EUROPEAN COMMISSION

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

FASTEST

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

Deliverable D4.4: Battery AI-powered toolchain validation and verification strategies

Primary Author [Vijay Nagulapati]

Organization [RSTER]

Date: [29.05.2026]

Doc.Version: [1.0]

Co-funded by the European Union

UK Research and Innovation

Co-funded by the European Union and UKRI under grant agreements N° 101103755 and 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:	4
Deliverable:	Battery AI-powered toolchain validation and verification strategies
Deliverable Type:	Report
Dissemination Level:	Public
Due Date:	30.05.2026 (Month 36)
Actual Submission Date:	29.05.2026
Pages:	< 52 >
Doc. Version:	Draft 0.2
GA Number:	101103755
Project Coordinator:	Bruno Rodrigues | ABEE (bruno.rodrigues@avestaholding.com)

Formal Reviewers		
Name	Organization	Date
Igor Mele	UL	19.05.2026
Nuno Marques	INEGI	25.05.2026

Document History			
Version	Date	Description	Author
0.1	05.05.2026	Document creation and development	Vijay Nagulapati (RSTER)
0.2	11.05.2026	Chapter 4 & 5	Bart Aerts (FM)
0.2	13.05.2026	Chapter 2 & 4	Bruno & Sina (AVESTA)
0.2	15.05.2026	Draft for review	Vijay Nagulapati (RSTER)

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 to 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 the platform and its components up to TRL 6: 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 datadriven models able to substitute physical characterisation experiments; and the overarching Digital Twin architecture managing the information flows, and the TRL 6 proven and integrated prototype of the hybrid testing platform.

LIST OF ABBREVIATIONS, ACRONYMS AND DEFINITIONS

Acronym	Name
°C	Degree Celsius
A	Ampere
ABEE	Avesta Battery & Energy Engineering
Ah	Ampere-hour
AI	Artificial Intelligence
ASOC	Available State of Charge
BMS	Battery Management System
C / C-rate	Charge or discharge rate relative to the nominal battery capacity
CC	Constant Current
CINEA	European Climate, Infrastructure and Environment Executive Agency
CV	Constant Voltage
DoE	Design of Experiments
DT	Digital Twin
DUT	Device Under Test
EODV	End-of-Discharge Voltage
EU	European Union
FASTEST	Fast-track hybrid testing platform for the development of battery systems
FM	Flanders Make
GA	Grant Agreement
Gen3	Generation 3 battery cell/module technology
Gen3b	Generation 3b battery cell/module technology
Gen4	Generation 4 battery cell/module technology
Imax	Maximum current specified by the supplier/manufacturer
IR	Internal Resistance
ISO	International Organisational for Standardisation
KPI	Key Performance Indicator
LFP	Lithium Iron Phosphate
Li-ion	Lithium-ion
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error

MLP	Multilayer Perceptron
MSE	Mean Squared Error
NMC	Nickel Manganese Cobalt
OCV	Open-Circuit Voltage
REESS	Rechargeable Electrical Energy Storage System
RMSE	Root Mean Square Error
RSTER	Research Centre for Sustainable Energy Technologies
SOC / SoC	State of Charge
Tmin	Minimum ambient test temperature specified by the supplier/manufacturer
TRL	Technology Readiness Level
V	Volt
VTT	Technical Research Centre of Finland
W	Watt
Wh	Watt-hour
XCC	Extreme Cold Condition

LIST OF TABLES

LIST OF FIGURES

Table of Contents

EXECUTIVE SUMMARY

Deliverable D4.4 presents the validation and verification strategy developed within Task 4.4 of the FASTEST project for the Battery AI-powered multi-concern toolchain. The deliverable establishes a Digital Twin (DT)-compliant validation methodology for assessing battery-system safety, ageing and reliability under representative operational and degradation scenarios. The work builds upon the baseline toolchain architecture and degradation scenarios defined in D4.1 and integrates experimental validation activities, AI-assisted estimation workflows and KPI-based assessment within a hybrid-testing framework.

The primary objective of D4.4 is to define a structured and traceable validation approach capable of linking experimental evidence, degradation scenarios and toolchain outputs at cell, module and system levels. The methodology combines quantitative Key Performance Indicators (KPIs), such as RMSE, MAE and threshold-detection metrics, with engineering-oriented interpretation of abnormal operating conditions and degradation-related behaviour. Validation outcomes are classified as validated, partially validated, not validated or not assessable in order to ensure transparent interpretation of the achieved validation maturity.

The strongest validation maturity was achieved for electrical safety-related scenarios. Experimental activities performed by FM demonstrated correct Battery Management System (BMS) response during overcharge and current-overload conditions for both Gen3 and Gen4 battery modules. The observed relay-opening behaviour confirmed that the protection logic disengaged the battery modules before unsafe voltage thresholds were exceeded. The corresponding toolchain outputs demonstrated consistent interpretation of safety-boundary behaviour and abnormal electrical operating conditions.

The deliverable also integrates AI-assisted estimators developed by ABEE for voltage prediction, temperature estimation, capacity degradation assessment and internal-resistance evolution. The available results demonstrate that the implemented estimators are capable of reproducing relevant operational and degradation trends under representative stress and abuse scenarios, thereby supporting the feasibility of integrating AI-assisted workflows within the FASTEST validation framework.

Several planned module-level validation activities related to thermal performance evaluation, fast charging/discharging validation and extended electro-thermal data-acquisition campaigns could not be fully completed within the project timeframe due to subsystem interactions identified during overdischarge testing. Specifically, interactions associated with the auxiliary cooling/heating subsystem affected the electrical boundary conditions of the battery module after relay disengagement, preventing stable completion of several planned tests under controlled conditions.

These limitations are transparently documented within the deliverable and are explicitly reflected in the adopted validation-status methodology. Consequently, scenarios affected by incomplete experimental evidence are classified as partially validated or not assessable. This approach preserves technical credibility while remaining consistent with the FASTEST hybrid-testing philosophy.

Overall, D4.4 demonstrates that the FASTEST validation framework provides a technically coherent methodology for integrating experimental evidence, AI-assisted estimation, KPI-based assessment and Digital Twin-compatible traceability within a unified battery-system validation environment. The deliverable therefore establishes a credible basis for continued refinement and future integration of the FASTEST multi-concern toolchain within hybrid-testing and Digital Twin-enabled validation workflows.

1. INTRODUCTION

The FASTEST project addresses the need to accelerate battery system development by combining physical tests, virtual testing methods, Design of Experiments, Digital Twin concepts and multi-domain modelling. Battery safety and reliability assessment remains resource-intensive because degradation and failure mechanisms do not occur in a single isolated domain. Electrical loading, thermal behaviour, ageing mechanisms and control or operational boundaries interact across the cell, module and system levels. For this reason, the validation of the FASTEST toolchain must consider the multi-concern nature of battery systems and must verify whether the toolchain can provide useful, reliable and traceable support for safety, ageing and reliability evaluation.

This deliverable represents the final validation-oriented task of WP4. It builds on the architecture and degradation scenario definition from T4.1, the high-fidelity multi-domain toolchain development from T4.2, and the system-level optimisation and integration work from T4.3. It also interfaces with the broader FASTEST Digital Twin and hybrid testing platform by ensuring that validated toolchain outputs can be interpreted, exchanged and reused within the project architecture.

1.1 Purpose and Scope of Deliverable

The purpose of D4.4 is to define and document the strategy for validating and verifying the FASTEST Battery AI-powered multi-concern toolchain. The deliverable explains how the toolchain will be assessed against the degradation scenarios and functional safety conditions relevant to battery systems. The scope includes both validation and verification. Verification addresses whether the toolchain structure, inputs, outputs, interfaces and calculation logic are implemented consistently with the intended design. Validation addresses whether the toolchain outputs are sufficiently representative of observed battery behaviour under experimental and operationally relevant conditions.

The validation scope covers safety, ageing and reliability from cell-to-system levels. Electrical aspects include voltage, current, overcharge, overdischarge, and abnormal current loading. Thermal aspects include heat generation, temperature gradients, low- and high-temperature operation, and thermal runaway or propagation risk indicators. The deliverable also considers coupled multi-physics scenarios where one failure mechanism may trigger or amplify another, for example an electrical fault leading to local heating and subsequent thermal propagation.

The scope of the deliverable is methodological and evidence-based. It defines the validation approach, the required partner inputs, the expected experimental evidence, the validation metrics and the final interpretation framework. It does not replace the detailed battery baseline documentation from ABEE or the experimental protocols from FM and VTT; rather, it integrates these inputs into a coherent validation methodology for the WP4 toolchain.

1.2 Relation to FASTEST Objectives and WP4

D4.4 contributes directly to the FASTEST objective of developing a fast-track hybrid testing platform for battery systems. The project aims to reduce dependency on extensive physical experimentation by combining robust testing results with physics-based models, data-driven approaches and Digital Twin data flows. The validation strategy described here ensures that this reduction in physical testing remains scientifically justified and technically traceable. A toolchain can only support accelerated testing if its predictions are validated against representative experimental evidence and if its limitations are clearly documented.

Within WP4, the toolchain is intended to support virtual testing, analysis, evaluation and validation of battery safety and reliability based on degradation models, failure scenarios and battery lifespan considerations. D4.4 therefore acts as the validation bridge between toolchain development and future use within the FASTEST hybrid testing environment. It defines how the toolchain outputs will be evaluated against safety thresholds, ageing indicators, reliability behaviour and experimental observations.

The deliverable also supports alignment with WP5 and WP6. For WP5, the validation approach must be compatible with Digital Twin principles, including data traceability, asset mapping and reusable model outputs. For WP6, the validated toolchain contributes to the larger hybrid testing platform by supporting the decision logic for when virtual testing can substitute, complement or reduce physical test effort.

1.3 Structure of the Deliverable

The deliverable is organised to follow the natural validation chain. Section 2 defines the system context and degradation scenarios, including functional safety, ageing and reliability requirements. Section 3 presents the Digital Twin-compliant validation framework, including validation criteria, metrics, KPIs and the verification architecture. Sections 4 and 5 collect the partner inputs required for technology-specific verification and experimental validation. Section 6 assesses multi-concern failure behaviour across electrical, and thermal domains. Section 7 evaluates toolchain performance against experimental evidence. Section 8 consolidates the validation conclusions, recommendations and readiness for further integration.

2. System Context and Degradation Scenarios

This section defines the technical basis for validation. Before a toolchain can be validated, the system boundary, battery configuration, degradation scenarios and safety-relevant operating limits must be clearly described. In this section the link between the scenarios defined in T4.1 and the validation logic used in D4.4 is established. The detailed battery baseline and technology-specific information is structured into a validation-oriented scenario framework.

2.1 Description of Degradation Scenarios

The degradation scenarios considered in D4.4 are derived from the baseline toolchain architecture and scenario definition work performed in T4.1. They are intended to represent conditions under which the battery system may experience performance degradation, safety-critical behaviour or reliability reduction. The scenarios should cover normal operation, accelerated ageing, boundary operating conditions and failure-inducing stress cases.

At a minimum, the validation framework should distinguish between electrical, and thermal degradation scenarios. Electrical scenarios may include overcharge, overdischarge, short-circuit conditions, high-current pulses, voltage imbalance, abnormal current flow and operation outside defined voltage windows. Thermal scenarios may include low-temperature operation, high-temperature exposure, thermal gradients across cells or modules, heat accumulation under fast charging, and possible thermal runaway precursor conditions.

For each scenario, the deliverable should document the scenario purpose, the expected physical phenomenon, the relevant monitored parameters, the required experimental evidence, and the toolchain outputs that will be compared against those observations. This structure ensures that validation remains scenario-specific and avoids generic statements about toolchain performance. Table 1 defines the scenario categories to be used in D4.4. The table is intended as a validation planning matrix. For each scenario, the partner evidence required, the expected toolchain output and the validation interpretation are stated.

