

Article

An Intelligent Cyber–Physical Framework for Electric Bus Batteries: Combining DQN Smart Charging with LSTM Digital Twins

Dragan Aleksendrić ^{1,*}, Veljko Vučinić ^{2,*} , Luca Seidel ²  and Eric Sax ² 

¹ Automotive Department, Faculty of Mechanical Engineering, University of Belgrade, Kraljice Marije 16, 11120 Belgrade, Serbia

² Institute for Information Processing Technologies, Karlsruhe Institute of Technology, Engesserstr. 5, 76131 Karlsruhe, Germany; luca.seidel@kit.edu (L.S.); eric.sax@kit.edu (E.S.)

* Correspondence: daleksendric@mas.bg.ac.rs (D.A.); veljko.vucinic@kit.edu (V.V.)

Highlights

What are the main findings?

- An LSTM-based digital twin accurately captures nonlinear electro-thermal battery dynamics.
- A look-ahead BMS utilizes a predictive thermal stress index for proactive cooling.

What are the implications of the main findings?

- The framework ensures robust cell-to-cell imbalance estimation and multi-step RUL projection.
- DQN smart charging clips high-rate current spikes to prevent localized electrode degradation.

Abstract

The operational reliability of electric buses depends heavily on the thermal stability and longevity of their battery packs. In urban public transit, highly dynamic driving cycles characterized by severe current surges during regenerative braking and prolonged constant-current charging phases induce massive localized thermal stress. Traditional battery management systems often fail to dynamically balance energy recovery with the physical constraints of cell matrices, leading to accelerated degradation and potential safety hazards. This paper proposes an integrated cyber–physical framework that merges intelligent control with advanced prognostic monitoring and predictive thermal management. The core of the control layer is a deep Q-network (DQN) agent designed for real-time smart charging and dynamic regenerative braking current limitation (smart derating). This control loop is tightly integrated with a long short-term memory (LSTM) network-based digital twin (DT) operating at the edge level to mirror complex electro-thermal dynamics and track the battery state of health (SOH). To ensure transparency and proactive thermal control, a nonlinear model predictive control (NMPC) strategy is implemented, utilizing SHapley Additive exPlanations (SHAP) to map core heat generation and dynamically adjust for internal thermal lag. The predictive NMPC-xAI loop maintained cell temperatures within a tight, optimal envelope, preventing delayed under-cooling and reducing peak temperature fluctuations.

Keywords: cyber–physical system; electric buses; digital twins; DQN; xAI



Academic Editor: Pierluigi Siano

Received: 5 August 2026

Revised: 20 October 2026

Accepted: 21 October 2026

Published: 23 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The electrification of urban public transportation is a cornerstone for the realization of smart, sustainable, and zero-emission cities. The global transition toward sustainable transportation systems depends heavily on the performance, reliability, and lifespan of lithium-ion batteries in electric vehicles [1]. Rapid and frequent variations in torque demands due to sudden acceleration and braking drastically cause peak current discharge and charge spikes within the battery. In addition, different driving styles and frequent stop-and-go patterns, substantial and fluctuating passenger loads, and aggressive regenerative braking of these vehicles create stress accumulation and electrochemical degradation within the battery pack. Regenerative braking is a crucial energy-saving technique in urban transit, but the fast response of the electric motor and braking pedal in high-mass vehicles generates high-rate charging currents. The intensity of the charging–discharging rate and the operating temperature of lithium-ion batteries make them vulnerable to failure, making the rate of current delivered to the battery by regenerative braking a serious concern [2]. This introduces severe thermal instability, particularly if the cells are operating at a high state of charge (SoC). The interactions between thermal, electrochemical, materials, and structural characteristics of batteries further complicate the challenges, but they also enable opportunities for developing innovative strategies of thermal management [3]. The self-heating of the battery triggers a series of exothermic reactions, which sequentially occur with the increase in temperature [4]. The internal parasitic chemical reactions within lithium-ion cells shorten service life and speed up capacity fade. To prevent thermal runaway, it is crucial to understand the heat generation mechanisms and design mitigation strategies that can efficiently dissipate the accumulated heat during operations. Battery systems need to maintain a uniform temperature across all battery modules. It is critical to ensure balanced aging, safety, and optimal performance. The temperature of electrochemical cells within battery systems is influenced by a combination of internal and external factors. Understanding these factors is crucial for optimizing battery performance, safety, and lifetime in electric buses, which operate under demanding conditions [3]. Irreversible heat is primarily generated due to internal resistance during charge and discharge processes. This phenomenon, known as the Joule effect, occurs when electrical current flows through resistive elements in the battery, such as the electrodes, current collectors, electrolyte, and interfaces [5]. Due to this, heavy-duty battery systems struggle to maintain uniform temperature distribution across all cells. This lack of thermal balance directly limits the efficiency of regenerative braking and triggers severe localized, nonlinear aging of specific battery modules.

Managing these localized thermal stresses represents a paramount challenge to the thermal stability, overall safety, and long-term state of health (SOH) of lithium-ion battery systems in public transit networks. To mitigate these severe thermal and electrical stresses, conventional battery management systems (BMSs) rely heavily on rule-based control strategies or standard proportional–integral–derivative (PID) control loops [6]. In commercial deployments, traditional approaches manage thermal stress by initiating pre-defined, conservative power derating protocols only after specific, static temperature thresholds have already been breached [7]. That is why advanced methodologies to simultaneously optimize lithium-ion battery longevity, electrochemical stability, and thermal safety under heavy-duty operational profiles are needed. Traditional physics-based models such as the electrochemical pseudo-two-dimensional or empirical Equivalent Circuit Models frequently struggle to preserve accuracy across the entire multi-variable operational envelope of commercial fleets, primarily due to intense computational complexity and the extreme nonlinearity of cell degradation dynamics under rapid transient states [8].

To circumvent these limitations, an AI-driven control loop is proposed in this paper. It integrates an LSTM network-based digital twin, DQN and proactive thermal control. Data-driven methodologies, specifically deep recurrent architectures, have demonstrated the capability to capture high-frequency electro-thermal transients directly from standard on-board telemetry. That is why LSTM network architectures were used in the paper. The fundamental architecture of the LSTM governed by specialized input, forget, and output gating loops allows the model to seamlessly propagate hidden states across thousands of sequential steps. It effectively mitigates the pervasive vanishing and exploding gradient challenges that plague conventional feedforward and standard recurrent neural networks (RNNs) when processing long, uninterrupted time-series data profiles [9].

Deep reinforcement learning (DRL) is seen as a flexible control paradigm for active vehicle energy management, smart terminal fast-charging, and dynamic powertrain modulation [10–12]. DRL controllers synthesize optimal control policies through continuous stochastic exploration of the operational environment [13]. Deep Q-networks have been widely adopted to control complex charging schedules, balance immediate passenger climate demands with grid constraints, and modulate real-time current instructions at the power electronics interface. The DQN agent can dynamically discover an optimal control strategy that balances vehicle transit efficiency with stringent electro-thermal constraints. It was used for smart charging to clip high-rate current spikes to prevent localized electrode degradation. Using robust trajectories of LSTM models in the diagnostic space and DQN algorithms in the control space, these two distinct methodologies continue to exist as parallel, isolated ecosystems within modern vehicle architectures, presenting significant integration bottlenecks under aggressive municipal public transit conditions.

The paper is structured to analyze related works in Section 2. Section 3 aims to explain the methodology and system architecture applied in the paper. The details about the experiment are shown in Section 4, while the development of the LSTM model for battery SOH is detailed in the Section 5. Section 6 is related to RUL prediction analysis. Section 7 explains the DQN model and agent control operations. Section 8 explains predictive chiller cooling system capabilities. Edge computing and data latency are analyzed in Section 9. Conclusion remarks are given in Section 10.

