and MAPE = 0.016 % (Table 4.5.6). These near-perfect figures arise because early, mid, and late cycles are mixed across train and test sets; they are retained as a sanity check but are not indicative of deployable performance.

Figure 13 shows the predicted (dashed) versus reference (solid) cell SOH trajectories for all six models under the chronological split. Faded lines represent the training portion; solid-coloured lines show the held-out test segment. LSTM and SVR track the late-life trajectory most closely; tree-based models exhibit a characteristic plateau artefact beyond the training distribution.

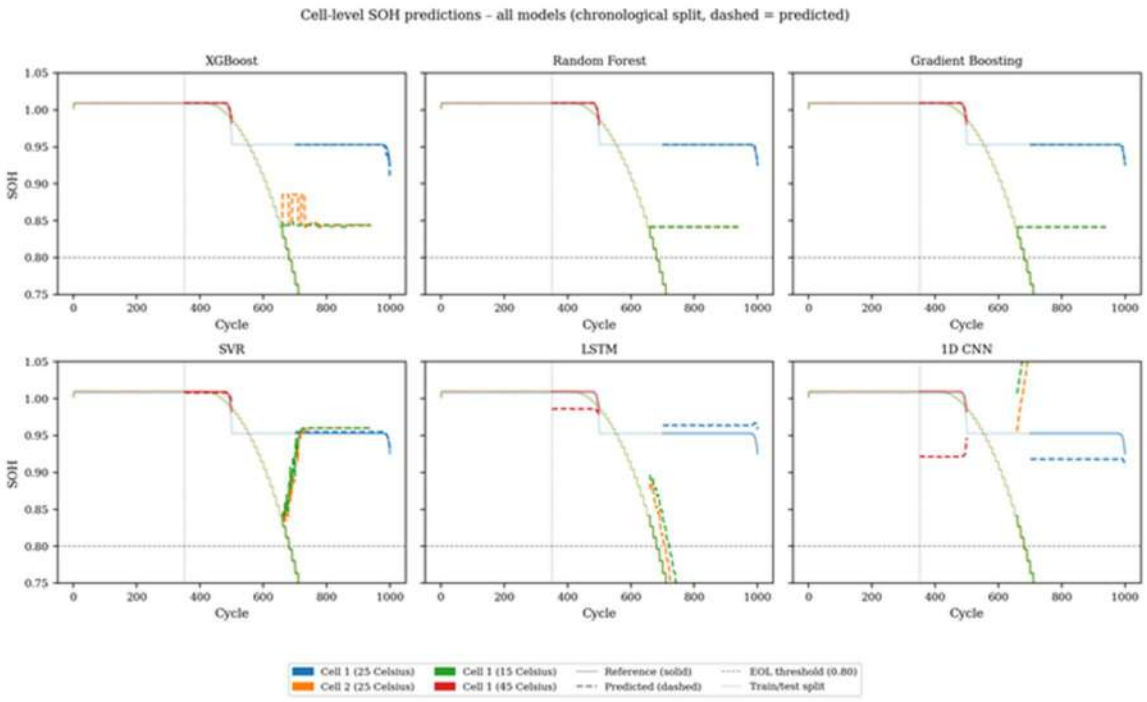

Figure 13: Cell-level SOH trajectories – all six models (chronological split). Solid lines: reference SOH (faded = train, opaque = test). Dashed lines: model predictions on the test segment. Dotted vertical line: train/test boundary.

Figure 14 presents the predicted-versus-reference SOH scatter for all models on the chronological test set. Points below the identity line indicate over-prediction of SOH (under-estimation of degradation), which is most pronounced in the sub-0.85 SOH region for tree-based models that cannot extrapolate outside the training value range.