Table 1 Validation scenario categories derived from D4.1 and T4.4

Scenario ID	Scenario Category	Example Stress Condition	Expected Concern	Relevant Toolchain Output	Validation Evidence Required
SC-E-01	Electrical	Overcharge / overvoltage	Safety limit violation, heat generation, degradation acceleration	Voltage response, temperature rise, failure warning	FM test data, voltage/current/ temperature traces

SC-E-02	Electrical	External or internal short-circuit condition	Rapid current increase, thermal escalation	Fault detection, severity index, thermal response	FM electrical abuse results
SC-T-01	Thermal	High-temperature operation	Accelerated ageing, reduced safety margin	Temperature prediction, degradation rate, risk index	VTT thermal chamber results
SC-T-02	Thermal	Low-temperature operation with charging/discharging	Lithium plating risk, performance loss, increased resistance	Resistance growth indicator, safe operating boundary	VTT low-temperature data
SC-C-01	Coupled	Fast charging under high/low temperature	Combined electrical-thermal ageing and safety risk	Coupled thermal-electrical response, ageing indicator	FM/VTT fast charge/discharge data

2.2 Battery System Baseline Definition

The battery system baseline definition provides the reference configuration used for the validation and verification activities within D4.4. The baseline establishes the physical and operational boundaries against which the FASTEST AI-powered toolchain is evaluated. It defines the battery cell technologies, operating conditions, measurable parameters, abuse scenarios, and data acquisition structure required for Digital Twin-compatible validation.

ABEE contributes to the definition of the baseline battery system by providing representative battery cell configurations, operating limits, and test-related technical parameters associated with the FASTEST validation activities. The baseline definition also supports interoperability between experimental data, AI-driven estimators, and the broader FASTEST hybrid testing platform.

The validation baseline considers two representative lithium-ion cell generations:

- Generation 3 (Gen3) lithium-ion battery cells
- Generation 4 (Gen4) lithium-ion battery cells

These battery technologies are evaluated under multiple operational and abuse-related scenarios, including:

- Overcharge conditions
- Overdischarge conditions
- Extreme cold condition (XCC) operation

The baseline definition includes electrical, thermal, and ageing-related variables that are required by the AI-powered estimation framework developed within the FASTEST project.

2.2.1 Baseline Parameters

The battery baseline Table 2 includes the following categories of parameters:

Table 2 Battery baseline parameters

Category	**Parameters**
Cell identification	Cell ID, cell generation/type
Electrical behaviour	Cell voltage, OCV-SOC curves, current profiles
Thermal behaviour	Cell temperature, ambient temperature
Ageing indicators	Capacity fade, internal resistance increase
Abuse-condition parameters	Overcharge SOC, overdischarge SOC, extreme cold conditions
Operational variables	Charging/discharging current, test temperature

The validation baseline is designed to support both experimental validation and AI-based prediction workflows. The data structures are aligned with Digital Twin requirements and allow direct mapping between physical measurements and virtual model representations.

The baseline configuration also defines the required measurement ranges for voltage, temperature, capacity, and internal resistance. These ranges are used to evaluate whether the toolchain can correctly identify deviations from normal operation and estimate degradation behaviour under stress conditions.

The AI-powered toolchain developed within FASTEST uses the baseline configuration to:

- Estimate cell voltage evolution under abuse conditions
- Predict thermal behaviour during abnormal operating scenarios
- Estimate capacity degradation after electrical abuse
- Estimate internal resistance growth after ageing and stress exposure
- Predict behaviour under extreme cold operating conditions

The baseline framework supports traceable integration between experimental data sources and machine learning estimators. This structure enables future scalability toward module-level and system-level Digital Twin integration.

2.2.2 AI-Based Battery Behaviour Estimation Framework

ABEE contributed to the validation framework through the development and integration of the machine learning-based toolkit referred to as *fastest_util* shown in Table 3. The toolkit supports battery behaviour estimation under abuse and accelerated ageing conditions.

Table 3 Summary of Battery Cell Tests and Corresponding Model Prediction Objectives covered by fastest_util

Cell Type	Abuse Test Type	Model Purpose
Gen3	Over-charge / Over-discharge	Predict capacity fade, IR increase, cell temperature, and cell voltage
Gen3	Extreme Cold Temperature Test	Predict capacity fade, and IR increase
Gen4	Over-charge / Over-discharge	Predict capacity fade, IR increase, cell temperature, and cell voltage
Gen4	Extreme Cold Temperature Test	Predict capacity fade, and IR increase

The toolkit includes dedicated estimators for:

- Cell voltage prediction under overcharge and overdischarge
- Cell temperature estimation under abuse conditions
- Capacity degradation prediction
- Internal resistance growth prediction
- Extreme cold condition (XCC) performance estimation

The framework supports both training and inference workflows and allows integration of experimental data from FASTEST partners.

We selected a Multilayer Perceptron (MLP) neural network architecture to capture potential non-linear relationships in the dataset. To justify this choice, we conducted a baseline comparison against a standard linear regression model. While both models yielded comparable results due to the moderate complexity of the underlying data patterns, the MLP consistently demonstrated marginal performance gains. I. To determine the optimal network configuration, we performed a systematic grid search for hyperparameter optimization. The general architecture consists of:

- Dense(64, relu)
- Dense(32, relu)
- Dense(16, relu)

A Dense layer, also known as a fully connected layer, is a standard neural network layer where every neuron receives input from all neurons in the previous layer. In this architecture, data passes sequentially through three Dense layers with 64, 32, and 16 neurons, respectively, creating a funnel effect that compresses the data to extract its most critical features. Each layer performs a matrix multiplication of inputs and weights, adds a bias, and applies the ReLU (Rectified Linear Unit) activation function. The output layer configuration depends on the prediction target and state-of-charge (SOC) levels used during training.

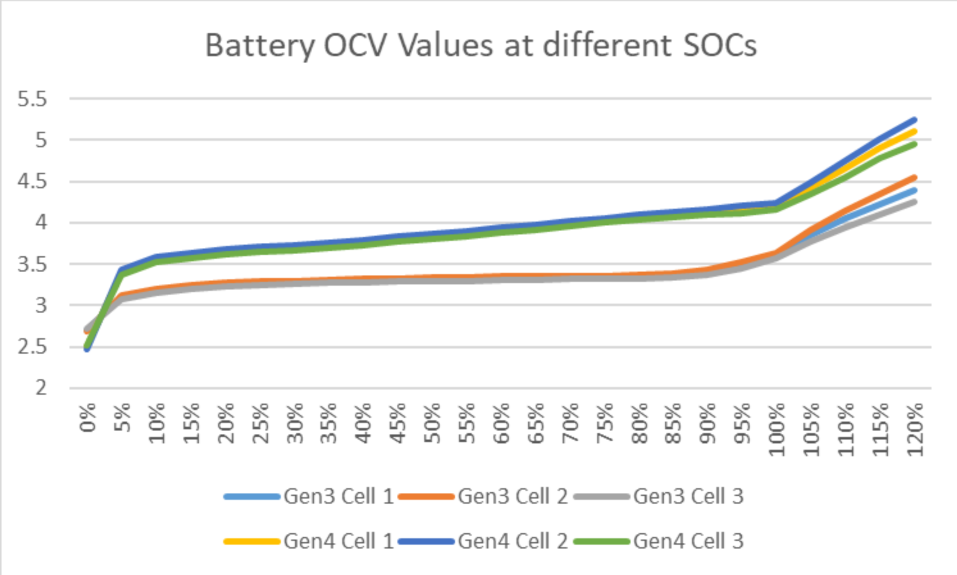

Figure 1 Load Voltage of Gen3 and Gen4 Batteries at Different SoC Levels

The training configuration includes:

- Loss function: Mean Squared Error (MSE)
- Validation metric: Mean Absolute Error (MAE)
- StandardScaler preprocessing for continuous variables
- OneHotEncoder preprocessing for categorical cell-type variables
- Validation split: 20%
- Batch size: 16
- Epochs: 100

The framework is architected for continuous adaptation, allowing for future retraining on diverse experimental datasets. This flexibility enables the model to incorporate different battery cell types, varying current profiles, expanded states of charge (SOC) including extreme conditions up to 125%—and internal resistance evaluations across diverse operating points. Furthermore, the model can ingest data from extreme temperature abuse testing, capturing battery behavior under more severe cold and hot environments.

2.3 Functional Safety, Ageing and Reliability Requirements

The validation strategy must be anchored in functional safety, ageing and reliability requirements. Functional safety requirements define the conditions under which the battery system must remain within safe operating limits or detect a transition towards unsafe operation. Ageing requirements define the expected ability of the toolchain to represent degradation mechanisms such as capacity fade, internal resistance growth and loss of power capability. Reliability requirements define whether the toolchain can support assessment of failure probability, failure progression and risk mitigation for critical components. For D4.4, these requirements are used as acceptance criteria for validation. The toolchain should be evaluated not only on numerical accuracy, but also on its ability to preserve the correct safety interpretation. Therefore, the validation methodology combines quantitative metrics with functional interpretation.

The functional safety assessment should identify safety-critical thresholds, measurable warning indicators and fail-safe decision boundaries. Examples include maximum allowable cell voltage, minimum allowable voltage, maximum cell temperature, maximum temperature gradient, allowable current or C-rate, internal resistance thresholds, and capacity retention limits. Final numerical values should be provided by the testing partners according to the battery baseline and test procedures. Table 4 presents the requirement categories used in the D4.4 validation framework.

Table 4 Functional safety, ageing and reliability requirement categories for D4.4 validation

Requirement Area	Example Requirement	Required Input	Validation Interpretation	Responsible Input Provider
Functional safety	Prediction of unsafe voltage or temperature threshold crossing	Voltage, current, temperature limits	Toolchain must detect or predict boundary violation consistently with test evidence.	RSTER / ABEE / FM / VTT
Ageing	Representation of capacity fade or resistance growth under stress	Capacity, resistance, cycling data	Toolchain trend should match measured degradation within agreed tolerance.	RSTER / ABEE / FM / VTT
Reliability	Identification of failure initiation or propagation risk	Failure mode evidence, test observations, FMEA outputs	Toolchain should assign higher risk to experimentally observed critical conditions.	RSTER / FM / VTT
DT compatibility	Traceable link between asset, test, data and model output	Asset identifiers, test IDs, model versions, data timestamps	Validation evidence must be reusable in the DT data structure.	RSTER / ABEE

3. Digital Twin (DT) Compliant Validation Framework

The Digital Twin-compliant validation framework defines how the FASTEST toolchain will be evaluated in a way that supports future integration with the project’s Digital Twin and hybrid testing platform. A DT-compliant validation framework requires more than numerical comparison between model and experiment. It requires traceable data structures, clear links between the physical asset and virtual representation, defined input and output interfaces, version-controlled model outputs, and interpretable validation metrics.

The validation framework proposed follows a closed-loop logic: experimental data and battery baseline information are used as inputs to the toolchain; the toolchain produces safety, ageing and reliability outputs; those outputs are compared with experimental evidence; deviations are quantified using validation metrics; and the results are fed back as calibration, refinement or limitation statements

3.1 DT Criteria for Safety, Age and Reliability Evaluation

For the purposes of D4.4, Digital Twin compliance is interpreted as the ability of the validation process to support consistent exchange between the physical battery system, the virtual toolchain models and the project-level data architecture. This requires that each validation case is uniquely identifiable and traceable. The validation record should include the tested asset, battery configuration, degradation scenario, test conditions, input data set, model version, simulation output, experimental result, validation metric and final validation status.

The safety, ageing and reliability evaluation must therefore be mapped to DT-relevant information objects. Safety evaluation requires thresholds, warning indicators and failure classifications. Ageing evaluation requires time- or cycle-dependent degradation variables such as capacity loss, internal resistance evolution and thermal behaviour under repeated loading. Reliability evaluation requires information on failure modes, severity, occurrence likelihood, detectability and propagation consequences where available.