2. Related Works

The increasing complexity of lithium-ion battery systems in modern electric vehicles (EVs) has introduced multi-faceted challenges encompassing thermal safety, operational reliability, lifecycle degradation, and economic sustainability [14]. To address these challenges, the convergence of high-fidelity digital twins (DTs), advanced machine learning (ML), and intelligent Control Theory has emerged as a dominant paradigm for next-generation BMSs.

2.1. Digital Twin Architectures and State-of-X Estimation

Modern cloud–edge computing frameworks have enabled the transition of BMSs from localized reactive controllers to centralized telemetry-driven DTs. Digital twin technology establishes dynamic, bidirectional virtual representations of the physical battery system by continuous sensing of voltage, current, and temperature (VIT data). A core service layer of the battery DT is the precise estimation of State-of-X (SoX) variables, namely SoC, SoH, and State-of-Power (SoP), which serve as critical inputs for real-time protection and diagnostics [15]. For instance, in [16] an online cloud-based DT framework leveraging adaptive H-infinity filtering and particle swarm optimization (PSO) to map cell-level dynamics was implemented. An effective approach for SoH estimation has been proposed in [17], where a battery DT based on long short-term memory is developed and used to virtually discharge the battery to estimate its capacity. To overcome the computational

limitations of conventional electrochemical models, deep learning architectures have gained widespread adoption. LSTM networks have proven exceptionally robust in capturing nonlinear electro-thermal relaxation phenomena, polarization transients, and multi-step-ahead Remaining Useful Life (RUL) projections [18,19]. Based on a comparative analysis of machine learning regressors for RUL prediction using real-world datasets, it is found that hybrid learning approaches are a promising avenue for further refinement [20]. In Ref. [21], a range of machine learning algorithms for SOC and SOH estimation were investigated. Their study found that random forest regression consistently outperformed traditional models, including linear regression, support vector machines, and gradient boosting, achieving R^2 values of 0.9999 and low error metrics (MAE: 0.0035, RMSE: 0.0097). Hybrid machine learning setups, such as those combining random forests and deep neural networks, have achieved low mean absolute errors (MAEs) for concurrent SoC and SoH tracking [22].

2.2. Reinforcement Learning and Thermal Management Control

Beyond passive diagnostics, the battery DT paradigm allows for active, health-aware control. The execution of high-rate regenerative braking current surges and prolonged fast-charging phases in urban transit fleets induces severe localized thermal gradients and degradation mechanisms, such as lithium plating [23]. Conventional control methodologies frequently fail to balance optimal energy harvesting with these strict electrochemical boundaries. To resolve this multi-objective optimization problem, reinforcement learning (RL), specifically deep Q-networks (DQNs), has been deployed to enforce dynamic boundary constraints through smart current-derating policies. Concurrently, the Predictive Index of Thermal Stress within a Model Predictive Control (MPC) framework has been introduced to manage the active thermal management system (TMS). Implementing predictive, look-ahead cooling policies via advanced TMS has demonstrated a potential reduction in overall battery maintenance and degradation costs by up to 62% [24].

2.3. Explainable AI (xAI) in Safety-Critical BMSs

As data-driven models and deep neural networks transition into safety-critical automotive control loops, the lack of transparency in “black-box” models introduces substantial validation risks. Explainable AI (XAI) methodologies, such as SHapley Additive exPlanations (SHAP), have recently been recognized as vital tools to map feature dependencies and provide physical interpretability to ML outputs [25]. However, as highlighted by [25], the integration of advanced xAI tools remains largely confined to offline manufacturing diagnostics or post-mortem performance monitoring. The real-time application of SHAP values to explain and dynamically modify control laws within an active cooling loop represents an open boundary in smart energy management.

2.4. Intelligent Control and Multi-Objective Charging Optimization

Within the battery DT framework, the fast-charging and energy recovery problems are frequently formulated as multi-objective optimization routines aimed at maximizing cell longevity while minimizing charging duration. This requires enforcing strict physical and electrochemical boundaries, such as the maximum permissible charging power, to prevent localized degradation pathways like lithium plating and accelerated capacity fade [15]. To solve these large-scale optimization tasks, deep reinforcement learning algorithms have leveraged cloud-level computational resources to enforce safety boundaries dynamically. For instance, Ref. [26] proposed a multiphysics-constrained fast-charging strategy that achieved a 15% extension in battery lifecycle at equivalent charging speeds by strictly monitoring thermal safety thresholds. To further optimize fluid and airflow architectures, a fuzzy model predictive control framework paired with reciprocating airflow mechanisms

was introduced in [27]. Their system successfully reduced auxiliary energy consumption by 76.4% and minimized temperature non-uniformity from 1.5 °C down to 0.6 °C, validating the impact of combining advanced control theory with thermal dynamics.

2.5. Critical Review of Current Industrial Limitations

Despite these academic advancements, modern industrial BMSs remain heavily bound to reactive control philosophies. Conventional derating strategies throttle the battery current only after a cell temperature or voltage sensor crosses a hazardous threshold. In high-dynamic urban transit operations, this reactive approach is fundamentally flawed. It either fails to prevent localized micro-scale thermal runaway caused by rapid regenerative current surges, or it introduces overly conservative current caps that severely restrict regenerative energy harvesting, thereby lowering the vehicle's net efficiency. Furthermore, while the concept of the battery digital twin is frequently discussed, existing commercial frameworks lack the interactive, low-latency cyber-physical architectures capable of executing simultaneous real-time state mirroring, proactive anomaly verification, and intelligent control adaptation. A profound gap remains in deploying zAI (xAI), federated learning, and online reinforcement learning within real-world application envelopes (such as vehicle-in-the-loop validation) while functioning under strict data constraints.

2.6. Summary of Literature and Identified Research Gaps

Despite the extensive advancements reported in the literature, a critical synthesis of the state of the art reveals three distinct research gaps that motivate this study:

Gap 1: The Disconnect Between Diagnostic Digital Twins and Real-Time Actuation.

Existing frameworks utilize LSTM-based digital twins primarily as isolated diagnostic or virtual sensing platforms (e.g., for tracking RUL or SoH). There is a profound lack of closed-loop cyber-physical architectures where the high-fidelity-state outputs of an edge-level LSTM twin are fed directly into an active reinforcement learning (DQN) control agent to dynamically execute predictive smart derating under dynamic fleet telemetry conditions.

Gap 2: Neglect of Core-to-Surface Thermal Lag in Active Cooling Loops.

Conventional thermal management systems (including PID and standard MPC variants) rely heavily on surface-mounted temperature sensors. The current literature largely ignores the dynamic variation of the internal core-to-surface thermal lag during sudden regenerative current surges. No existing study utilizes an adaptive look-ahead window within a Nonlinear MPC to pre-emptively chill the liquid loop based on anticipated internal heat generation before the thermal wave manifests on the cell surface.

Gap 3: Lack of Real-Time Control Optimization Using xAI Insights.

While frameworks like SHAP are increasingly used to explain why a machine learning regressor predicts a certain thermal or degradation state, they are strictly treated as offline interpretive tools. The utilization of global SHAP rules to distill complex, non-differentiable ensemble models into explicit, real-time control laws for compressor modulation remains entirely unaddressed in contemporary automotive control applications.

2.7. Main Contributions of This Study

High-Fidelity Cyber-Layer Digital Twin: An optimized LSTM sequential neural network model was implemented to function as a real-time virtual sensing platform on the cyber layer. This twin successfully captures nonlinear electro-thermal relaxation trends and polarization transients, providing a robust baseline for cell-to-cell imbalance estimation and multi-step-ahead state forecasting without demanding heavy on-board computational overhead.

Predictive and Adaptive AI Control Loop: A DQN agent equipped with a unique, nonlinear boundary-enforcing reward function was developed for proactive smart derating. Unlike reactive industrial current-clipping, this agent dynamically throttles fast-charging

and high-rate regenerative current spikes (operating within a precise -120 A to -40 A window) based on short-term predictive thermal and electrical stress profiles, seamlessly balancing aggressive energy recovery with active electrochemical preservation.