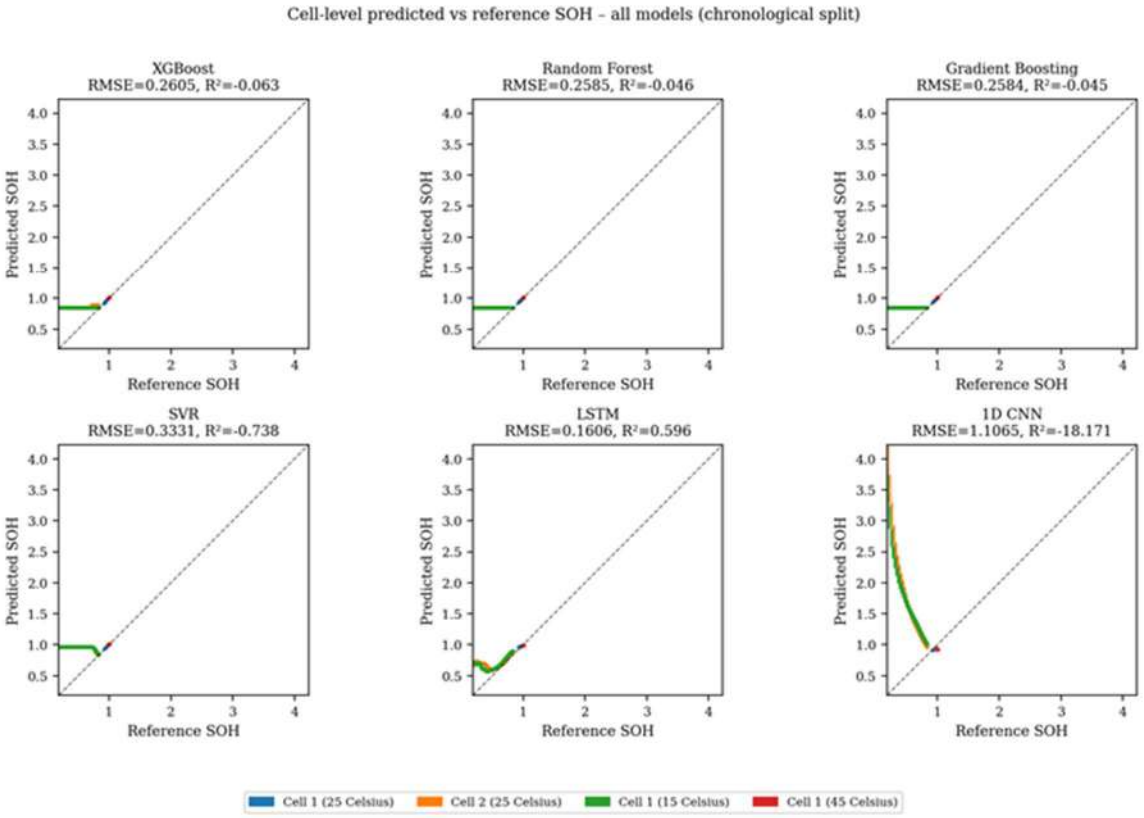

Figure 14: Cell-level predicted vs reference SOH scatter – all six models (chronological split). Dashed line: perfect prediction (identity). RMSE and R² annotations per panel.

Figure 15 shows the SOH prediction residual (predicted − reference) by cycle for all six models. For models that generalise poorly to late-life cycles, the residual grows sharply after the capacity knee at approximately cycle 700 for the two cells that reached EOL within the measurement window.