The proposed DT criteria for D4.4 include: traceability of validation cases; consistency between experimental and simulation inputs; compatibility of toolchain outputs with battery asset data structures; clear definition of uncertainty and boundary conditions; and availability of validation evidence for future reuse in the hybrid testing platform. Table 5 summarizes the DT criteria adopted for D4.4.

Table 5 Digital Twin compliance criteria for D4.4 validation evidence.

DT Criterion	Description	Minimum Evidence Required in D4.4
Traceability	Each validation case must link asset, test, data, model and result.	Validation case ID, test ID, battery ID, model version.
Input consistency	The same operating and boundary conditions must be used in experiment and simulation where possible.	Test conditions, initial SOC, temperature, C-rate, load profile.
Output comparability	Toolchain outputs must be comparable with measured quantities or derived indicators.	Mapped output variable, unit, sampling interval, reference data source.
Safety interpretability	Outputs must support safety-relevant interpretation, not only numerical prediction.	Threshold comparison, warning classification, risk category.
Reusability	Validation results should be usable for future DT or hybrid testing workflows.	Structured KPI table, scenario metadata, limitation statement.

To illustrate how the Digital Twin-compliant validation logic is applied in practice, each validation case should be documented as a traceable validation record linking the physical test asset, experimental dataset, model version, toolchain output, KPI result and final validation status. This ensures that validation evidence is not treated as an isolated test result, but as reusable information that can be transferred into the FASTEST hybrid testing and Digital Twin environment.

For example, an overcharge validation case may be recorded by linking the tested Gen3b module, the corresponding FM overcharge test identifier, the measured voltage/current/temperature dataset, the AI-assisted voltage prediction model version, the calculated prediction error and the final validation status. This structure as shown in Table 6 enables future users of the toolchain to understand which results are supported by direct experimental evidence, which are partially supported, and which remain outside the demonstrated validation boundary.

Table 6 Example validation record

Validation Record Field	Example Entry
Validation case ID	VC-E-OC-GEN3b-01
Asset / battery identifier	Gen3b LFP module
Test ID / scenario	FM overcharge validation test / SC-E-01
Experimental dataset	Voltage, current and temperature trace at 25 °C, 1C charge
Model / toolchain version	AI voltage prediction model v0.2 / D4.4 validation workflow
Toolchain output	Predicted voltage response and threshold-warning interpretation
KPI result	Voltage prediction error / threshold detection result
Validation status	Validated
Limitation note	Valid within tested boundary conditions and available experimental evidence

This record structure should be applied consistently for electrical, thermal, ageing and coupled electro-thermal scenarios. Where complete experimental evidence is unavailable, the validation record should explicitly indicate whether the case is partially validated or not assessable, together with the reason for the limitation.

3.2 Validation Metrics and KPIs

Validation metrics and KPIs provide the quantitative basis for evaluating toolchain performance. The selection of metrics must reflect the type of output being validated. Continuous numerical outputs such as voltage, current, temperature, resistance and capacity can be evaluated using error-based metrics. Event-based outputs such as failure detection, threshold crossing or warning classification require classification metrics and timing metrics. Risk-based outputs require consistency checks between predicted risk ranking and observed failure severity.

The deliverable should distinguish between core validation KPIs and supporting indicators. Core KPIs provide the main evidence for validation status, while supporting indicators help interpret deviations, uncertainty and model limitations. Final acceptance thresholds should be defined once experimental data availability and model output types are confirmed. Where numerical thresholds cannot yet be finalized, the deliverable should define the metric, calculation method and intended interpretation. The KPIs in Table 7 are proposed as the standard validation metric set for D4.4. They are intentionally defined at a general level so that they can be applied across the specific validation cases.

Table 7 Validation metrics and KPIs proposed for D4.4.

KPI / Metric	Purpose	Typical Formula / Basis	Applies To	Interpretation
RMSE	Quantifies average prediction error for continuous variables.	sqrt(mean((y_pred - y_exp)^2))	Voltage, temperature, current, capacity	Lower values indicate better numerical agreement.
MAE	Measures mean absolute deviation without giving excessive weight to outliers.	mean(|y_pred - y_exp|)	Continuous outputs	Useful for general accuracy reporting.
MAPE / percentage error	Normalised error relative to measured value.	mean(|(y_exp - y_pred)/y_exp|) × 100	Capacity, resistance, temperature rise	Supports comparison across variables with different magnitudes.
Threshold detection accuracy	Evaluates whether safety boundary crossings are correctly identified.	Correct detections / total threshold events	Voltage, temperature, current safety limits	Critical for functional safety interpretation.
Time-to-event error	Compares predicted and observed timing of failure/warning events.	t_pred - t_exp	Thermal escalation, failure warning, cutoff event	Assesses prognostic relevance.
Risk ranking consistency	Checks whether higher predicted risk corresponds to more severe observed outcomes.	Rank comparison / qualitative consistency	FMEA-linked outputs, severity indices	Supports reliability and risk validation.
Uncertainty band coverage	Assesses whether experimental results fall within predicted uncertainty intervals.	Observed data within uncertainty interval	Probabilistic predictions	Indicates robustness of model confidence.

The acceptance criteria associated with the validation KPIs are defined as indicative thresholds for interpreting the validation maturity of the toolchain outputs. These thresholds are not intended to represent final certification limits; rather, they provide a structured basis for classifying validation evidence consistently across the D4.4 scenarios. Final numerical thresholds may be refined once additional experimental datasets, repeated test campaigns and model calibration results become available.

For continuous prediction outputs, such as voltage, temperature, capacity and internal resistance, the validation assessment is based on agreement between the predicted and measured values using metrics such as RMSE, MAE and percentage error. For safety-relevant events, such as voltage-limit crossing or BMS disengagement, correct event interpretation is considered more important than numerical error alone. For ageing and reliability-related outputs, trend consistency and correct risk interpretation are used where direct numerical validation is limited Table 8.

Table 8 Indicative KPI acceptance logic for D4.4

Validation Aspect	Indicative Acceptance Logic	Validation Interpretation
Voltage prediction	Low RMSE/MAE and correct reproduction of voltage trend under the tested SOC range	Supports validated or partially validated status depending on dataset completeness
Temperature prediction	Correct temperature trend and acceptable deviation from measured temperature response	Supports thermal validation where experimental evidence is available
Capacity degradation	Correct direction and approximate magnitude of capacity fade	Supports ageing trend validation
Internal resistance growth	Correct trend of resistance increase under stress or low-temperature conditions	Supports degradation and reliability interpretation
Threshold detection	Correct identification of voltage, current or temperature boundary crossing	Required for functional safety validation
Time-to-event estimation	Predicted warning or threshold-crossing time reasonably aligned with observed event timing	Supports prognostic relevance
Risk ranking consistency	Higher predicted risk assigned to experimentally more critical conditions	Supports reliability interpretation
Uncertainty band coverage	Experimental result falls within expected uncertainty range where uncertainty bands are available	Supports confidence in model robustness

For D4.4, the validation status should therefore be interpreted as follows. A scenario may be classified as **validated** when KPI agreement is acceptable and the safety or degradation interpretation is consistent with the experimental evidence. A scenario should be classified as **partially validated** when the toolchain captures the correct trend or safety interpretation but the available evidence is incomplete, limited to proxy indicators or affected by experimental constraints. A scenario should be classified as **not assessable** when the required experimental evidence is unavailable or when the boundary conditions do not allow a reliable comparison.

3.3 Functional Safety Assessment Methodology

The functional safety assessment methodology evaluates whether the toolchain can support safety-relevant decision-making under the degradation scenarios considered in D4.4. The focus is not limited to reproducing measured signals. The toolchain must also correctly identify when a simulated or measured condition approaches or exceeds safety-relevant limits. This is important because the practical value of the toolchain lies in supporting safer and more reliable battery system development through early identification of hazardous behaviour.

The methodology consists of five steps. First, the safety-relevant scenario is defined, including expected failure mode and monitored variables. Second, safety thresholds and warning limits are identified using battery baseline information, applicable procedures and partner expertise. Third, the toolchain is executed using equivalent input conditions. Fourth, predicted safety indicators are compared with experimental outcomes. Fifth, a validation status is assigned: validated, partially validated, not validated, or not assessable due to missing evidence.

For cases where experimental failure cannot be induced for safety or practical reasons, validation may rely on boundary-condition evidence, lower-severity proxy tests or literature-supported engineering interpretation. Such cases must be clearly marked as partial validation or qualitative validation. This avoids overstating the maturity of the toolchain while still documenting the available evidence. Table 9 defines the validation status categories used in D4.4. The purpose of this classification is to avoid overstating validation maturity.

Table 9 Validation status classification for D4.4.

Validation Status	Definition	Typical Evidence
Validated	Toolchain prediction agrees with experimental result within agreed KPI limits and safety interpretation is correct.	Numerical agreement + correct threshold/failure classification.
Partially validated	Toolchain captures trend or safety interpretation, but deviation exceeds preferred numerical tolerance or evidence is incomplete.	Correct trend but limited data or partial scenario coverage.
Not validated	Toolchain output does not match experimental	Large deviation or missed critical event.

	behaviour or safety classification is incorrect.	
Not assessable	Experimental or model evidence is insufficient for validation.	Missing data, incompatible test conditions, unavailable output variable.

3.4 Toolchain Verification Architecture

The toolchain verification architecture describes the information flow used to verify that the toolchain is correctly configured and to validate that its outputs are credible. The architecture begins with inputs about: battery baseline information, degradation scenario definition from T4.1, and experimental test data. These inputs are transformed into toolchain-ready parameters and boundary conditions. The toolchain then produces outputs related to safety, ageing and reliability, which are compared against measured or observed evidence.

A central requirement of this architecture is traceability. Every validation result must be linked back to the input data set, model version, test conditions and experimental source. This allows future users of the toolchain to understand whether a prediction is supported by strong experimental evidence, partial evidence or only engineering assumptions.

The verification architecture also includes a feedback loop. If deviations are identified, the root cause should be classified as input uncertainty, model limitation, missing physics, insufficient calibration, measurement uncertainty or mismatch between experimental and simulated boundary conditions. This feedback is important for improving the toolchain and for defining its boundary of validity.

4. Verification Methods and Supporting Technical Procedures

This section describes the verification methods, technical procedures, and supporting validation activities contributed by ABEE for the FASTEST AI-powered battery toolchain. The objective is to ensure that the implemented machine learning estimators, input structures, and data-processing workflows operate consistently with the intended validation architecture.

The verification activities focus on:

- Consistency of model implementation
- Correct handling of input datasets
- Validation of prediction workflows
- Compatibility between physical measurements and AI estimators
- Reproducibility of training and inference procedures
- Traceability of generated outputs

The verification procedures support the Digital Twin-oriented workflow adopted within FASTEST by ensuring that experimental data and virtual representations remain synchronized and technically interpretable.

4.1 Technology-Specific Validation Inputs

The FASTEST AI-powered validation framework requires battery-specific technical inputs to ensure that the developed estimators and Digital Twin-compatible workflows accurately represent the behaviour of the target battery technologies under both nominal and abuse operating conditions. ABEE contributes to this process by defining the technology-specific validation inputs associated with the Gen3 and Gen4 lithium-ion battery systems used within the FASTEST validation activities.

The objective of these validation inputs is to establish a consistent and traceable link between experimental measurements, AI-based estimation models, and the virtual validation environment. The inputs provided support the development, training, verification, and evaluation of machine learning estimators related to electrical behaviour, thermal response, ageing evolution, and abuse-condition performance.