Explainable Proactive Thermal Management System: A novel approach was implemented by incorporating the SHAP methodology directly into the control loop of a Nonlinear Model Predictive Control (NMPC) framework. By explaining and mapping the nonlinear relationship between internal core heat generation, ambient temperatures, and chiller cooling capacity, the system successfully calculates the dynamic core-to-surface thermal lag. This enables pre-emptive chiller modulation, eliminating the traditional “staircase” sensor response delays, suppressing high-frequency actuator chatter, and maintaining cell temperatures within a strict, optimal operating envelope.

3. Methodology and System Architecture

The proposed architecture operates across two distinct domains: the physical layer (the vehicle and its telemetry) and the cyber layer (the intelligent algorithms), see Figure 1. The centralized telemetry infrastructure was used as an evaluative validation testbed for the proposed solutions. The sequence begins with data preprocessing to filter out instrumentation noise and scale raw parameters into valid physical units. A comprehensive architecture diagram of the CPS framework for electric bus battery management is shown in Figure 1. The diagram is split into two primary layers: the physical layer (electric vehicle—edge) and the cyber layer (cloud/edge server). The physical layer contains the electric bus, its battery pack, sensors (voltage, current, temperature), and the on-board BMS controller. To bridge the gap between raw data acquisition and deep learning state estimation, the telemetry stream is expanded beyond direct sensor measurements. While voltage (V), current (I_{raw}), and battery temperature (T_{bat}) are polled directly from the physical cell sensors, the vehicle velocity (v) is acquired via the onboard fleet management system. Furthermore, an explicit electro-thermal feature representing the instantaneous rate of Joule heat generation (Q_{gen}) is mathematically formulated and extracted at the preprocessing edge. The cyber layer houses the LSTM digital twin and the DQN control agent. It illustrates the continuous closed-loop data flow: sensor measurements are uploaded to the cloud, the LSTM processes them to predict states, the DQN computes optimal control actions, and these instructions are sent back to the physical vehicle to adjust charging and regenerative braking currents. While cloud-based processing is highly effective for long-term prognostic forecasting (e.g., SOH and RUL trend mapping), real-time closed-loop control via the DQN agent demands deterministic, low-latency execution loops at the edge level. High telemetry lag could theoretically delay current-derating actuation, leading to transient thermal boundary breaches before the control policy takes effect. To mitigate this latency constraint, the framework presented in Figure 1 is structured as a dual-layer edge–cloud topology. Under this architecture, the DQN controller operates concurrently across both layers through a master–slave configuration. On the edge level, the local DQN agent executes instantaneous, low-latency control actions (online inference) using fixed neural network weights to ensure real-time responsiveness on the vehicle’s ECU. Simultaneously, on the cyber (cloud) level, the master DQN controller continuously ingests fleet telemetry data to perform computationally intensive policy training and weight optimization. Once a superior control policy is converged upon in the cloud, the updated neural network weight coefficients are transmitted via an over-the-air (OTA) update back to the edge layer, dynamically refreshing the local controller’s parameters without introducing on-board computational overhead.

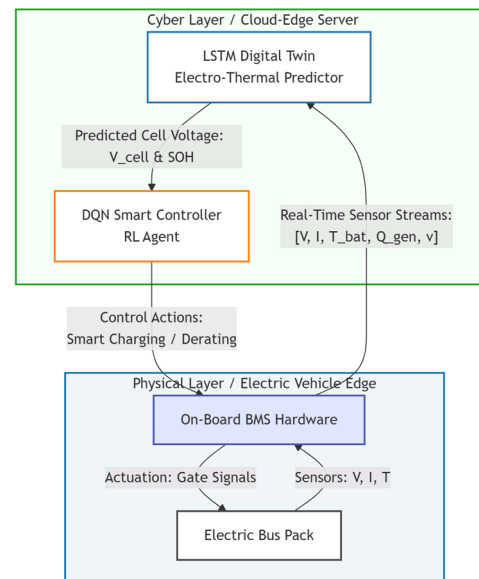


Figure 1. Cyber-physical system for intelligent transport—a general concept.

DQN and LSTM interaction is shown in Figure 2a. This figure illustrates the specific data-level interaction between the LSTM network and the deep Q-network agent. The LSTM network acts as a state estimator, accepting historical sequential time-series data: voltage $V_{(t-k,t)}$, current $I_{(t-k,t)}$, temperature $T_{(t-k,t)}$, vehicle speed $v_{(t-k,t)}$ and instantaneous rate of Joule heat generation $\dot{Q}_{gen(t-k,t)}$, formulated and extracted at the preprocessing edge. The core thermal degradation driver in lithium-ion packs is the Joule heating, governed by the first derivative of thermal energy over time:

$$\frac{dQ}{dt} = \dot{Q}_{gen}(t) = I_{raw}^2(t) R_{in} \quad (1)$$

where $R_{in} = 0.015 \, \Omega$. The output is the cell voltage $\dot{V}_{cell(t+1)}$. This predictive insight, combined with the current SOC_t , SOH_t , heat generation $\dot{Q}_{gen(t-k,t)}$, and ambient temperature forms the complete state space (S_t) for the DQN.

The LSTM is a specialized RNN architecture explicitly designed to model sequential and time-series data with long-term dependencies. It can be used for processing historical sequences of current, voltage, and temperature to predict nonlinear electro-thermal dynamics and SOH evolution [28]. Traditional RNNs suffer from the vanishing or exploding gradient problem, which restricts their ability to learn long sequences [9]. LSTM resolves this challenge by introducing a cell state c_t that acts as an information conveyor belt across time with minimal linear interactions, governed by three specialized control gates [9,29]:

1. **Forget Gate (f_t):** Determines which information from the previous cell state should be discarded based on the current input x_t and the previous hidden state h_{t-1} .

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

2. **Input Gate (i_t):** Decides which new information will be stored in the cell state, while concurrently generating a candidate state vector:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$

$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \quad (5)$$

3. Output Gate (o_t): Dictates the next hidden state h_t , which represents the network's output signal at the current time step:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t \times \tanh(c_t) \quad (7)$$