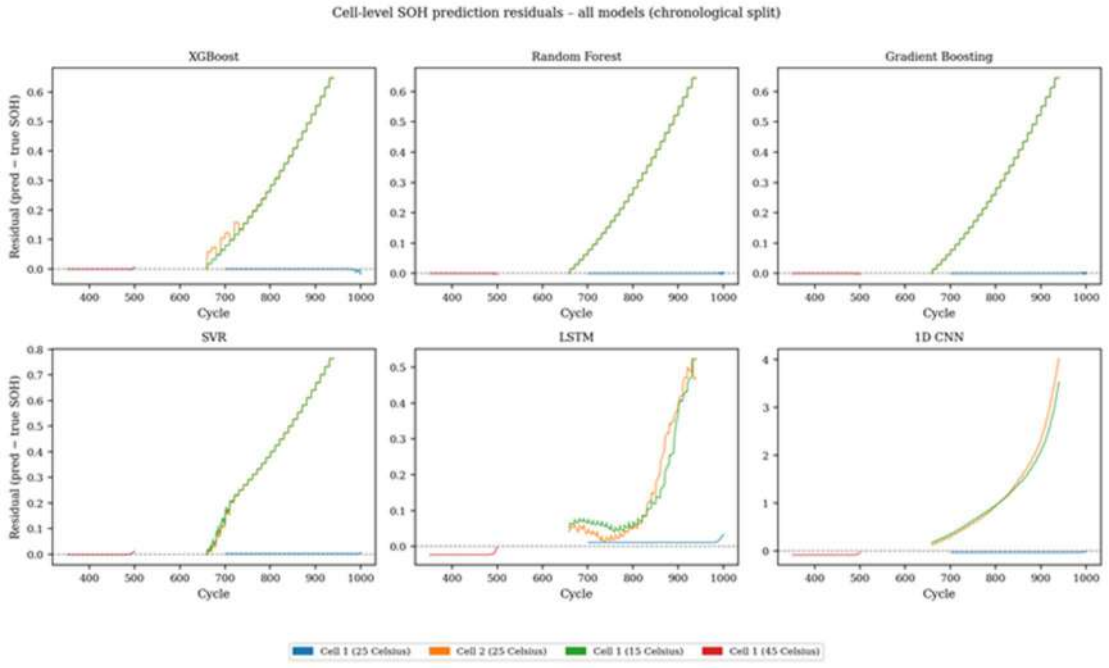

Figure 15: Cell-level SOH prediction residuals by cycle – all six models (chronological split). Each colour corresponds to one cell. Dashed horizontal line: zero residual.

4.5.5 Module-Level SOH and RUL Prediction

Module SOH is defined as SOH_module = Q_module / 30 Ah. In the 16s1p configuration all cells share the same discharge current, so the module capacity (Ah) equals the individual cell capacity (Ah) and SOH_module = SOH_cell numerically. The SOH prediction model therefore applies directly to module-level estimation without modification, hence the results in Tables 4.5.5 and 4.5.6 apply over here.

What changes at module level is the physical interpretation of the prediction error (see Figure 4.5.4, top-right panel, for the module energy degradation trajectory). The LSTM chronological RMSE of 0.1606 SOH corresponds to 4.82 Ah at the module terminals. At the module nominal voltage of 52 V this represents an energy error of 0.250 kWh, out of a module gross capacity of 1.752 kWh. The mean absolute EOL-cycle error at module level equals the cell-level figure (34.0 cycles) because the cycle counter is common to both.

Figure 16 translates the SOH predictions of all six models into module discharge capacity (Ah) for the chronological test segment. The y-axis is expressed on the 30 Ah module nominal basis, with the EOL line at 24 Ah. The relative ranking of models is unchanged from the cell-level plot; the difference between models becomes physically meaningful in terms of Ah error at the module terminals.