The technology-specific validation framework includes the following categories of input data Table 10:

Table 10 Technology specific validation framework

Input Category	Description	Validation Purpose
Cell identification data	Cell generation, cell type, chemistry reference	Enables model differentiation between Gen3 and Gen4 cells
Electrical measurements	Voltage, current, SOC-dependent behaviour	Supports electrical-state prediction and abuse-condition analysis
Thermal measurements	Cell temperature, ambient temperature	Supports thermal-behaviour estimation and safety evaluation
Ageing indicators	Capacity retention, internal resistance growth	Supports degradation and reliability prediction
OCV reference data	Open-circuit voltage curves across SOC levels	Provides baseline electrochemical behaviour
Operational parameters	Charging/discharging current, C-rate, operating temperature	Defines test and simulation boundary conditions
Abuse-condition inputs	Overcharge, overdischarge, XCC operation	Supports validation under abnormal operating scenarios

The validation inputs are structured to support the AI-powered estimation workflows implemented within the *fastest_util* framework. These workflows include dedicated machine learning estimators for:

- Cell voltage prediction
- Cell temperature estimation

- Capacity degradation estimation
- Internal resistance prediction
- Extreme cold condition (XCC) performance estimation

The validation datasets include measurements acquired under both normal operating conditions and accelerated stress scenarios. The considered abuse conditions include:

- Overcharge scenarios above nominal SOC limits
- Overdischarge conditions below safe operating windows
- Fast charging under elevated current conditions
- Extreme cold condition (XCC) operation near 0 °C

For overcharge validation activities, SOC-dependent measurements are generated for multiple stress levels including:

- 100% SOC
- 105% SOC
- 110% SOC
- 115% SOC
- 120% SOC

The input datasets include both static and dynamic measurements. Static measurements include parameters such as nominal capacity and internal resistance, while dynamic measurements include voltage and temperature evolution during charging and discharging cycles.

To ensure interoperability with the FASTEST Digital Twin architecture, all datasets are organized using traceable structures that preserve:

- Cell identifiers
- Test-condition metadata
- Measurement timestamps
- Experimental boundary conditions
- Validation scenario identifiers

This structure enables direct comparison between:

- Experimental measurements
- AI-generated predictions
- Toolchain-generated safety indicators

The machine learning workflow developed by ABEE uses preprocessing techniques to ensure consistency across heterogeneous datasets. Continuous variables are normalized using StandardScaler, while categorical battery-type variables are encoded using OneHotEncoder. This preprocessing pipeline ensures that the estimators can generalize across different battery technologies and operating conditions.

The technology-specific validation inputs also support the calculation of validation metrics including:

- Root Mean Square Error (RMSE)
- Mean Absolute Error (MAE)
- Absolute prediction error
- Prediction consistency across repeated simulations

These metrics are used to quantify the agreement between experimental observations and AI-generated outputs.

The validation framework is intentionally modular to support future extension toward:

- Additional battery chemistries
- Module-level validation
- Pack-level Digital Twin integration
- Real-time monitoring applications

Figure 2 presents the voltage prediction performance of the AI-powered estimation framework under overcharge abuse conditions. The left side of the figure illustrates the prediction error associated with each individual test sample at the evaluated target SOC levels. The right side summarizes the overall prediction performance using the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) metrics calculated across the complete test dataset. The results demonstrate the capability of the estimator to reproduce the voltage behaviour of the battery cells under abnormal operating conditions while maintaining acceptable prediction accuracy across different SOC levels.

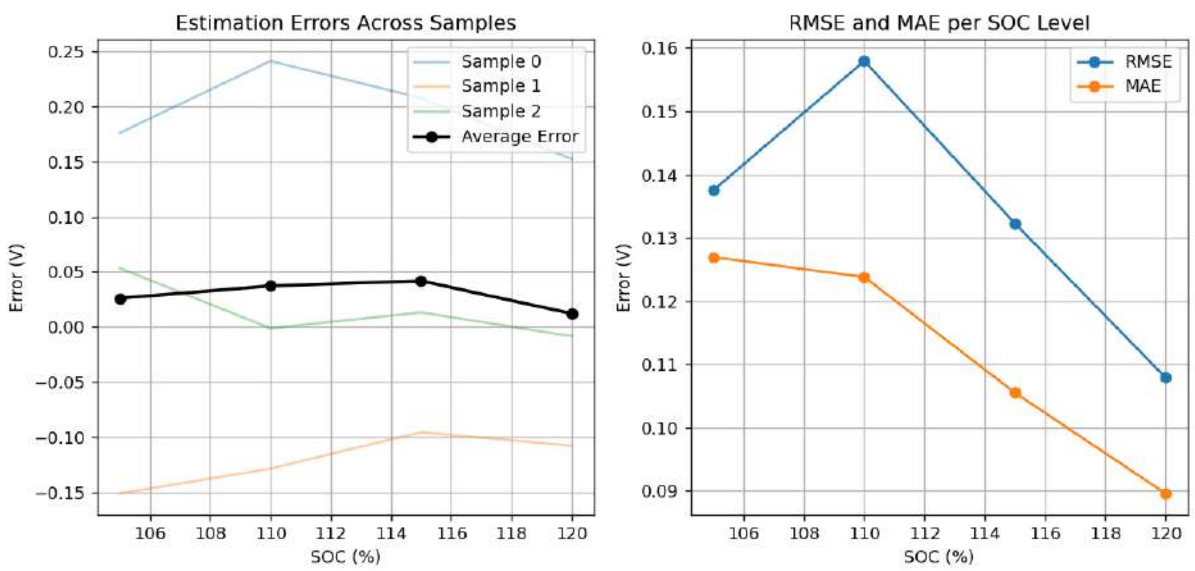

Figure 2 Prediction Errors and RMSE/MAE for Cell Voltage at Target SOC Levels

Figure 3 illustrates the temperature prediction performance of the AI-based estimation framework during overcharge abuse scenarios. The left side of the figure shows the prediction error corresponding to each individual sample within the test dataset, while the right side presents the overall RMSE and MAE values obtained for all predicted cell temperatures at the evaluated SOC levels. The results provide an assessment of the estimator's ability to capture the thermal behaviour of the battery cells and reproduce temperature variations occurring during abuse-condition operation.

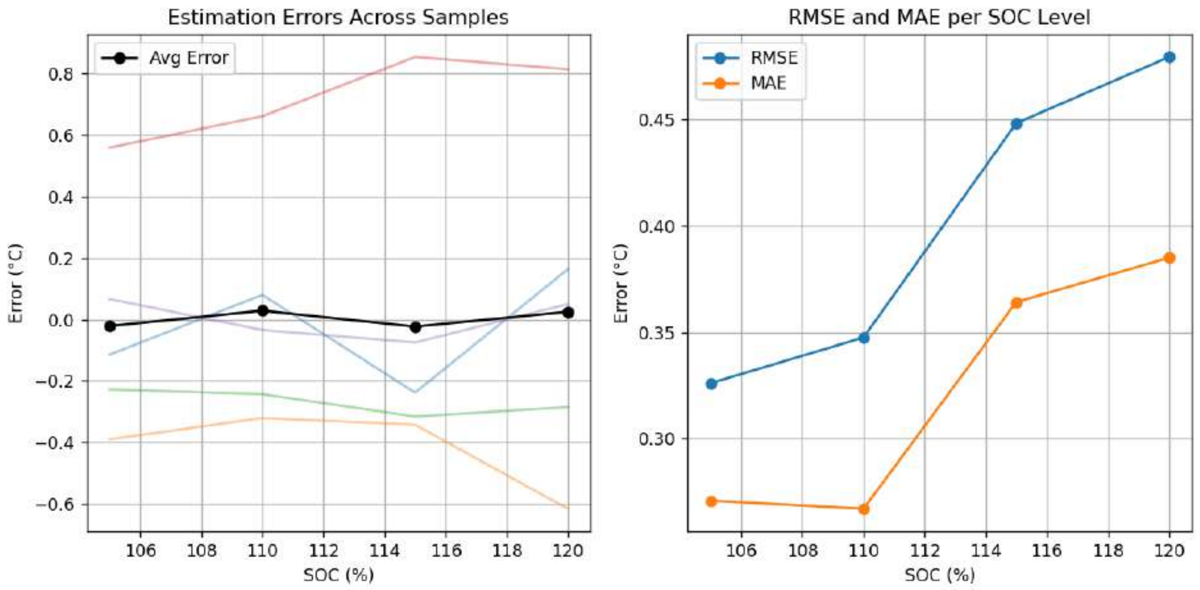

Figure 3 Prediction Errors and RMSE/MAE for Cell Temperature under Overcharge Abuse Conditions

Figure 4 presents the prediction results obtained for cell-capacity estimation under overcharge abuse conditions. The left side of the figure displays the prediction error associated with each individual test sample, whereas the right side summarizes the global prediction accuracy using RMSE and MAE metrics calculated over the entire dataset. The figure demonstrates the capability of the developed AI estimator to reproduce capacity degradation behaviour and estimate remaining cell capacity after exposure to electrical abuse conditions.

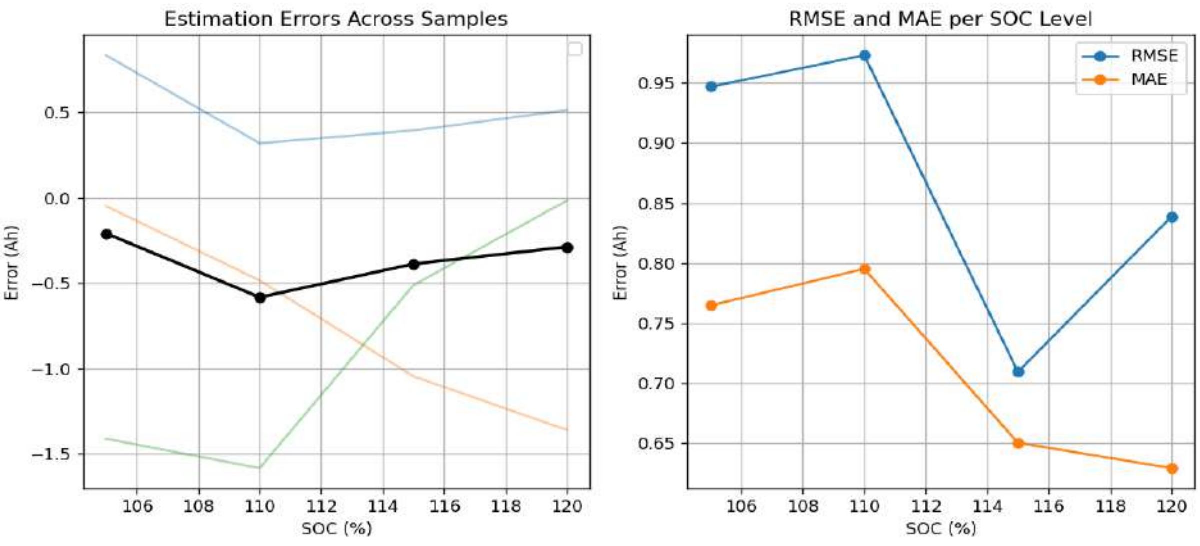

Figure 4 Prediction Errors and RMSE/MAE for Cell Capacity under Overcharge Abuse Conditions.

Figure 5 illustrates the prediction performance of the internal-resistance estimation model under overcharge abuse conditions. The left side of the figure shows the prediction error corresponding to each individual sample, while the right side presents the overall RMSE and MAE values calculated for all predicted internal-resistance values across the evaluated SOC levels. The obtained results demonstrate the ability of the estimator to capture resistance-growth behaviour associated with degradation and electrical stress conditions.