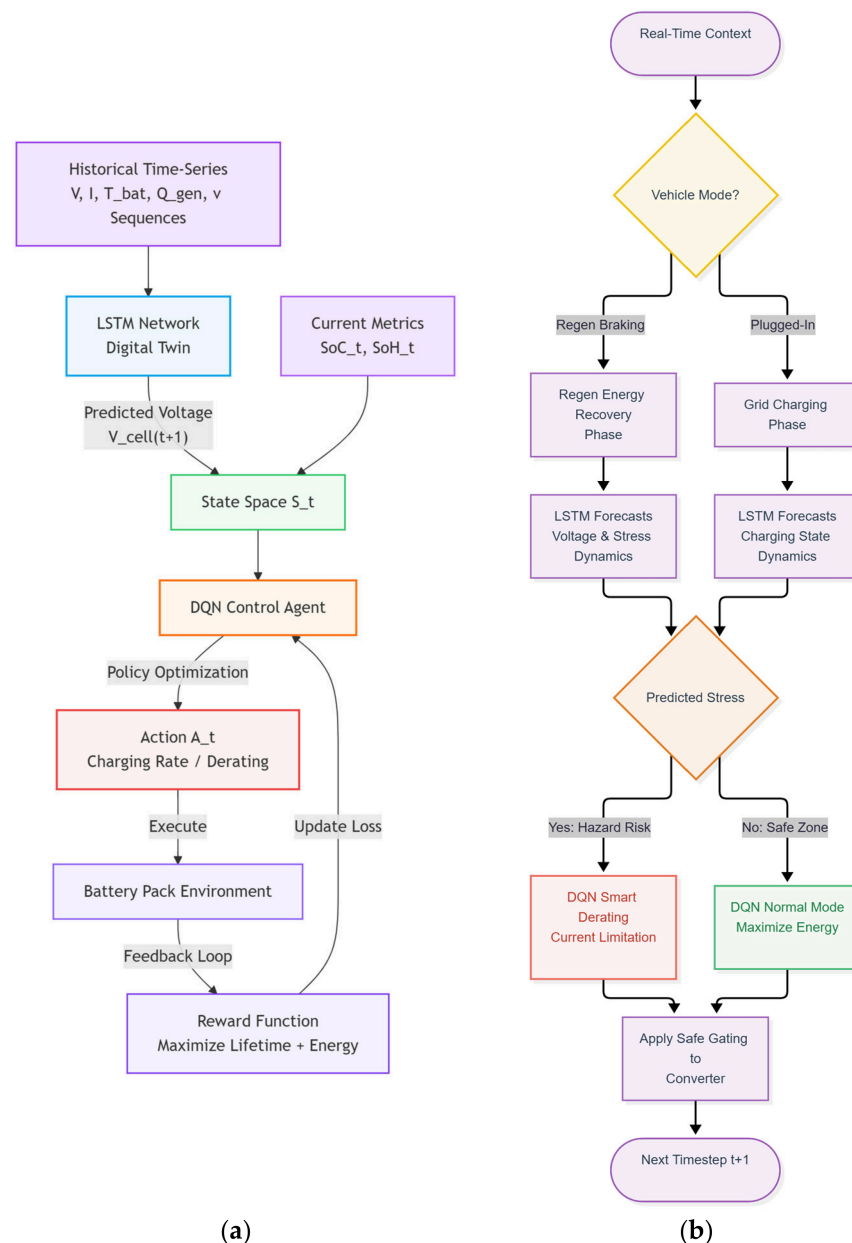


Figure 2. DQN and LSTM interaction block diagram (a), and smart derating and charging concept (b).

To implement the high-fidelity sequential state tracking, the LSTM network topology was structured in this framework as a Deep Multi-Input Single-Output regression framework. This setup is specifically optimized for multivariate electro-thermal time-series mapping where high-dimensional temporal features are projected onto a singular physical

domain. The network architecture consists of the following consecutive layers: 1. Sequence Input Layer: Configured with a fixed dimension of 5, corresponding directly to the expanded time-series feature stream: actual battery current (I_{raw}), battery temperature (T_{bat}), average pack voltage (V), instantaneous Joule heat generation (Q_{gen}), and vehicle velocity (v); 2. LSTM layer: Formulated with 100 hidden units to effectively capture the complex, nonlinear long-term temporal dependencies, polarization effects, and chemical relaxation transients of the lithium-ion cells without suffering from vanishing gradient anomalies; 3. Fully Connected Layer: Configured with an output dimension of strictly 1, which dynamically collapses and maps the high-dimensional hidden state vectors back into a continuous, 1D physical target scale; 4. Regression Output Layer: Executes the final step-by-step prediction of the expected terminal cell voltage $\hat{V}_{\text{cell}}$. This single-output regression facilitates continuous, highly responsive tracking of the residual error for immediate anomaly isolation. To guarantee smooth convergence and eliminate numerical instability during backpropagation, the input variables were normalized using a robust z-score stabilization technique, incorporating a small regularization factor (10^{-2}) to protect the algorithm against constant sensor noise baselines.

The DQN agent evaluates this state, passes it through its Q-network to maximize the customized reward function, and selects an optimal action (A_t), representing either a charging current command or a regenerative braking limit, which is then applied to the battery environment. The dynamic management of massive charging and regenerative current spikes is formulated as a Markov decision process. Traditional battery management systems utilize static lookup tables or rule-based proportional–integral–derivative (PID) controllers, which often over-constrain the system or fail during rapid transient events [30]. To introduce adaptability, a DQN agent is developed to dynamically optimize current limits through a state space and an action space. The state space agent observes the operational environment defined by the vector S_t . The action space domain consists of discrete levels of current restriction commands (smart derating), switching between a highly conservative safety limit (-40 A) and a maximum energy capture envelope (-120 A) [31]. The smart derating and charging concept is shown in Figure 2b.

A workflow and conceptual logic diagram demonstrates a decision-making pipeline that resolves the conflict between maximum energy harvesting and thermal safety. The system continuously polls the LSTM's critical temperature forecasts. If a severe current surge (e.g., from harsh regenerative braking or high-current fast charging) threatens the physical cell constraints, the DQN agent triggers a dynamic derating protocol to cap the current. Otherwise, the agent optimizes for maximum efficiency and regenerative energy recovery, keeping operations within safe electro-thermal boundaries.

4. Experiment Results

4.1. Vehicles Specifications

To evaluate the performance of the proposed cyber–physical framework and the LSTM digital twin prediction performance, a fleet of 10 serial-production solo (12 m) heavy-duty transit buses were deployed in active urban public transportation. The precise structural, electrochemical, and electromechanical parameters of the baseline vehicle configuration are detailed as follows: The vehicle has a curb weight of approximately 13,000 kg with a maximum gross vehicle weight rating of 19,500 kg to account for high-occupancy crush passenger loads during peak transit hours. The energy storage system configuration represents the high-energy battery architecture. The total onboard nominal energy capacity is modelled at 240 kWh (high-power version), consisting of roof-mounted modular lithium-ion cell chemistry packs, designed to endure high-rate rapid charging up to 220 kW. This results in a total of 864 individual battery cells distributed across six roof-mounted packs.

Each pack contains six cell modules (24 cells per module, totaling 36 modules for the entire vehicle). The BMS monitors the thermal state at the module level, activating chilling when any of these 36 modules breaches the temperature threshold. The nominal operational voltage of the integrated battery system is 690 V DC. The baseline vehicle regulates thermal transients via a liquid-glycol circulation loop coupled with an active chiller unit integrated into the central heating, ventilation, and air conditioning (HVAC) system. Under standard factory operating conditions, this cooling system is regulated by a conventional on/off hysteresis controller. The baseline logic initiates active liquid circulation only after the pack's average temperature exceeds 25 °C.

4.2. Experimental Data

To validate the LSTM digital twin and DQN controller action, real vehicle data related to the battery voltage are used (see Figure 3). The typical data related to vehicle speed (v), current (I) and battery temperature (T) are shown in Figure 4 during vehicle operation. In addition, heat generation Q_{gen} (W) was calculated based on current (I) variation; see Figure 5. To resolve the structural time-step misalignment and prevent training instability within the sequential LSTM architecture, the data alignment pipeline was executed prior to model configuration. The telemetry data streams were downsampled to a uniform 1 Hz framework. Figure 3 shows typical measured battery voltage variations over a day for vehicle No. 6. Figure 3 depicts a high-resolution dataset comprising approximately 1.9×10^5 continuous telemetry samples, for only one day, reflecting heavy transient transit stress characterized by consecutive acceleration and intense regenerative braking cycles. The data shown in Figures 4 and 5 were limited to the specified timeframe to provide visible experimental results to understand the complex correlation between the vehicle speed, current variation, measured battery temperature, and calculated generated heat.