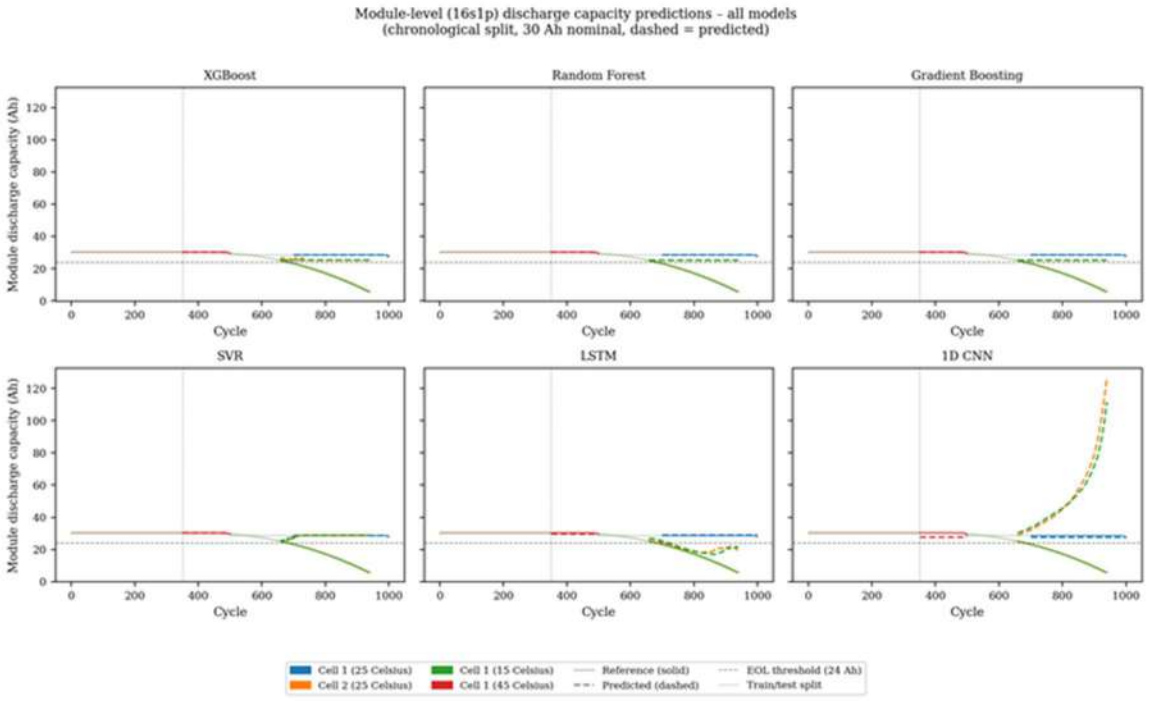

Figure 16: Module-level (16s1p) discharge capacity trajectories – all six models (chronological split, 30 Ah nominal). Solid lines: measured module capacity (faded = train, opaque = test). Dashed lines: predicted capacity. Horizontal dashed line: EOL at 24 Ah.

Figure 17 compares the prediction errors of all six models in Ah terms at the module level (30 Ah basis). The blue and green bars show chronological RMSE and MAE respectively; the red bars show the random-split RMSE for reference. The near-zero random-split bars confirm that the interpolation-leakage effect is

present at module level as well as cell level, and that chronological performance is the only figure of merit relevant to deployment.

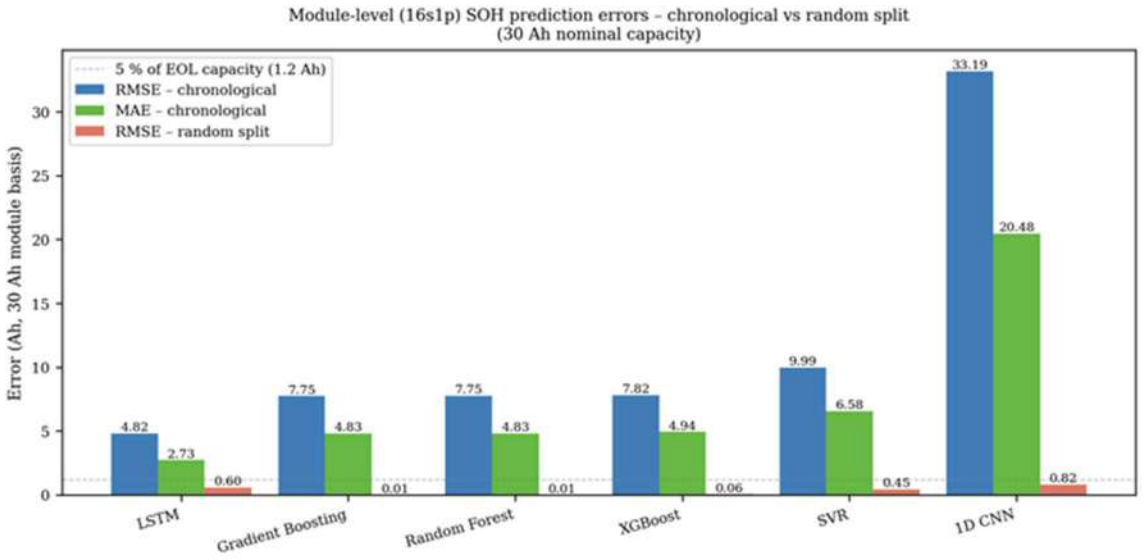

Figure 17: Module-level prediction error comparison – all six models (30 Ah module basis). Blue: chronological RMSE (Ah). Green: chronological MAE (Ah). Red: random-split RMSE (Ah).