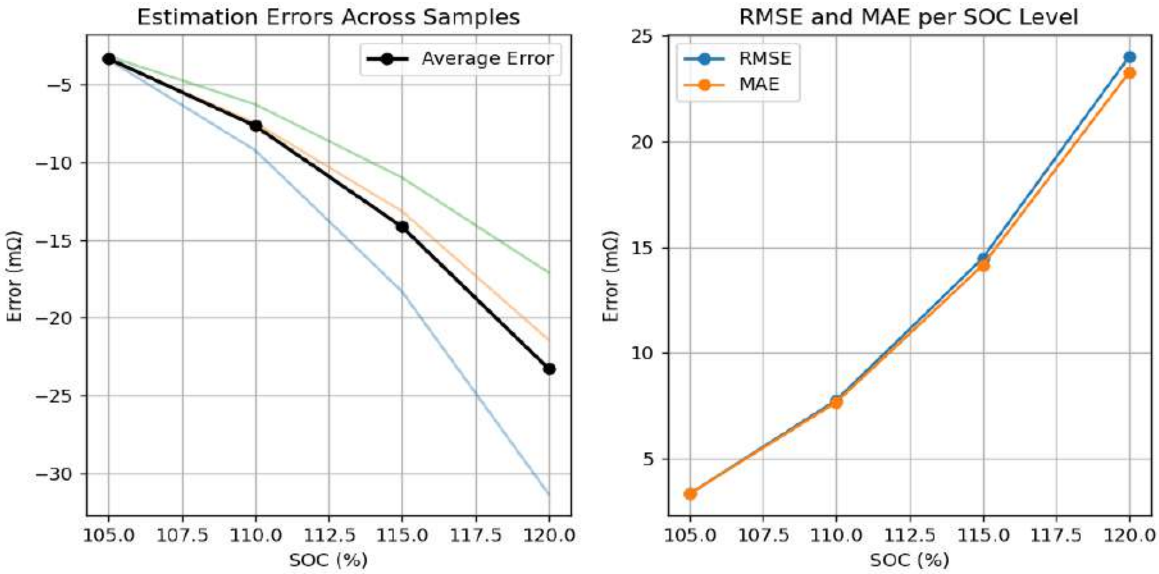

Figure 5 Prediction Errors and RMSE/MAE for Cell Internal Resistance under Overcharge Abuse Conditions.

Figure 5Figure 6 presents the absolute prediction error obtained for the cell internal-resistance estimator under extreme cold operating conditions at 0 °C. The figure illustrates the prediction deviation for each individual sample included in the test dataset. The results provide an evaluation of the estimator’s capability to reproduce internal-resistance behaviour at low temperatures, where battery performance and electrochemical behaviour are strongly affected by reduced thermal conditions.

For this validation scenario, the prediction results achieved an RMSE value of 0.354360 mΩ and an MAE value of 0.231080 mΩ across the complete test dataset. These results indicate that the estimator is capable of maintaining stable prediction performance under extreme cold operating conditions.

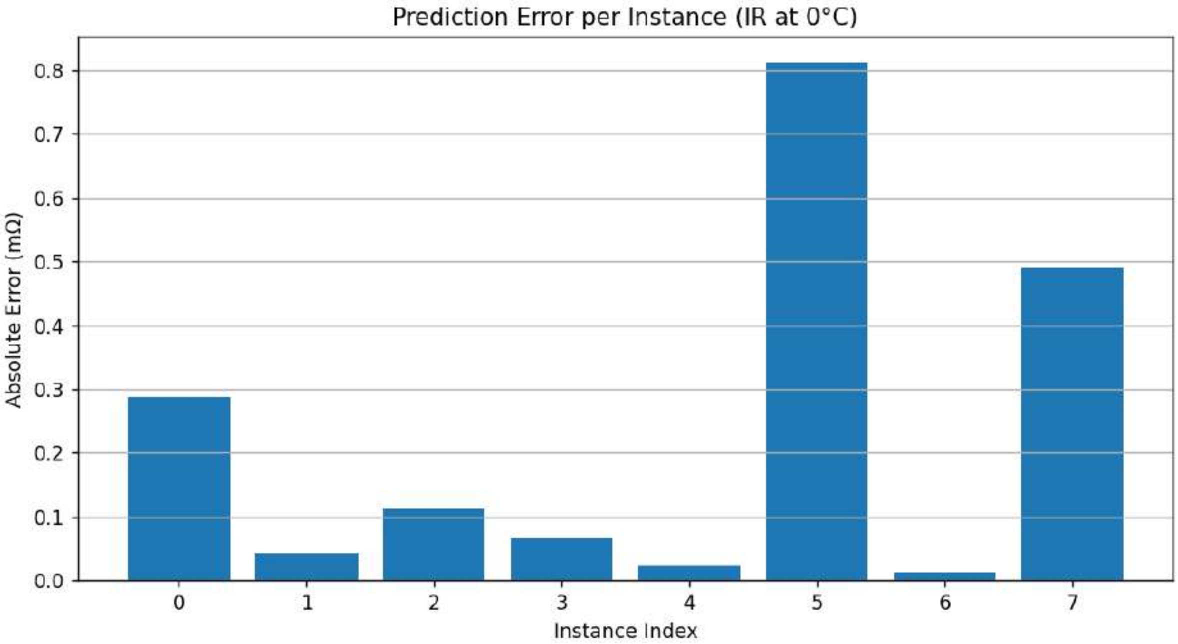

Figure 6 Absolute Prediction Errors for Cell internal resistance at 0 °C.

Figure 7 illustrates the absolute prediction error associated with the cell-capacity estimation model under extreme cold operating conditions at 0 °C. The figure presents the prediction deviation for each individual sample included in the test dataset and provides insight into the estimator's ability to reproduce low-temperature capacity behaviour.

For this validation scenario, the obtained results show an RMSE value of 0.0690 °C and an MAE value of 0.0637 °C acrossall evaluated test samples. The relatively low prediction-error values demonstrate the capability of the AI-powered estimator to maintain accurate capacity predictions under low-temperature operating conditions.

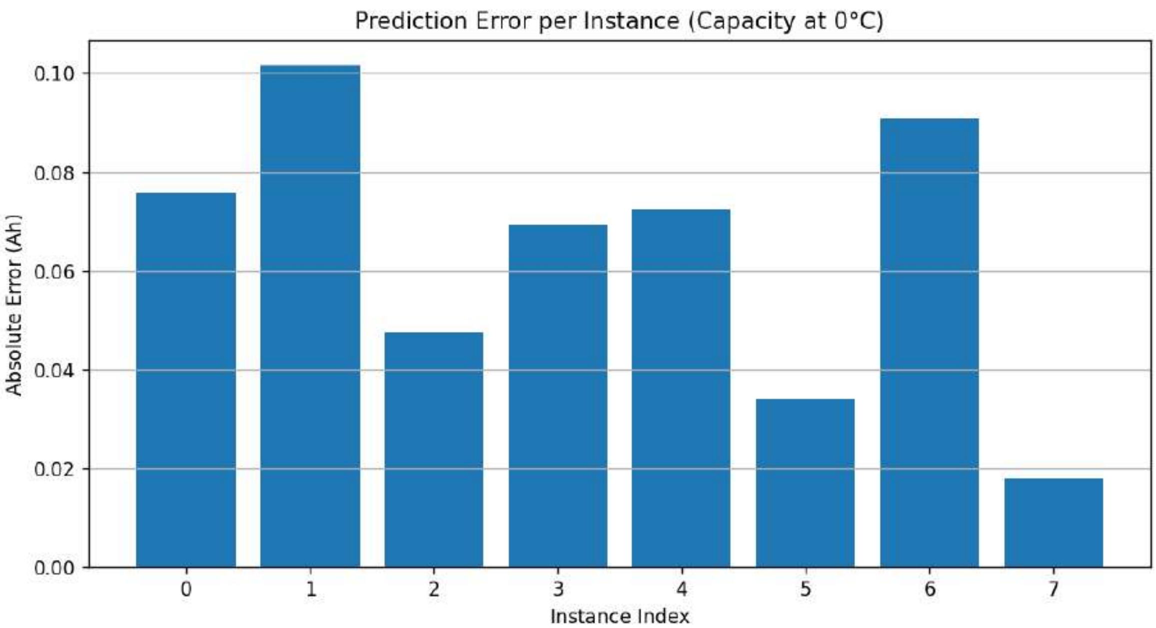

Figure 7 Absolute Prediction Errors for Cell Capacity at 0 °C

4.2 Internal Procedures and Coordination Framework

ABEE established an internal coordination and verification framework to ensure that the AI-powered battery validation activities performed within the FASTEST project remain consistent, traceable, and interoperable with the broader Digital Twin and hybrid testing architecture. The framework defines the internal procedures used for data handling, model training, prediction generation, verification activities, and coordination with project partners contributing experimental and validation data.

The objective of the coordination framework is to ensure that all AI-based estimators developed within the FASTEST toolchain operate according to a reproducible and technically validated workflow. The framework also supports alignment between software development activities, experimental validation procedures, and Digital Twin integration requirements.

The internal workflow implemented by ABEE follows a structured sequence consisting of:

1. Experimental data collection
2. Dataset preprocessing and organization
3. Input validation and consistency checks
4. AI model training or model loading
5. Prediction generation
6. Verification and error analysis
7. Export of validated outputs
8. Integration into the FASTEST validation environment

The framework supports collaboration between ABEE and the FASTEST validation partners by establishing standardized data structures and common validation procedures. Experimental measurements provided by FM and VTT are processed using unified preprocessing pipelines to ensure compatibility with the AI-powered estimation framework.

4.2.1 Data Preparation and Preprocessing Procedures

The preprocessing workflow includes the normalization and organization of battery datasets before they are used for training or inference. The implemented procedures include:

- Removal of incomplete or inconsistent records
- Alignment of SOC-dependent measurements
- Normalization of continuous variables using StandardScaler
- Encoding of battery-type variables using OneHotEncoder
- Separation of training and testing datasets
- Validation of measurement ranges and operational boundaries

These procedures ensure that the datasets remain consistent across multiple battery technologies and abuse-condition scenarios.

The framework also supports structured handling of:

- Voltage measurements
- Temperature measurements
- Capacity measurements
- Internal resistance measurements
- Ambient-condition variables
- Abuse-condition indicators

4.2.2 Model Management and Reproducibility

To ensure reproducibility of the validation activities, the AI framework supports both retraining and inference-only workflows.

When pretrained models are available, the framework automatically loads:

- Trained neural-network models
- Scaler configurations
- Encoder configurations
- Stored validation parameters

This architecture reduces computational overhead during repeated validation activities and guarantees consistency between multiple validation iterations.

The trained model artifacts are stored in structured directories to preserve:

- Model version information
- Associated preprocessing configurations
- Training parameters
- Validation metrics
- Timestamped outputs

This approach supports version control and traceability of all generated results.

4.2.3 Verification and Output Management

The generated outputs are automatically evaluated using quantitative verification metrics including:

- Root Mean Square Error (RMSE)
- Mean Absolute Error (MAE)
- Absolute prediction error
- Prediction consistency indicators

The framework supports export of:

- Prediction results
- Validation metrics
- Error distributions

- CSV-based evaluation files
- Visualization-ready outputs

These outputs are used to compare AI-generated predictions against experimental observations and to evaluate the overall performance of the FASTEST AI-powered toolchain.

4.3 Baseline Model Verification Strategy

The baseline model verification strategy developed by ABEE aims to ensure that the AI-powered estimators implemented within the FASTEST toolchain operate consistently with the intended validation architecture and produce technically reliable outputs for safety, ageing, and performance evaluation. The verification activities focus on confirming the correctness of the implemented workflows, the reproducibility of prediction results, and the consistency between AI-generated outputs and expected battery behaviour under representative operating and abuse conditions.

The verification strategy addresses both software-oriented verification and engineering-oriented validation. Software-oriented verification evaluates whether the implemented machine learning workflows correctly process input datasets, execute model inference procedures, and generate reproducible outputs. Engineering-oriented validation evaluates whether the generated predictions remain physically interpretable and sufficiently representative of expected battery-system behaviour.

The verification framework is applied to the machine learning estimators developed within the *fastest_util* toolkit, including:

- Voltage prediction estimators
- Temperature prediction estimators
- Capacity degradation estimators
- Internal resistance growth estimators
- Extreme cold condition (XCC) estimation models

The strategy supports validation activities associated with:

- Overcharge conditions
- Overdischarge conditions
- High-current charging scenarios
- Extreme cold operating conditions
- Ageing-related degradation behaviour

4.3.1 Verification Objectives

The baseline verification strategy evaluates whether:

- AI models can reproduce stable prediction outputs
- Prediction results remain physically consistent

- Estimators generalize across Gen3 and Gen4 battery technologies
- Generated outputs support safety and degradation assessment workflows
- Validation metrics remain within acceptable limits

The verification activities therefore combine quantitative evaluation methods with engineering interpretation of battery behaviour.