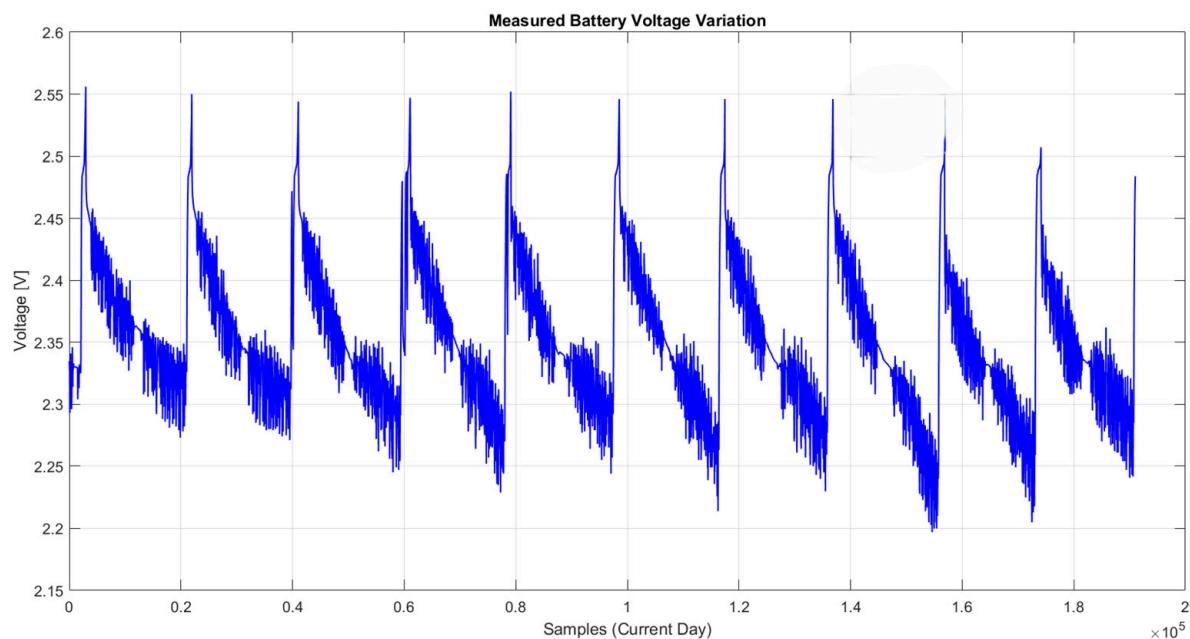


Figure 3. Measured battery voltage variation per day—an example.

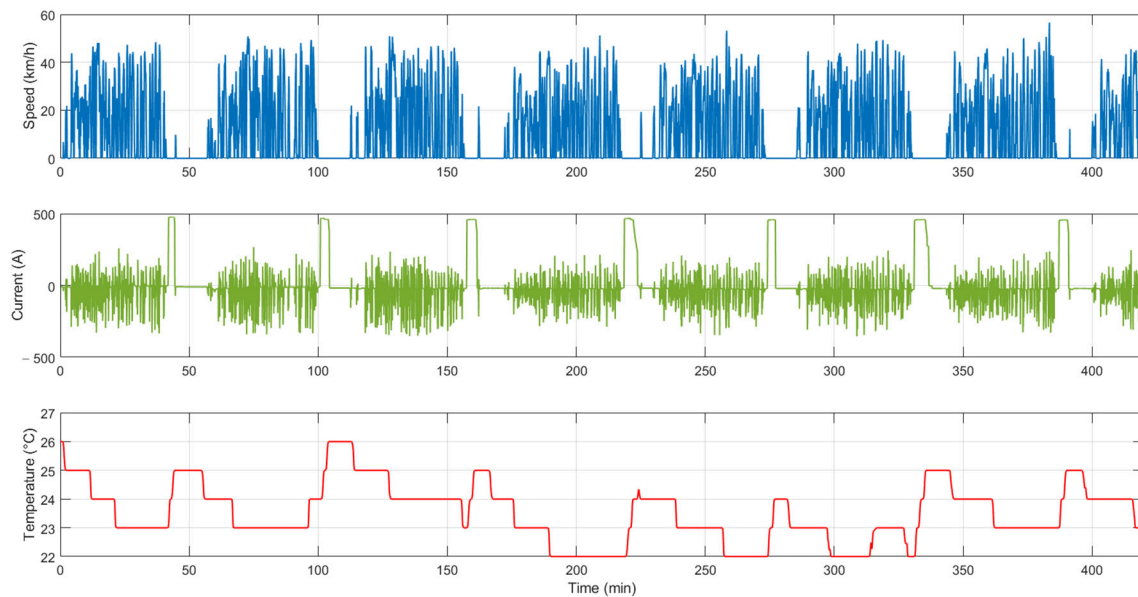


Figure 4. Measured vehicle speed (km/h), current (A), and battery temperature (°C) in the specified timeframe (425 min) of the vehicle operation.

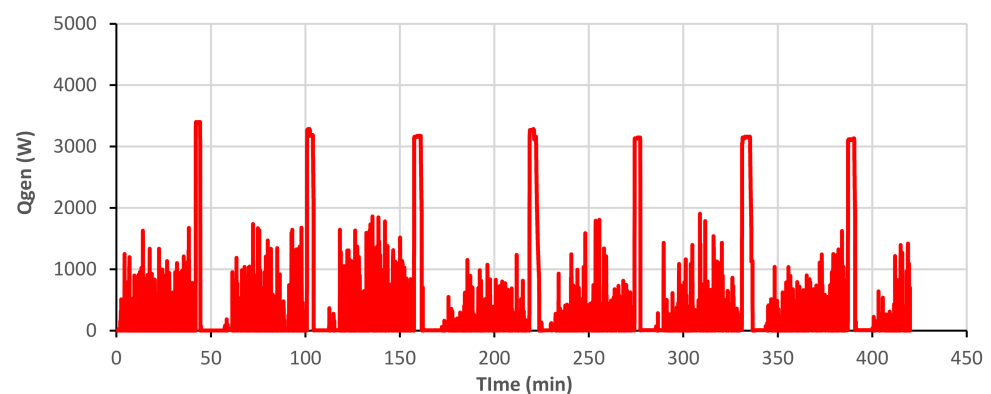


Figure 5. Q_{gen} —calculated battery heat generation versus current (I ; Figure 4—middle panel) variation in the specified timeframe (425 min) of the vehicle No. 6 operation.

5. LSTM Model Development and Validation

While the DQN agent acts as the online decision-maker, real-time validation and prognostic health monitoring are executed by a sequence-to-sequence neural network serving as the vehicle's cyber-layer digital twin. Standard feedforward neural networks inherently lack temporal awareness; conversely, lithium-ion battery states exhibit highly dynamic, sequential memory paths where internal polarization, degradation, and core heat accumulation depend closely on historical duty cycles. To process the multivariate empirical telemetry streams, voltage, current, and temperature (V , I , T), vehicle velocity (v), and calculated internal heat generation (Q_{gen}), whose nature change was shown in Figures 3–5, the LSTM network architecture was deployed. The LSTM's unique hidden-state gating mechanisms effectively mitigate vanishing gradient challenges during the backpropagation-through-time processing of long, highly volatile transit profiles [9]. The LSTM digital twin was structured using a multi-layer deep topology denoted by a 5-20-10-1 architecture. This setup consists of a five-dimensional input layer, a first hidden LSTM layer with 20 units, a second hidden LSTM layer with 10 units, and a single linear output layer optimized for terminal voltage prediction. To ensure full algorithmic reproducibility, the hyperparameter configuration established during the optimization protocol is defined as follows: sequence

length (look-back window) 60 samples (corresponding to a 60 s temporal history at a 1 Hz sampling frequency); optimizer—adaptive moment estimation; initial learning rate—0.001; learning rate schedule—0.2 every 10 epochs; batch size—256 samples; max. epochs—50; regularization—0.0001; early stopping—five epochs. The neural framework was trained using a high-resolution repository spanning 10 operational days of fleet telemetry (averaging 12 driving hours per day). To strictly honor the data partitioning rules established in Section 4, the global dataset was divided into a 70/15/15 split (70% for training, 15% for early-stopping validation, and 15% reserved as an entirely independent, unseen testing set). The model performance over the independent dataset was RMSE = 0.0084 V, and $R^2 = 99.35\%$.