Figure 9 (shown in the preprocessing sub-section) displays the derived module voltage, energy fraction, and SOH trajectories for the four available cells. The module voltage trajectory, rising from the 40 V cut-off to the 58.4 V maximum, illustrates the 16× cell-voltage scaling. The SOH trajectory is identical to the cell-level trajectory because the module capacity equals the cell capacity in the 16s1p topology.

4.5.6 Pack-Level SOH and RUL Prediction

Pack SOH is defined as SOH_pack = Q_pack / 60 Ah, where the two-module parallel pack has a nominal capacity of 60 Ah at 52 V (3.504 kWh gross), as shown in Figure 10. Under the representative-module assumption, both modules age identically, so pack SOH is numerically equal to module and cell SOH (see Figure 11, bottom panels, for the physical degradation trajectories at pack level). The SOH prediction model applies directly; the same normalised RMSE and R^2 values hold as at cell and module level which is seen in Table 5 and Table 6).

Table 5: Pack-level SOH prediction metrics – chronological split (60 Ah pack basis, 52 V).

Model	RMSE (SOH)	RMSE (60 Ah basis)	MAE (SOH)	MAPE (%)	R^2
LSTM	0.1606	9.63 Ah	0.0911	26.81	0.596
Gradient boosting	0.2584	15.50 Ah	0.1611	44.16	-0.045
Random forest	0.2585	15.51 Ah	0.1611	44.18	-0.046
XGBoost	0.2605	15.63 Ah	0.1646	44.73	-0.063
SVR	0.3331	19.99 Ah	0.2194	57.27	-0.738
1D CNN	1.1065	66.39 Ah	0.6827	191.81	-18.171

Table 6: Pack-level SOH prediction metrics – random split baseline (60 Ah pack basis).

Model	RMSE (SOH)	RMSE (60 Ah basis)	MAE (SOH)	MAPE (%)	R^2
Gradient boosting	0.0002	0.01 Ah	0.0001	0.02	1.000
Random forest	0.0005	0.03 Ah	0.0002	0.03	1.000
XGBoost	0.0022	0.13 Ah	0.0007	0.10	1.000
SVR	0.0149	0.89 Ah	0.0017	0.19	0.992
LSTM	0.0199	1.19 Ah	0.0188	2.30	0.986
1D CNN	0.0274	1.64 Ah	0.0221	2.67	0.974

At pack level the absolute capacity error doubles relative to cell and module level because the nominal capacity is 60 Ah instead of 30 Ah. The LSTM chronological RMSE of 0.1606 SOH corresponds to 9.63 Ah at the pack terminals, or 0.501 kWh in energy terms (Table 5). Under the random split the best estimator (Gradient boosting) achieves RMSE = 0.0143 Ah (0.00024 SOH, Table 6), confirming that near-zero interpolation errors do not translate to acceptable energy errors at system level.

Figure 18 shows the predicted (dashed) versus measured (solid) pack discharge capacity for all six models under the chronological split. The y-axis is expressed on the 60 Ah pack nominal basis, with the EOL line at 48 Ah. The model ranking mirrors the cell and module levels; the doubled Ah scale makes the absolute spread between the best and worst models more readily apparent.