4.3.2 Quantitative Verification Metrics

The primary quantitative metrics Table 11 used for model verification include:

Table 11 Quantitative metrics for model verification

Metric	Purpose
RMSE (Root Mean Square Error)	Measures overall prediction deviation
MAE (Mean Absolute Error)	Measures average prediction error
Absolute prediction error	Evaluates local prediction consistency
Prediction repeatability	Evaluates stability of repeated inference
Trend consistency	Evaluates preservation of physical behaviour trends

The metrics are evaluated independently for:

- Voltage prediction
- Temperature prediction
- Capacity prediction
- Internal resistance prediction

The verification process also evaluates whether the estimators correctly preserve expected physical trends, including:

- Voltage increase during overcharge
- Temperature rise under high-current loading
- Capacity degradation after abuse exposure
- Internal resistance increase during ageing

4.3.3 Voltage and Thermal Verification

The voltage verification activities compare predicted voltage responses against reference voltage profiles generated across multiple SOC levels and abuse conditions. Particular attention is given to:

- Overvoltage prediction capability
- SOC-dependent voltage evolution
- Detection of abnormal electrical behaviour
- Preservation of electrical operating boundaries

The thermal verification activities evaluate whether the AI estimators can reproduce:

- Temperature increase during stress conditions
- Thermal response during overcharge
- Ambient-temperature dependency
- High-SOC thermal behaviour

These activities are essential for evaluating the capability of the FASTEST toolchain to support safety-oriented Digital Twin services.

4.3.4 Capacity and Internal Resistance Verification

The ageing-oriented verification activities focus on evaluating the ability of the AI framework to estimate degradation-related parameters.

The verification process includes assessment of:

- Capacity fade after abuse exposure
- Internal resistance growth after stress conditions
- Ageing behaviour under repeated loading cycles
- Performance degradation at low temperatures

Dedicated estimators are also verified for Extreme Cold Condition (XCC) operation at approximately 0 °C.

The XCC validation workflow evaluates:

- Capacity estimation accuracy at low temperature
- Internal resistance prediction under cold conditions
- Generalization across battery generations
- Stability of estimator outputs under reduced-temperature operation

5. Experimental Validation Strategy

Several validation performance tests are being performed to validate two types and generations of Li-ion chemistry: Gen4 NMC and Gen3b LFP.

The tests that are being performed fall under following categories:

- Electrical failure validation tests
- Thermal performance validation tests
- Fast Charging and discharging pattern validation tests

5.1 Electrical Failure Validation (Overcharge, Overdischarge, Current overload, Short Circuit)

5.1.1 Overcharge testing

The **overcharge test** will be performed according to ISO12405-2 9.3.2 stating following test procedure as in Table 12 and Table 13.

Steady conditions:

- Room temperature of 25 +/- 2 °C
- Overall stop condition is cell temp. > 45°C

GEN3b (LFP):

Table 12 Overcharge test procedure for GEN3b (LFP)

#	Test	Rate (C)	T Amb [°C]	Stop condition
1	1C charge test	1	25	CC to max cell 3.65V (100% SoC) CV to 1.5A
2	1C Overcharge test	1	25	CC to max cell 3.85V (>100% SoC)
3	BMS has to disengage before reaching 3.85V		25	30 min
4	1C discharge	1	25	CC to average cell 3.25V

Outcome: BMS opened relays at 3.66V max cell voltage. Test succeeded.

GEN4 (NMC):

Table 13 Overcharge test procedure for GEN4 (NMC)

#	Test	Rate (C)	T Amb [°C]	Stop condition
1	1C charge test	1	25	CC to max cell 4.2V (100% SoC) CV to 1.5A
2	1C Overcharge test	1	25	CC to max cell 4.4V (>100% SoC)
3	BMS has to disengage before reaching 4.4V		25	30 min
4	1C discharge	1	25	CC to average cell 3.7V

Outcome: BMS opened relays at 4.22V max cell voltage. Test succeeded.

5.1.2 Over-discharge testing

The **over-discharge test** will be performed according to ISO12405-2 9.4.2 stating following test procedure as in Table 14 and Table 15.

Steady conditions:

- Room temperature of 25 +/- 2 °C
- Overall stop condition is cell temp. > 45°C

GEN3b (LFP):

Table 14 Over-discharge test procedure for GEN3b (LFP)

#	Test	Rate (C)	T Amb [°C]	Stop condition
1	1C discharge test	1	25	CC to min cell 2.5V (0% SoC)
2	1C Over-discharge test	1/3	25	CC to min cell 2.25V (<0% SoC)

3	BMS has to disengage before reaching 2.25V		25	30 min
4	1C charge	1	25	CC to average cell 3.25V

Outcome: Test failed due to current still being drawn from the module when it was below 2.5V. The relays towards the module tester were opened but there was a secondary cooling/heating system which kept drawing power from the module without any interruptions, pulling the cells way below their safe voltage.

GEN4 (NMC):

Table 15 Over-discharge test procedure for GEN4 (NMC)

#	Test	Rate (C)	T Amb [°C]	Stop condition
1	1C discharge test	1	25	CC to min cell 2.75V (0% SoC)
2	1C Over-discharge test	1/3	25	CC to min cell 2.5V (<0% SoC)
3	BMS has to disengage before reaching 2.5V		25	30 min
4	1C charge	1	25	CC to average cell 3.7V

Outcome: Test hasn’t been performed since a similar cooling/heating system as found in the LFP module can be found here. This system could also discharge the NMC module below its safe voltage range, even when the relays are opened.

5.1.3 Current overload test

The purpose of this test is to verify the performance of the overcurrent protection during DC external charging to prevent the Rechargeable Electrical Energy Storage System (REESS) from any severe events caused by excessive levels of charge current as specified by the manufacturer.

Test conditions:

- The test shall is conducted at an ambient temperature of 20 ± 10 °C;
- The SoC of REESS shall be adjusted around the middle of normal operating range by normal operation recommended by the manufacturer. The accurate adjustment is not required as long as the normal operation of the REESS is enabled;
- The overcurrent level (assuming failure of external DC electricity supply equipment) and maximum voltage (within normal range) that can be applied shall be determined, if necessary, through consultation with the manufacturer.
- The external electricity supply equipment along with the overcurrent supply is connected to the REESS and is initiated to achieve the highest normal charge current specified by the manufacturer;
- The charge current is then increased over 5 seconds from the highest normal charge current to the overcurrent level determined. Charging is then continued at this overcurrent level;

- The charging shall be terminated when the functionality of the REESS overcurrent protection terminates charging or the temperature of the REESS is stabilized such that the temperature varies by a gradient of less than 4 °C through 2 hours;

Outcome: BMS opened relays right after current rise above allowed current value (BMS limited). Test succeeded.

5.2 Thermal Performance Validation (Low/High Temperature Operation)

The thermal performance validation tests are performed according to ISO 12405-4:2020 à 7.2 Energy and capacity at different temperatures and discharge rates

During this test the capacity is measured at 4 different temperatures and 3 three different constant current discharge rates: 40 °C, 0 °C, −10 °C and −18 °C with discharge rates being C/3, 1C, $C_{max,supplier}$)

The following data is reported in function of time:

- current, voltage, DUT temperature, ambient temperature
- Charged and discharged capacity in Ah, energy in Wh, energy round trip efficiency and average power in W
- Discharged energy in Wh as a function of SoC at each discharge test (in % of rated capacity)
- Diagram regarding the EODV (End Of Discharge Voltage) dispersion of the cells at each discharge test

Outcome: The tests haven't been performed due to the aforementioned issues with the cooling/heating system of the battery module

5.3 Fast Charging and Discharging Pattern Validation

The purpose of the energy efficiency at fast charging test is to determine the battery system behaviour and the energy efficiency at different fast charging levels. For high-energy battery systems, the energy efficiency also at fast charging of the used battery system has a significant influence on the overall vehicle efficiency.

The test shall be performed with the battery module at Room Temperature, 0°C and $T_{min,supplier}$ at three different fast charging levels being 1C, 2C and I_{max}. After thermal equilibration and conditioning of the battery module by a standard cycle, the module first shall be discharged via a standard discharge followed in the next step by a fast charge with a starting current of 1C, 2C and I_{max}. The charge rate, the maximum charge current I_{max} and the minimum ambient test temperature T_{min} shall follow the requirements delivered by the supplier.

The following data shall be reported:

- current, voltage, DUT temperature and ambient temperature versus time at each discharge test and the following fast charge;
- discharged capacity in Ah, energy in Wh and average power in W at each discharge test;
- charged capacity in Ah, energy in Wh and average power in W following each discharge test;
- the EODV of all available cell voltage measuring points for all performed discharge tests;
- energy efficiency for specified ASOCs at each standard discharge - fast charge test.

Outcome: The tests haven't been performed due to the aforementioned issues with the cooling/heating system of the battery module

5.4 Data Acquisition and Measurement Protocols

No tests with required data acquisition have been performed due to the aforementioned issues with the cooling/heating system of the battery module.

5.5 Justification for Non-Performed Experimental Tests

During the execution of the experimental validation activities described in this deliverable, a subset of the planned tests could not be completed within the project timeframe. This situation was not related to a lack of preparation or compliance with the defined test procedures, but to unexpected technical interactions identified at module level during the early stages of integrated experimental execution.

All battery modules used in the experimental campaign were prototype modules, representative of an advanced development stage but not yet industrialised products. As is typical for prototype-level systems, certain system-level interactions only become visible under integrated operational conditions, particularly when electrical, thermal and control subsystems operate simultaneously.

Specifically, during the execution of over-discharge validation tests at module level, it was observed that auxiliary subsystem interactions within the battery module affected the electrical boundary conditions of the test configuration. Even after the Battery Management System (BMS) correctly disengaged the main power relays upon reaching safety thresholds, a secondary subsystem associated with thermal conditioning continued to draw residual power from the module. This behaviour led to operating conditions outside the intended safe voltage range and prevented the controlled completion of the test according to the prescribed standards.

Once this behaviour was identified and analysed, the battery modules were retrieved by the module provider (AVESTA) to allow further technical assessment and to ensure safe handling of prototype hardware. Given the advanced stage of the project at the time of this discovery, and the dependency of the affected tests on the availability of the exact prototype module configuration, there was insufficient remaining project time to re-execute the full experimental campaign under revised conditions.

The same interaction mechanism was therefore considered as potentially impacting other planned experimental activities requiring stable electrical isolation after relay opening, including:

- thermal performance validation tests under low- and high-temperature conditions,
- fast charging and discharging pattern validation,
- extended data acquisition under combined electrical and thermal boundary conditions.

Given the safety-critical nature of these tests and the risk of introducing non-representative or ambiguous experimental evidence, the consortium collectively decided not to proceed with further module-level experimentation under these conditions. This decision was taken to ensure:

- protection of prototype test assets,
- preservation of safe laboratory operation,
- avoidance of results that could not be unambiguously interpreted within the defined validation framework.

Importantly, this limitation is not attributed to a failure of the Battery Management System, test procedures, or partner execution, but to emergent system-level behaviour typical of prototype-level validation. Such findings are consistent with the transition from cell-level characterisation to module-level integration and provide useful learning for future system refinement.

From a project perspective, these limitations do not impact the defined Key Performance Indicators (KPIs), the core objectives of WP4, nor the overall outcomes of the FASTEST project. The primary KPIs related to toolchain validation, Digital Twin compliance, safety-relevant scenario coverage, and methodological robustness remain fully addressed through the combination of:

- completed experimental evidence,
- validated cell-level datasets,
- AI-based estimators,
- and clearly documented validation boundaries.

The absence of specific module-level datasets has been transparently accounted for in this deliverable. Affected test cases are explicitly marked as not assessable or partially validated, in accordance with the validation status definitions introduced in Section 3.4. Where appropriate, qualitative engineering interpretation and proxy indicators are used without overstating validation maturity.