The digital twin dynamically maps these nonlinear electro-thermal transients of the physical pack to predict the expected cell voltage output based on the raw sensor inputs:

$$\hat{V}_{cell}(t+1) = \mathcal{F}_{LSTM}\left(V_{t-k,t}, I_{raw(t-k,t)}, v_{t-k,t}, T_{bat(t-k,t)}, \dot{Q}_{gen(t-k,t)}\right) \quad (8)$$

The fault isolation and anomaly verification are conducted by assessing the instantaneous residual error between the physical sensor streams (V_{cell}) and the digital twin prediction:

$$Error(t+1) = |V_{cell}(t+1) - \hat{V}_{cell}(t+1)| \quad (9)$$

To avoid false positives caused by measurement noise, a dynamic boundary threshold is established. This threshold adaptively scales based on the running mean (μ_{error}) and standard deviation (σ_{error}) of the tracking residual:

$$\gamma(t) = \mu_{error} + k\sigma_{error} \quad (10)$$

where the running mean captures the systematic calibration drift of the digital twin, while the standard deviation reflects the stochastic sensor noise and high-frequency battery dynamics. The multiplication factor (k) acts as a sensitivity tuning parameter, empirically selected as (2.5) to ensure a 98.7% confidence interval under a normal distribution, thereby minimizing false alarms while maintaining high sensitivity to localized cell degradation or sensor anomalies. If the estimated $Error(t)$ remains strictly below this dynamically computed threshold $\gamma(t)$, the state estimation is verified as nominal (free of internal short circuits or cell sensor faults). If the threshold is breached consecutively for a designated time window, an anomaly flag is generated, signaling the DQN agent to immediately execute safety-oriented derating protocols. This verified data stream is subsequently filtered to track short-term SOH variations and evaluate capacity regeneration trends across a multi-day prognostic horizon. To mathematically justify this dynamic threshold boundary ($k = 2.5$), a dual-statistical approach was employed to analyze the estimation residuals ($\mu = 0$, $\sigma = 0.09165$) using a dual approach. First, the Lilliefors test was performed on the complete dataset, yielding a test statistic of $D = 0.015$. This demonstrates that the maximum vertical divergence between the empirical distribution function and theoretical Gaussian cumulative distribution is restricted to a mere 1.5%, confirming a practical fit to a normal distribution for automotive engineering applications. Second, to rule out sample-size distortion, the Lilliefors test was replicated across 10 randomized sub-samples of $N = 1000$ observations each. These sub-samples yielded a mean p -value of 0.22, which safely exceeds the conventional significance threshold (0.05). This statistical verification confirms that the residual distribution does not fundamentally deviate from normality, thereby mathematically anchoring the validity of the chosen confidence intervals and dynamic threshold bounds. The dynamic threshold framework ($\mu + 2.5\sigma$) performs local detrending, ensuring local stationarity and noise whitening by filtering out operational drifts.

To validate the prognostic capabilities of the proposed framework, the outputs of the LSTM-based digital twin were compared directly against unknown high-frequency physical sensor data gathered from the electric buses. A dedicated digital twin instance has been developed for each of the 10 monitored vehicles, since individual battery packs exhibit divergent aging behaviors due to distinct operational routes, passenger load variations, and driver styles. For the purpose of validation, results related to the vehicles defined as No. 6 and No. 5 have been chosen. These vehicles were chosen to represent a different operational conditions regarding the configuration of the travel routes, the number of acceleration/braking/charging events, and different driving styles.

Figure 6 demonstrates the comparison between the physical voltage sensor stream (grey solid line) and the predictive state tracking of the LSTM digital twin (red dashed line). Across the entire multi-step daily profile, the two curves overlap almost perfectly. The model successfully avoids the common pitfall of over-smoothing high-frequency components. Even during sharp voltage drops down to 2.2 V caused by peak traction demands, and rapid voltage spikes up to 2.55 V induced by aggressive regenerative braking current surges, the LSTM network precisely maps the complex internal polarization and electrochemical relaxation dynamics of the pack without any observable phase lag. Figure 6 shows the current voltage profile state of the battery installed in vehicle no. 6. The voltage profile is strongly influenced by the current vehicle's battery state but also the driving style of the vehicle on the specific day and frequent stop-and-go patterns. Every vehicle has its own digital twin and accordingly its own battery voltage profile. To validate this approach regarding the capability of the digital twin to model the complex change of the voltage profile during the vehicle's use, the digital twin's prediction performance was tested over the independent dataset using vehicle No. 5. Figure 7 shows comparison between measured voltage data and the digital twin predictions. It is evident that a completely new voltage profile was well predicted by the digital twin of the battery installed in vehicle No. 5. Obviously, the battery performance differed significantly from vehicle to vehicle (see Figures 6 and 7).

Leveraging continuous state estimation, the framework executes near-real-time fault detection by evaluating the instantaneous voltage residual error. Using this approach, anomalies related to system operation can be detected. Detection of anomalies regarding electric vehicle operation is very important [32]. Throughout the nominal test cycles, this residual tracking error remains close to 0 mV and satisfies the dynamic sensitivity threshold constraint $\gamma(t)$. This near-zero error verifies that the physical battery pack currently operates free of internal short circuits, cell-to-cell connection degradation, or localized sensor faults. To test the digital twin dynamic sensitivity versus threshold constraint $\gamma(t)$, controlled fault injection was conducted offline on the independent test dataset. Three different faults were simulated: (1) Sensor drift fault—a sudden +60 mV step-offset applied to the continuous cell terminal voltage stream; (2) Connection degradation/high resistance—an operational voltage drop proportional to the input current excitation was triggered (>55 mV); (3) Simulation of Internal Short Circuit (ISC)—a micro-short internal cell discharge pathway leading to a localized continuous 52 mV drop below the nominal twin prediction envelope. Based on that, precision, recall and false-alarm rate can be calculated. The evaluation matrix metrics results are shown in Table 1. According to Table 1, the digital twin reacted at an acceptable level versus controlled fault injection.

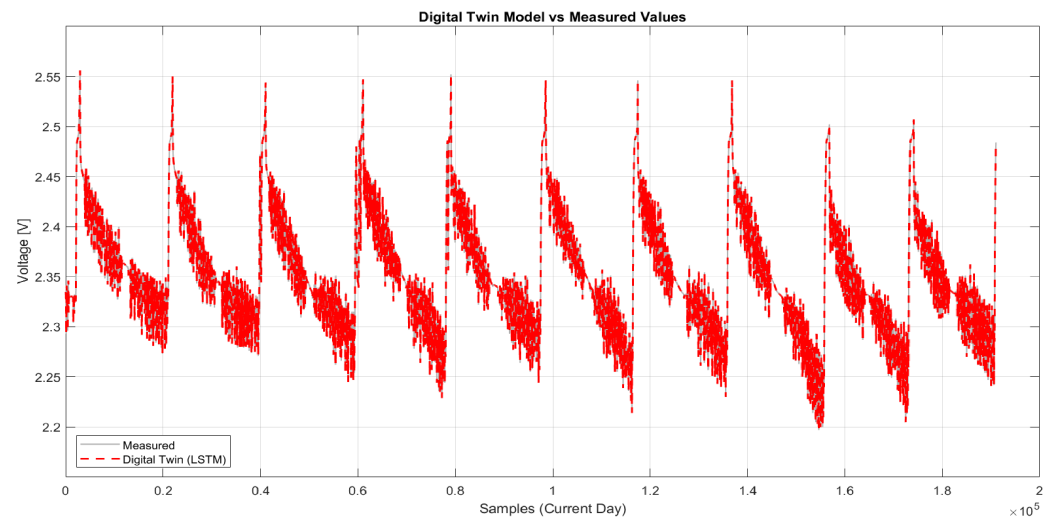


Figure 6. Validation of digital twin model prediction versus measured voltage value—vehicle No. 6, day 16.