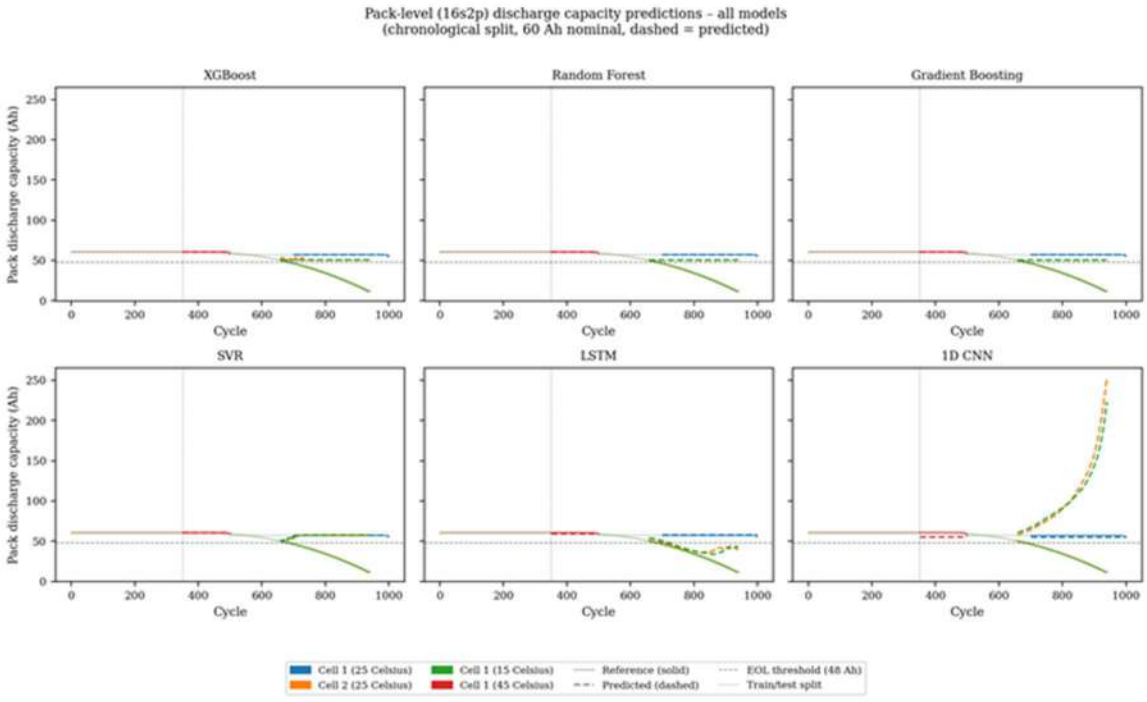

Figure 18: Pack-level (16s2p) discharge capacity trajectories – all six models (chronological split, 60 Ah nominal). Solid lines: measured pack capacity (faded = train, opaque = test). Dashed lines: predicted capacity. Horizontal dashed line: EOL at 48 Ah.

Figure 19 compares the prediction errors of all six models in Ah terms at pack level (60 Ah basis). The chronological RMSE and MAE values are exactly double the module-level figures because the nominal pack capacity is twice the module capacity. The random-split RMSE bars remain near zero, illustrating that the interpolation-leakage advantage does not diminish at pack scale.