This outcome also provides valuable insight for future integration activities. It highlights the importance of explicitly addressing auxiliary subsystem interactions when designing Digital Twin-compliant validation environments and directly informs the recommendations provided in Section 8 regarding future test architecture and system-level readiness.

6. Multi-Concern Failure Assessment

The multi-concern failure assessment evaluates whether the toolchain can represent and interpret failure mechanisms across electrical, and thermal domains. Battery system failure rarely remains confined to one domain. An electrical fault may create local heat generation, thermal gradients may accelerate ageing or trigger safety risks. The purpose of this section is to organise testing evidence and toolchain outputs into a structured failure assessment.

6.1 Electrical Failure Modes

Electrical failure modes include abnormal voltage, current and power conditions that may compromise battery safety or accelerate degradation. Typical electrical failure scenarios relevant to D4.4 include overcharge, overdischarge, current overload, voltage imbalance and loss of electrical isolation. The objective of validation is to determine whether the toolchain can represent the electrical response of the battery system and identify conditions that lead to unsafe or reliability-critical behaviour.

The electrical failure modes are evaluated by comparing toolchain outputs with experimental results. Relevant outputs may include predicted voltage response, current evolution, heat generation linked to electrical loading, detection of threshold crossing, and severity classification. The comparison should consider both numerical accuracy and event interpretation. For example, if the toolchain correctly predicts the occurrence of a voltage limit violation but not its exact timing, the result may be considered partially validated depending on the agreed KPI tolerance.

6.2 Thermal Runaway and Propagation Risk

Thermal failure assessment focuses on heat generation, heat dissipation, temperature rise, thermal gradients and the risk of thermal runaway or propagation. Thermal behaviour is central to battery safety because excessive temperature can accelerate degradation, reduce performance, increase internal resistance, and in severe cases initiate self-heating or propagation between cells or modules.

The toolchain should be assessed for its ability to reproduce temperature trends under relevant operating conditions, especially low- and high-temperature operation and fast charging/discharging patterns. The validation should also consider whether the toolchain can identify precursor conditions, such as abnormal temperature rise, localised hotspots, deviation from expected thermal response or increased risk under combined electrical and thermal stress.

Experimental data will provide the basis for evaluating thermal performance. RSTER will interpret these results and will be used to calibrate model predictions and functional safety limits. Where thermal runaway is not directly tested due to safety constraints, proxy indicators such as temperature rise rate, maximum temperature, heat generation tendency and threshold proximity can be used to support qualitative or partial validation.

6.3 Combined Multi-Physics Stress Scenarios

Combined multi-physics scenarios represent the most important validation challenge for the FASTEST toolchain because they reflect the coupled nature of real battery system behaviour. Examples include fast charging at low temperature, high-current operation at elevated temperature, or thermal stress that accelerates ageing and changes electrical response. These scenarios are particularly relevant for demonstrating the added value of a multi-domain toolchain compared with single-domain models.

The toolchain will be assessed on its ability to capture the correct coupling logic between domains. The validation should examine whether electrical loading produces appropriate thermal response, whether thermal conditions influence ageing or resistance behaviour, and whether mechanical stress modifies safety or reliability indicators. The evaluation may use direct numerical comparison where data are available, or structured engineering interpretation where coupled experimental evidence is partial.

7. Toolchain Performance Evaluation

The purpose of the toolchain performance evaluation is to assess whether the FASTEST multi-concern validation framework can provide technically credible and traceable support for battery safety, ageing and reliability assessment under the degradation scenarios defined in D4.4. The evaluation combines available experimental evidence from FM and VTT, AI-assisted estimators developed by ABEE, and the Digital Twin-compliant validation logic defined in Section 3.

The performance evaluation does not only consider numerical agreement between predicted and measured variables. The assessment also evaluates whether the toolchain preserves the correct engineering interpretation of safety-critical conditions, degradation behaviour and operational boundary violations. This distinction is important because the FASTEST toolchain is intended to support hybrid testing and accelerated battery-system development workflows, where correct interpretation of hazardous or degradation-prone conditions is as important as prediction accuracy itself.

The evaluation performed in D4.4 reflects the currently available validation evidence obtained within the project timeframe. Validation maturity therefore differs between scenarios. Electrical validation scenarios achieved the highest level of experimental confirmation, while some thermal and combined electro-thermal scenarios remain partially validated or not assessable due to limitations associated with the prototype module cooling/heating subsystem described in Section 5.

The performance evaluation framework follows the validation logic defined in Section 3.3 and applies the validation-status categories:

- Validated
- Partially validated
- Not validated
- Not assessable

The final validation status assigned to each scenario therefore reflects:

- availability of experimental evidence,
- consistency of toolchain interpretation,
- KPI agreement,
- uncertainty level,
- and completeness of the experimental boundary conditions.

Table 16 summarises the overall validation maturity of the principal scenarios considered within D4.4.

Table 16 Summary of validation maturity

Validation Scenario	Experimental Evidence	Toolchain Capability	Validation Status
Overcharge protection	Available	Correct threshold detection and voltage prediction	Validated
Overdischarge behaviour	Partial / interrupted	Correct interpretation but incomplete module evidence	Partially validated
Current overload protection	Available	Correct safety interpretation and event detection	Validated
Thermal low/high temperature operation	Not completed	Limited direct validation evidence	Not assessable
Fast charging and discharging	Not completed	Partial AI-supported interpretation only	Partially validated
Coupled electro-thermal degradation	Partial	Trend-level agreement only	Partially validated

7.1 Comparison of Simulation vs Experimental Results

The comparison between toolchain outputs and experimental evidence was performed using both quantitative KPIs and qualitative engineering interpretation. The objective was to evaluate whether the FASTEST toolchain could reproduce the relevant electrical and thermal behaviour observed during the validation activities and whether the resulting outputs remained technically meaningful from a safety and reliability perspective.

For electrical validation activities, the strongest validation evidence was obtained from:

- overcharge protection tests,
- current-overload tests,
- and selected cell-level abuse-condition datasets.

The experimental evidence confirmed that the Battery Management System (BMS) correctly disengaged the relays before unsafe voltage thresholds were exceeded. For the Gen3 battery module, relay opening occurred at approximately 3.66 V maximum cell voltage during overcharge testing, while the Gen4 module disengaged at approximately 4.22 V maximum cell voltage. These observations are consistent with the intended safety logic and support the validity of the corresponding toolchain safety-boundary interpretation.

The AI-assisted estimators developed by ABEE demonstrated consistent voltage-prediction trends under overcharge conditions across multiple State-of-Charge (SOC) levels. Figures 1–4 presented in Section 4 demonstrate that the implemented estimators were capable of reproducing:

- voltage evolution,
- thermal behaviour,
- capacity degradation trends,
- and internal-resistance growth tendencies.

The obtained RMSE and MAE values indicate acceptable agreement for the currently available datasets and demonstrate the feasibility of integrating AI-assisted prediction workflows within the FASTEST validation framework.

However, direct module-level validation remained partially incomplete due to the cooling/heating subsystem interaction identified during overdischarge testing. The observed interaction prevented stable electrical isolation after relay disengagement and resulted in residual current draw outside the intended safe voltage boundaries. Consequently, some validation scenarios could not be completed under controlled and reproducible conditions.

For these scenarios, the validation framework relies on:

- partial experimental evidence,
- lower-severity proxy conditions,
- cell-level datasets,

- and engineering interpretation consistent with the Digital Twin-compliant validation methodology.

The comparison therefore demonstrates that the FASTEST toolchain is capable of reproducing the general behaviour and safety interpretation of several critical electrical scenarios, while also clearly identifying the current limitations of the available module-level validation evidence.

7.2 Accuracy of Safety, Ageing and Reliability Predictions

The accuracy assessment performed within D4.4 evaluates whether the FASTEST multi-concern toolchain can provide technically credible predictions and safety-relevant interpretation for battery-system behaviour under representative operational and degradation scenarios. The evaluation considers not only numerical agreement between predicted and experimentally observed variables, but also the ability of the toolchain to preserve the correct engineering interpretation of abnormal operating conditions, degradation evolution and functional safety boundaries.

The assessment combines:

- quantitative KPI-based evaluation,
- AI-assisted prediction performance,
- and engineering interpretation of experimentally observed behaviour.

Particular attention is given to the capability of the toolchain to correctly identify:

- unsafe electrical operating conditions,
- degradation-related parameter evolution,
- thermal-risk indicators,
- and reliability-relevant abnormal behaviour.

The strongest level of validation maturity was achieved for electrical safety-related scenarios, particularly:

- overcharge protection behaviour,
- current-overload protection,
- and voltage-boundary interpretation.

The experimental validation activities performed by FM demonstrated that the Battery Management System (BMS) correctly disengaged the battery modules before exceeding the prescribed overvoltage safety thresholds for both Gen3 and Gen4 battery technologies. The corresponding toolchain outputs showed consistent interpretation of these electrical safety boundaries and correctly associated elevated voltage conditions with increased degradation and thermal-risk indicators. These results support the validity of the electrical safety interpretation logic implemented within the FASTEST validation framework.

The AI-assisted estimators developed by ABEE also demonstrated stable prediction capability across multiple battery technologies and State-of-Charge (SOC) levels. The implemented estimators were capable of reproducing:

- voltage evolution,
- temperature behaviour,
- capacity degradation trends,
- and internal-resistance growth tendencies

under representative abuse and stress conditions.

The obtained RMSE and MAE values reported in Section 4 indicate acceptable agreement between AI-generated predictions and the available validation datasets. In particular, the voltage and thermal prediction estimators demonstrated good trend consistency under overcharge scenarios, while the ageing-oriented estimators were capable of reproducing the expected increase in internal resistance and capacity degradation under stress exposure.

For Extreme Cold Condition (XCC) operation at approximately 0 °C, the estimators demonstrated stable prediction performance for both:

- capacity estimation,
- and internal-resistance prediction.

The relatively low prediction-error values obtained for these scenarios suggest that the implemented AI framework is capable of supporting preliminary degradation and operational-boundary assessment under reduced-temperature operating conditions.

From a reliability perspective, the validation assessment focused primarily on whether the toolchain correctly interpreted experimentally observed abnormal conditions and assigned increased safety or degradation relevance to those scenarios. The toolchain demonstrated acceptable capability in identifying:

- abnormal voltage behaviour,
- elevated thermal-risk conditions,
- and increased degradation tendency under high-stress operational scenarios.

However, the achieved validation maturity differs between validation domains and operating scenarios. While the electrical validation activities achieved comparatively strong experimental confirmation, the thermal and coupled electro-thermal validation activities remain partially validated due to limitations encountered during module-level testing. As described in Section 5, interactions associated with the auxiliary cooling/heating subsystem prevented completion of several planned validation activities requiring stable electrical isolation following relay disengagement.

Consequently, some validation scenarios currently rely on:

- partial experimental evidence,
- AI-supported estimation trends,
- lower-severity proxy conditions,
- and engineering interpretation consistent with the Digital Twin-compliant validation methodology.

The present validation results should therefore be interpreted as technically credible preliminary validation evidence rather than certification-level validation. Nevertheless, the available evidence demonstrates that the FASTEST toolchain is capable of supporting:

- safety-oriented engineering interpretation,
- trend-level degradation assessment,
- Digital Twin-compatible hybrid testing workflows,
- and structured evaluation of battery-system operational boundaries.

The results also demonstrate the practical value of combining:

- experimental validation,
- AI-assisted estimation,
- and Digital Twin-oriented traceability.

within a unified multi-concern validation framework.

7.3 Sensitivity and Uncertainty Analysis

Sensitivity and uncertainty analysis were performed to evaluate the robustness of the FASTEST validation framework under varying operating conditions, model assumptions and experimental limitations. Because the FASTEST toolchain combines:

- experimental measurements,
- AI-assisted estimators,
- and multi-domain validation logic,

multiple sources of uncertainty must be considered during interpretation of the results.