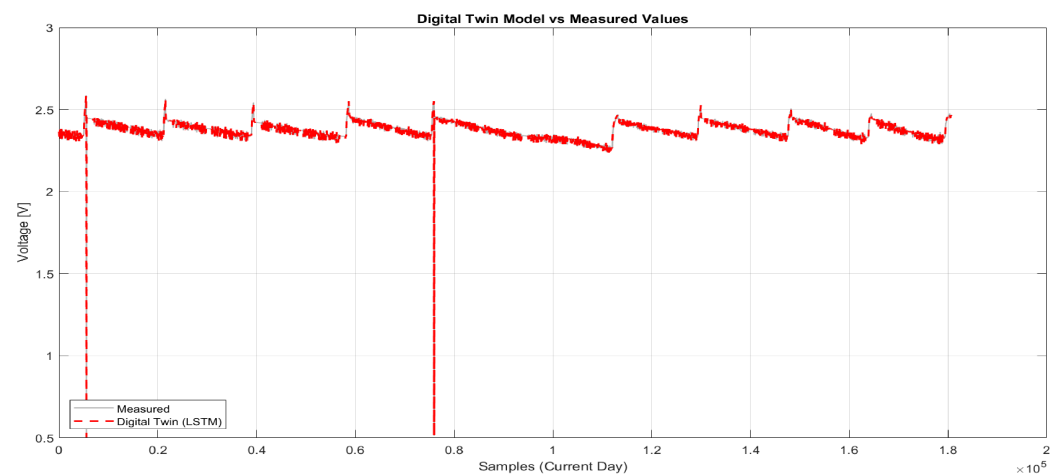


Figure 7. Validation of digital twin model prediction versus measured voltage values—vehicle No. 5, day 15.

Table 1. Evaluation matrix.

Anomaly	Fault	Precision [%]	Recall [%]	False-Alarm Rate [%]	Designated Time Window (s)
Sensor drift	+60 mV	99.1	98.4	0.7	3
Connection degradation	>55 mV	98.2	97.1	1.1	3
ISC	Voltage drop 52 mV	97.8	96.8	1.02	5

The mathematical validity of the 1.24% theoretical false-alarm rate relies on the localized stationarity and independence of the nominal tracking errors. Within this framework, because the sequential LSTM twin successfully captures the battery's temporal polarization memory, the nominal residuals undergo a whitening process, collapsing into independent white noise. Furthermore, by deploying the dynamic moving-window threshold $\gamma(t) = \mu + 2.5\sigma$, the framework filters out systematic operating drifts, keeping the residual variance locally stationary and ensuring a highly robust fault detection specificity.

6. RUL Prediction

The cyber layer integrates a long-term prognostic engine to determine the health evolution of the asset. Figure 8 depicts the predictive capacity fade layer of the framework across a consecutive 16-day operational sample. The blue marker line shows the digital twin SOH estimation over 16 days used in this case. Utilizing this sequence of physical historical data points, the prognostic engine executes a multi-step forward autoregressive projection (represented by the red dashed line) toward the designated End-of-Life (EOL) safety threshold, fixed at 80% SOH. Rather than relying on static electrochemical aging laws, the diagnostic cyber layer utilizes the entire historical SOH data points tracked by the digital twin. To project the health trajectory into the future, the trained LSTM network is deployed in a recursive autoregressive configuration. At time step t , the network evaluates the historical rolling window of degradation parameters to predict the next prospective health state $\widehat{SOH}(t+1)$. This predicted value is then iteratively appended back into the input sequence vector to recursively resolve further forward horizons:

$$\widehat{SOH}(t+m) = F_{LSTM}(\widehat{SOH}(t+m-1), \dots, \widehat{SOH}(t+m-k)) \quad (11)$$

where m represents the forward-looking day index, and k is the depth of the recursive memory window (entire historical operational horizon—16 days in this case). The recursive loop terminates immediately upon hitting the recognized battery EOL safety threshold, defined as 80% of the nominal capacity baseline. The RUL in the next 30 days is subsequently computed as the exact temporal delta between the current operational day and the target threshold intersection:

$$RUL = \min \left\{ m \mid \widehat{SOH}(t+m) \leq SOH_{EOL} \right\} \quad (12)$$

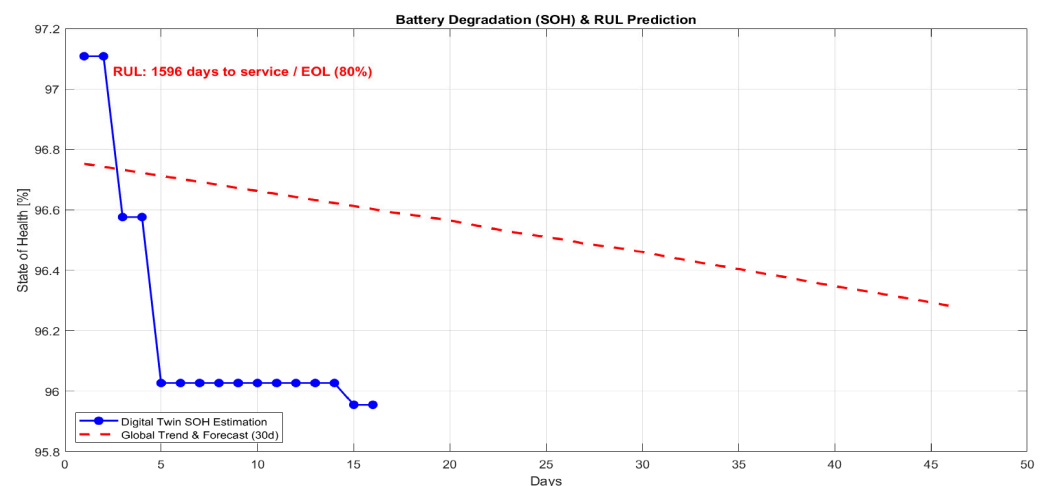


Figure 8. Digital twin SOH estimation and RUL, an indicative macro-trend prediction—vehicle No. 6.

The prognostic architecture generates a RUL global trend and forecast, not only over the short-range sample of 16 days but also by continuously evaluating the battery pack's lifetime trajectory, indicating an expected baseline of 1596 remaining days prior to reaching the 80% EOL capacity threshold. Crucially, the proposed architecture functions as an adaptive moving-horizon rolling estimation framework. The 1596-day output is not presented as an absolute chronological lifetime guarantee, but rather as an instantaneous prognostic velocity vector representing the battery's health trajectory under the specific environmental stress conditions experienced during that 16-day historical window. It must be explicitly stated that predicting a multi-year EOL asset replacement threshold from a 16-day

trace dataset represents an indicative macro-trend rather than a verified chronological certainty. In real-world public transit operations, battery capacity fade is highly nonlinear and sensitive to unobserved future anomalies, cell-to-cell thermal mismatch, and severe seasonal ambient shifts. Therefore, the primary focus of the prognostic engine is locked on high-fidelity short-term prediction (a rolling 30-day operational horizon) where accuracy remains statistically acceptable. To validate the predictive integrity of the prognostic layer, the model's accuracy was quantified across expanding temporal horizons. Because long-term capacity fade under dynamic public transit stress is highly nonlinear, the prognostic engine utilizes a dual-horizon evaluation strategy to prevent speculative overconfidence. For the high-fidelity short-term prediction—operating on a rolling 30-day moving horizon—the model tracks capacity degradation trends with high statistical accuracy. Evaluated against the independent validation set, the short-term RUL forecast achieved a RMSE of 0.42% and a mean absolute error (MAE) of 0.31% (expressed in normalized state of health capacity units). Within this immediate window, the uncertainty envelope remains highly contained, yielding a narrow 95% confidence interval (CI).

However, as the extrapolation horizon extends from the 30-day rolling window toward the projected EOL threshold at 1596 days, the prediction accuracy inherently decays. This decay is a direct consequence of cumulative error propagation within the recurrent neural architecture, coupled with the stochastic nature of unobserved future driving cycles, seasonal ambient temperature shifts, and cell-to-cell thermal mismatches. Rather than presenting the 1596-day output as a static chronological certainty, the framework defines it as an instantaneous prognostic velocity vector representing the battery's health trajectory under the specific environmental stress conditions experienced during that historical window. The reconstruction of the long-term lifetime trajectory is mathematically derived by evaluating the instantaneous prognostic velocity vector, expressed as $(dSOH/dt)$. This localized degradation gradient is extracted from the sequential outputs of the digital twin and projected forward until it intersects the baseline (80%) End-of-Life (EOL) capacity threshold, yielding the estimated 1596 remaining days. This macro-trend functions as a rolling, adaptive moving-horizon estimate that continuously updates with incoming daily telemetry to accommodate nonlinear operational shifts. The SOH capacity is evaluated via integrated Ampere-hour calculations during terminal-station charging sequences, where the digital twin cross-validates the localized capacity fade against the long-term polarization voltage drift.