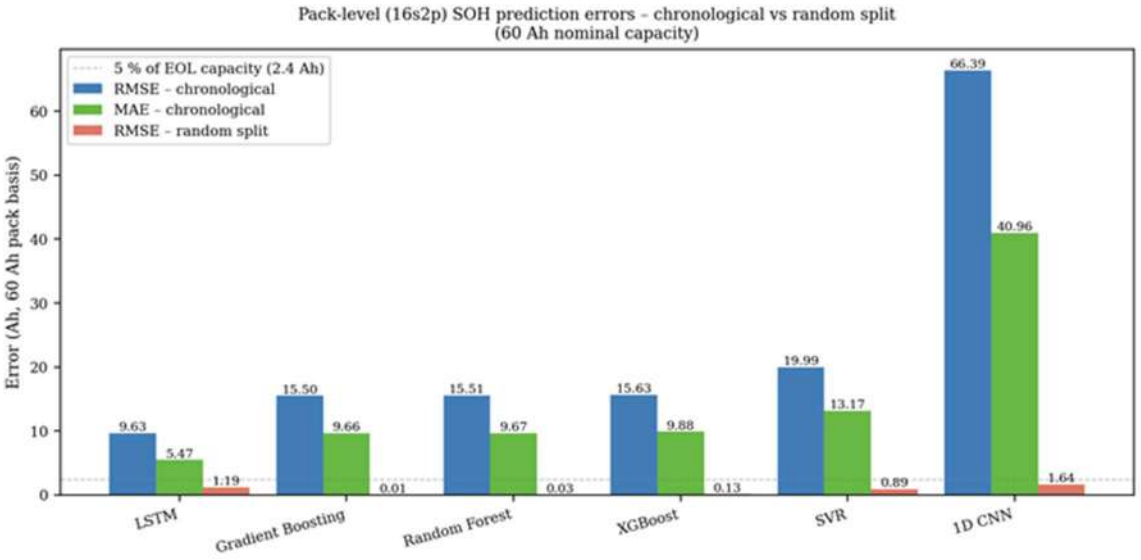

Figure 19: Pack-level prediction error comparison – all six models (60 Ah pack basis). Blue: chronological RMSE (Ah). Green: chronological MAE (Ah). Red: random-split RMSE (Ah).

Pack RUL is estimated from the predicted pack SOH trajectory using the same 0.80 EOL threshold applied at cell and module level. The EOL cycle number is numerically identical at all three hierarchy levels because the cycle counter and SOH value are shared. At pack level, the EOL state corresponds to a remaining usable capacity of 0.80 × 60 = 48 Ah (2.803 kWh). The mean absolute EOL-cycle error from the chronological LSTM is 34.0 cycles, equal to the cell- and module-level figures; the per-cell RUL trajectories inferred from the LSTM SOH predictions are shown in Figure 19.

4.5.7 Cross-level Prediction Summary

Table 7 brings together the key prediction metrics for the LSTM under the chronological split across all three hierarchy levels. The normalised SOH metrics (RMSE, MAE, R^2) are identical at every level because SOH is a dimensionless ratio whose value is the same at cell, module, and pack scale. The absolute capacity and energy errors grow with the nominal capacity of each level. Figure 11: Chronological 70/30 train-test split used for the primary SOH/RUL evaluation across all three levels. shows how the same underlying degradation curves translate into different physical capacity and energy values at each level.

Table 7: Cross-level SOH/RUL prediction summary for the LSTM – chronological split.

Level	Nominal capacity	SOH RMSE	Capacity RMSE	Energy RMSE	SOH MAE	R^2	MAPE (%)	Mean |EOL error| (cycles)

Cell	30 Ah, ~3.25 V avg	0.1606	4.82 Ah	0.0157 kWh	0.0911	0.596	26.81	34.0
Module (16s1p)	30 Ah, 52 V	0.1606	4.82 Ah	0.250 kWh	0.0911	0.596	26.81	34.0
Pack (16s2p)	60 Ah, 52 V	0.1606	9.63 Ah	0.501 kWh	0.0911	0.596	26.81	34.0

4.5.8 Pack Aggregation Assumptions

The results above are derived under the representative-trajectory assumption: both modules in the parallel pack are assumed to age identically. Under this assumption pack SOH, module SOH, and cell SOH are numerically equal cycle-by-cycle, and the single SOH prediction model covers all three levels without modification. Pack capacity, current, and energy are obtained by the topology scaling in Table 1.

In a physical pack with independent current paths for each module, module ageing will diverge over time due to manufacturing spread and local temperature gradients. Pack capacity will then be limited by the weaker module rather than the average, and the representative-cell assumption will overestimate usable pack capacity. A physically accurate pack-level model in that case requires either separate SOH estimates for each module or a dedicated pack-level measurement campaign. The present analysis cannot capture module-divergence effects because only one physical cell trajectory is available per operating condition.

The codebase includes cell-to-module and module-to-pack conversion functions that propagate all measured quantities (voltage, current, capacity, energy) according to the series/parallel topology. Additional pack topologies can be evaluated by adjusting the module_spec and pack_spec entries in the experiment configuration file.