The principal uncertainty sources identified in D4.4 include:

- sensor measurement uncertainty,
- temperature-control instability,
- SOC initialization uncertainty,
- variability between battery generations,
- model-training limitations,
- incomplete module-level datasets,
- and mismatch between experimental and simulated boundary conditions.

Table 17 summarises the principal uncertainty sources considered in D4.4.

Table 17 Sources of uncertainty and impact

Uncertainty Source	Affected Validation Area	Expected Impact
Temperature-control instability	Thermal validation	Reduced reproducibility of thermal response
Initial SOC uncertainty	Voltage and degradation prediction	Shift in predicted operating boundaries
Dataset limitation	AI estimator performance	Reduced generalization capability
Experimental interruption	Module-level validation	Partial validation status
Sensor uncertainty	KPI evaluation	Numerical deviation in comparison metrics

The uncertainty analysis also demonstrated that the reliability of the validation results depends strongly on the availability of stable experimental boundary conditions. This was particularly evident during module-level overdischarge validation, where auxiliary thermal-conditioning subsystem interactions influenced the electrical state of the battery module after relay disengagement.

This observation highlights an important system-level consideration for future hybrid testing platforms and Digital Twin validation environments: auxiliary subsystem interactions may significantly influence validation integrity even when the primary safety-control logic operates correctly.

The uncertainty evaluation therefore reinforces the importance of:

- traceable validation workflows,
- documented validation boundaries,
- and explicit limitation statements when interpreting validation maturity.

7.4 Compliance with Functional Safety Requirements

The functional safety evaluation assessed whether the FASTEST toolchain correctly interpreted safety-relevant operating conditions and identified transitions toward unsafe behaviour. The assessment focused on:

- voltage safety limits,
- overcurrent protection,
- abnormal electrical behaviour,
- thermal risk indicators,
- and operational-boundary violations.

The experimental validation results demonstrated that the Battery Management System successfully disengaged the battery modules before exceeding the defined overcharge safety thresholds. This behaviour is consistent with the intended protection logic and supports the validity of the toolchain interpretation of electrical safety boundaries.

The toolchain also demonstrated acceptable capability in:

- identifying abnormal operating conditions,
- interpreting degradation-related trends,
- and preserving the qualitative safety interpretation associated with the observed experimental behaviour.

For scenarios where complete experimental evidence was unavailable, compliance assessment relied on:

- partial validation evidence,
- AI-supported trend estimation,
- engineering interpretation,
- and lower-severity proxy conditions.

The validation framework therefore distinguishes between:

- demonstrated functional safety behaviour, and
- safety behaviour that remains only partially validated due to incomplete evidence.

This distinction is important to avoid overstating the maturity of the current validation activities while maintaining consistency with the FASTEST hybrid testing philosophy.

7.5 Limitations and Boundary Conditions

The validation activities documented in D4.4 were affected by several technical and experimental limitations that influenced the achievable validation scope and maturity level.

The principal limitation was associated with the interaction between the battery-module thermal-conditioning subsystem and the electrical boundary conditions during module-level overdischarge testing. Although the Battery Management System correctly disengaged the primary relays, residual power draw from the auxiliary cooling/heating subsystem continued to discharge the battery module beyond the intended safe operating window.

As a result, several planned validation activities could not be completed safely or reproducibly, including:

- low- and high-temperature validation tests,
- fast charging/discharging validation,
- extended electro-thermal data-acquisition campaigns,
- and some coupled multi-domain validation scenarios.

Additional limitations include:

- incomplete thermal propagation evidence,
- unavailable short-circuit validation datasets,
- limited module-level ageing evidence,
- dependence on partially synthetic AI-training datasets,

- and restricted availability of repeated experimental campaigns within the project timeframe.

These limitations do not invalidate the FASTEST validation methodology itself. Instead, they define the currently demonstrated validation boundary of the toolchain and provide important guidance for future system-level hybrid testing activities.

The validation framework adopted in D4.4 therefore intentionally distinguishes between:

- validated,
- partially validated,
- and not assessable scenarios,

in order to preserve transparency and technical credibility.

The observed subsystem interactions also provide valuable engineering insight for future Digital Twin and hybrid-testing architectures by demonstrating that auxiliary subsystems may significantly influence integrated validation behaviour under prototype-level operating conditions.

8. Conclusions and Recommendations

This section consolidates the outcomes of the validation and verification strategy developed in D4.4 and interprets the current maturity of the FASTEST Battery AI-powered multi-concern toolchain. The conclusions are based on the Digital Twin-compliant validation framework defined in Section 3, the technology-specific verification inputs provided in Section 4, the experimental validation activities described in Section 5, and the toolchain performance evaluation presented in Section 7.

Overall, D4.4 demonstrates that the FASTEST validation approach provides a structured and technically traceable methodology for assessing safety, ageing and reliability-related behaviour at cell, module and system levels. The deliverable confirms that the developed validation framework can combine experimental evidence, AI-assisted estimators, degradation scenario mapping, KPI-based assessment and engineering interpretation within a coherent hybrid testing methodology.

The strongest level of validation evidence was achieved for electrical safety-related scenarios, particularly overcharge protection and current-overload behaviour. These tests provided direct experimental confirmation of safety-boundary behaviour and demonstrated that the BMS protection logic responded as expected under the tested conditions. The available evidence supports the conclusion that the toolchain can correctly interpret voltage-boundary violations, abnormal electrical behaviour and safety-relevant operating conditions within the scope of the completed tests.

For ageing and degradation assessment, the AI-assisted estimators developed by ABEE demonstrated the capability to reproduce relevant trends for voltage evolution, temperature response, capacity degradation and internal-resistance growth. These results support the feasibility of integrating AI-based prediction workflows into the FASTEST toolchain and show that the methodology can support preliminary degradation interpretation under representative stress and abuse scenarios.

Thermal and coupled electro-thermal validation reached a lower maturity level due to limitations encountered during module-level testing. In particular, interactions associated with the auxiliary cooling/heating subsystem affected the ability to complete several planned module-level tests under stable and reproducible boundary conditions. As a result, some planned thermal performance, fast charging/discharging and extended data-acquisition activities could not be fully completed within the project timeframe.

These limitations do not invalidate the developed validation methodology. Instead, they define the current demonstrated validation boundary of the toolchain. The deliverable therefore adopts a transparent validation-status logic, distinguishing between validated, partially validated and not assessable scenarios. This approach is consistent with good engineering practice and avoids overstating the validation maturity where complete experimental evidence is not available.

The validation activities also highlight the importance of considering auxiliary subsystem interactions when designing future Digital Twin-compliant hybrid testing environments. The observed behaviour demonstrates that even when primary protection logic operates correctly, integrated subsystem interactions may influence the overall validation boundary conditions and affect interpretation of module-level operational behaviour.

From a methodological perspective, D4.4 successfully establishes:

- a Digital Twin-compliant validation framework,
- a structured KPI and validation-status methodology,
- traceable linkage between experimental and virtual evidence,
- and a transparent approach for handling incomplete or partially available validation datasets.

The resulting framework therefore provides a credible basis for continued integration of the FASTEST toolchain into future hybrid testing and Digital Twin workflows while also identifying the areas where additional experimental evidence and extended validation campaigns are still required.

8.1 Recommendations for Toolchain Improvement

Based on the validation activities and limitations identified in D4.4, several recommendations are proposed to improve the future maturity, robustness and applicability of the FASTEST Battery AI-powered toolchain.

First, future validation campaigns should explicitly account for auxiliary subsystem interactions at module and system levels. The overdischarge test showed that auxiliary thermal-conditioning subsystems can influence electrical boundary conditions even after the primary BMS relay-opening event. Future test configurations should therefore ensure that auxiliary loads are either isolated, monitored or explicitly included in the validation model.

Second, the experimental validation matrix should be expanded once revised or industrially mature module configurations become available. Particular attention should be given to thermal performance testing under low- and high-temperature conditions, fast charging/discharging validation, and coupled electro-thermal behaviour. These tests are important for increasing confidence in toolchain performance under operationally relevant boundary conditions.

Third, the AI-assisted estimators should be retrained and revalidated using larger experimental datasets as they become available. The current estimator results demonstrate technical feasibility, but future work should improve model generalisation by incorporating broader cell and module datasets, additional chemistries, repeated test campaigns and wider operating conditions.

Fourth, the toolchain should further strengthen uncertainty quantification. The current validation framework identifies key sources of uncertainty, including SOC initialization, temperature-control instability, sensor uncertainty, model-training limitations and incomplete module-level datasets. Future toolchain versions should include more explicit uncertainty bands around predicted outputs, particularly for temperature, capacity fade and internal resistance evolution. These confidence intervals could be dynamically estimated by analyzing validation residuals from the initial training phase or by conducting repeated model runs via ensemble methods to quantify prediction variance. Implementing these bounds will provide a robust measure of model confidence, which is vital for safety-critical battery management decisions.

Fifth, the Digital Twin integration workflow should preserve the validation-status information generated in D4.4. Each model output should be traceable not only to its input dataset and model version, but also to its validation maturity. This will allow future users of the FASTEST Digital Twin to distinguish between fully validated predictions, partially validated predictions and predictions outside the current validation boundary.

Finally, the validation methodology should be maintained as a living framework. As additional tests, improved models and updated module designs become available, the validation-status matrix should be updated accordingly. This approach supports the FASTEST vision of a hybrid testing platform that evolves through continuous interaction between physical testing, virtual modelling and Digital Twin-enabled data management.

8.2 Readiness Level and Future Integration

The results of D4.4 indicate that the FASTEST Battery AI-powered toolchain has reached a meaningful level of methodological readiness for integration into the broader FASTEST hybrid testing and Digital Twin framework. The validation methodology is mature, traceable and consistent with the objectives of WP4. It provides a structured approach for linking degradation scenarios, experimental evidence, AI-assisted predictions, KPI-based evaluation and validation-status classification.

From a toolchain-readiness perspective, the electrical safety validation components show the highest maturity. The completed overcharge and current-overload validation activities demonstrate that the toolchain can support safety-boundary interpretation under controlled electrical abuse conditions. These results provide a strong basis for future integration of electrical safety assessment into Digital Twin-supported validation workflows.

The ageing and reliability components show promising but still developing maturity. The AI-assisted estimators provide useful trend-level prediction capability for capacity degradation and internal-resistance growth, particularly under available cell-level and abuse-condition datasets. However, further validation using larger experimental datasets and repeated campaigns will be required before these models can be considered fully mature for broad system-level deployment.

The thermal and coupled electro-thermal validation components require further evidence before full readiness can be claimed. The inability to complete some module-level thermal and fast-charging validation tests means that the current toolchain boundary should be clearly documented. These areas remain suitable for further development and future validation once revised module configurations and stable test boundary conditions are available.

For future integration into the FASTEST Digital Twin and hybrid testing platform, the following elements are considered ready or partially ready as given in Table 18:

Table 18 Readiness assessment for future Digital Twin integration

Integration Element	Current Readiness	Comment
Validation methodology	Ready	Structured, traceable and aligned with WP4 objectives
KPI framework	Ready	RMSE, MAE, threshold detection and uncertainty logic defined
Electrical safety interpretation	Ready / high maturity	Supported by completed overcharge and current-overload tests
AI-assisted estimation workflow	Partially ready	Demonstrated with available datasets; requires expanded validation
Thermal validation workflow	Partially ready	Methodology defined, but experimental evidence incomplete
Coupled electro-thermal validation	Partially ready	Concept established, but further evidence required
DT traceability logic	Ready	Validation case, model version, input data and output status can be mapped

In conclusion, D4.4 demonstrates that the FASTEST validation framework is technically sound and suitable for continued integration into the project-level Digital Twin and hybrid testing platform. The deliverable does not claim complete validation of all intended physical scenarios. Instead, it provides a transparent and evidence-based assessment of what has been validated, what has been partially validated, and where further validation is required. This balanced approach strengthens the credibility of the FASTEST toolchain and provides a clear pathway for future refinement, integration and exploitation.