4.5.9 Conclusion

The multi-level SOH/RUL analysis demonstrates that a single data-driven model trained on cell-level cycle features covers the cell, 16s1p module, and 16s2p pack levels simultaneously, because the normalised SOH metric is numerically identical across all three levels under the representative-cell assumption. Table 7 summarises the LSTM chronological metrics at all three levels. The absolute physical errors scale with the nominal capacity of each level: the LSTM chronological RMSE of 0.1606 SOH corresponds to ≈ 4.8 Ah at cell and module level and ≈ 9.6 Ah (0.50 kWh) at pack level.

The chronological evaluation is the only benchmark relevant to deployment: it reveals that late-life extrapolation is the key challenge, with the LSTM achieving the best chronological performance (R^2 = 0.596) while tree-based models excel only in the random-split interpolation baseline. The mean absolute EOL-cycle error from the chronological LSTM is 34.0 cycles at all three hierarchy levels. These results demonstrate the feasibility of data-driven SoH and RUL modelling, while

also indicating that robust deployment will require larger datasets, strict chronological validation, and direct module- and pack-level measurements.

The main limitations of this analysis are: (1) only four usable cell trajectories across three operating conditions are available, limiting generalisation; (2) the representative-trajectory assumption for the two parallel modules ignores module-divergence effects that will affect pack-level performance in practice; and (3) no direct module- or pack-level cycling measurements exist to validate the scaling relationships experimentally. Additional multi-cell datasets covering a wider temperature and SoC-window range, combined with physical module measurements, would be required before claiming robust deployed pack-level RUL performance.

5. CONCLUSION

This deliverable has presented the development and validation of data-driven and physics-assisted battery models for virtualization of performance and ageing testing within the FASTEST project. Multiple modelling approaches were implemented, including LSTM-based SoC estimation, machine-learning-based SOH and RUL prediction, and physics-informed deep operator networks for electrode-specific SOH estimation.

The results demonstrate that data-driven models can achieve high accuracy across different chemistries, operating conditions, and ageing stages while offering significant reductions in computational cost compared to conventional numerical simulations. The integration of physics-based knowledge improves model robustness, interpretability, and suitability for extrapolation beyond the training domain.

All developed models were prepared for integration into the FASTEST Digital Twin and hybrid testing platform, including export as FMUs and compatibility with co-simulation and HiL environments. The work therefore represents a concrete step towards test virtualization at TRL 6, contributing directly to reduced testing time, lower development cost, and accelerated battery innovation.

Remaining challenges include the availability of broader multi-cell datasets, validation under more diverse real-world conditions, and extension towards pack-level deployment with cell-to-cell variability. These aspects will be addressed in subsequent FASTEST activities and follow-up research. Overall, D3.3 confirms the viability and value of data-driven modelling as a core pillar of next-generation battery testing and development workflows.

6. REFERENCES

[1] P. Brendel, I. Mele, A. Rosskopf, T. Katrašnik und V. Lorentz, „Parametrized physics-informed deep operator networks for Design of Experiments applied to Lithium-Ion-Battery cells.," *Journal of energy storage, 128, 117055,* 2025.

[2] P. Brendel, C. Straub, A. Rosskopf, V. Lorentz und F. Dietrich, „Physics-Informed Operator Learning for Parameter Estimation in Lithium-Ion-Battery Models Enhanced by Global Experimental Design and Local Identifiability Analysis. Preprint. Available at SSRN 6156315," 2026.

[3] C. H. Chen, F. Brosa Planella, K. O'Regan, D. Gastol, W. D. Widanage und E. Kendrick, „Development of experimental techniques for parameterization of multi-scale lithium-ion battery models," *Journal of The Electrochemical Society, 167(8), 080534,* 2020.

[4] I. Lopetegi, G. L. Plett, M. S. Trimboli, L. Oca, E. Miguel und U. Iraola, „A new battery soc/soh/esoh estimation method using a pbm and interconnected spkfs: Part ii. soh and esoh estimation," *Journal of The Electrochemical Society, 171(3), 030518,* 2024.