7. DQN Agent Control Operation

7.1. Predictive Thermal Stress

To provide the proactive mitigation capabilities of the closed-loop cyber-physical framework and the co-dependency between vehicle dynamics and battery temperature gradients, the battery thermal stress was introduced as a basis for DQN control strategies. The cumulative thermal stress induced during harsh stop-and-go public transit cycles and battery charging is formulated as a predictive thermal stress index ($I_{\text{stress}}(t)$). This index evaluates the moving window integral of the product between the instantaneous heat generation rate and the vehicle velocity $v(t)$, expressed in the discrete operational domain as:

$$I_{\text{stress}}(t) = \frac{1}{N} \sum_{k=t-N}^t \left(\frac{I_{\text{raw}}(k)^2 \cdot R_{\text{in}}}{Q_{\text{max}}} \cdot \Phi(k) \right) \quad (13)$$

where $N = 500$ denotes the moving-average window size utilized for edge-level computation. The variable $Q_{\text{max}} = 3500$ W represents the maximum internal heat generation rate, calibrated to match the peak current experienced during ultra-fast charging at the station, while $v_{\text{max}} = 70$ km/h represents the maximum permissible operational speed of

the transit bus. The scaling factor $\Phi(k)$ is a state-dependent kinematic operator introduced to dynamically decouple the velocity influence based on the vehicle's operational state.

$$\Phi(k) = \frac{v(k)}{v_{\max}}, \text{ if } v(k) > 0 \text{ (dynamic driving regime), or } 1, \text{ if } v(k) = 0 \text{ (static driving regime).}$$

By utilizing this formulation, the index simultaneously captures the multi-physical thermal stress under both dynamic and static vehicle conditions. During standard driving and regenerative braking intervals, both the internal Joule heating and the external convective airflow variations (caused by the vehicle speed) are nonlinearly scaled between 0 and 1. Conversely, when the bus is stationary at the charging station, the velocity drops to zero; in this specific operational state, the battery pack transitions from a dynamic kinetic-thermal stress regime to a purely static thermodynamic loading regime. To ensure comprehensive battery thermal management system (BTMS) protection, the control logic automatically enforces $\Phi(k) = 1$, overriding the kinetic component and evaluating the thermal index solely as a function of the localized core heat accumulation ratio $Q_{\text{gen}}/Q_{\max}$.

Figure 9 shows the thermal index change (lower panel—in log scale) versus measured input quantities: the vehicle speed (upper panel—blue solid line), current (I) during driving, braking and charging (upper panel—red dashed line). The battery temperature is also measured and shown (middle panel) to illustrate co-dependence between these important input parameters. It is obvious that during battery charging ($I = 467$ A, for example) at the charging station, the thermal index has the highest value close to 1—which is extremely high for the battery thermal stress index (Figure 9). At a vehicle speed of $v_{\max} = 30$ km/h and $Q_{\text{gen}} = 1500$ W, the thermal index is around 0.18. The thermal stress index was taken as a basis for DQN derating control and later for the battery cooling control.

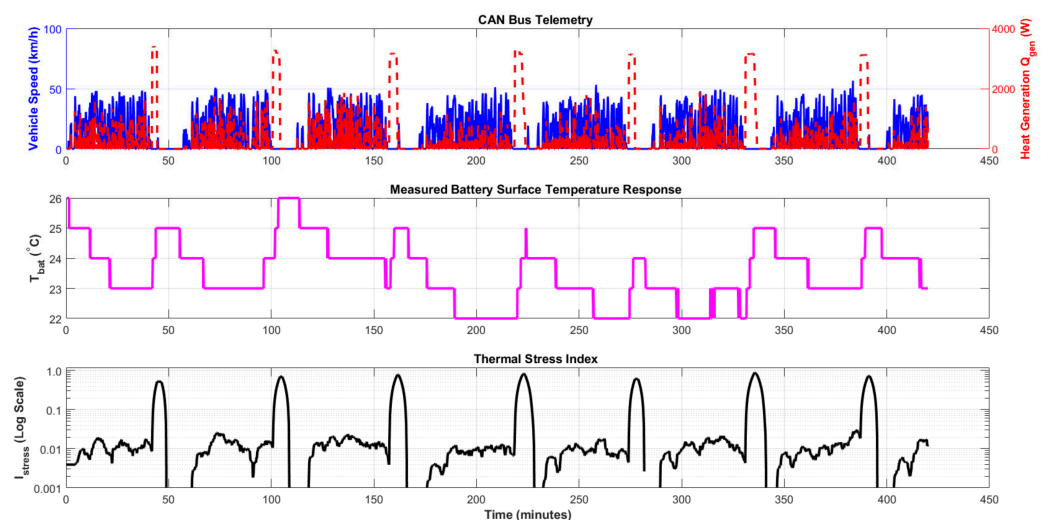


Figure 9. The battery thermal stress calculation (lower panel) vs. measured vehicle (No.6) speed and current (I_{raw}) variation (upper panel). The real values of the battery's temperature are shown in the middle panel.

The DQN controller systematically mitigates severe internal thermal gradients between the electrode matrix and the outer cell casing. This restriction prevents local hot-spot crystallization and guarantees long-term electrochemical capacity preservation. The practical utility of this thermal stress index is manifested in two parallel, intelligent control actions orchestrated by the DQN agent:

- (1) **Intelligent Current Limitation (Smart Derating)**—The stress index indicates that the cumulative thermal pressure is approaching a critical boundary; the DQN agent updates its policy to heavily penalize high-current actions. During subsequent aggressive regenerative braking phases, the agent instantly caps the negative current envelope. It reduces the term at the source, preventing localized hot spots within the

cell electrodes while still harvesting the maximum permissible recuperation energy. Any change of the regenerative braking effect will request change in the friction braking strategy. That is why introduction of intelligent hybrid braking shall rely on a battery management system based on the battery thermal stress. The proposed approach offers broad possibilities for introducing various strategies for intelligent control of hybrid braking, specifically in the context of balancing the regenerative braking effect and the battery pack's thermal footprint. The importance of introducing such intelligent braking mechanisms has been extensively highlighted in [33,34].

- (2) Predictive Cooling Actuation—Instead of waiting for the battery pack to breach the typical factory threshold, the DQN agent reads the rising edge of the stress index. It preemptively activates the liquid-glycol circulation pumps and ramps up the active HVAC chiller unit before the bulk temperature escalates. This proactive cooling creates a thermal buffer within the liquid loop, shifting the thermal management from a reactive hysteresis state to an anticipatory optimization paradigm.

7.2. DQN Reward Function

It is known that traditional systems look only at static temperature, which suffers from severe thermal lag; by the time a physical sensor reads a temperature spike, the internal microstructural degradation has already occurred [11]. The vehicle's speed profile in cities is characterized by a stochastic stop-and-go public transit pattern with frequent rapid acceleration phases and harsh deceleration cycles. Under conventional reactive battery management, these sudden acceleration and deceleration events cause extreme, unmitigated kinetic energy injections via unregulated regenerative braking. This severe transient loading drives the battery temperature upward. To counteract this phenomenon, the thermal stress index serves as the core control foundation for the DQN agent.

The DQN agent's actions are defined by integrated reward function. The DQN agent penalizes high-current actions not only during subsequent aggressive regenerative braking phases but also during charging. To establish how the reinforcement learning agent internalizes thermal and electrochemical safety constraints, the correlation between the mathematical reward architecture and the resulting control actions was evaluated. Figure 9 displays the geometric shape of the boundary-enforcing DQN reward function. The core mechanism (reward function R) forcing the agent to prioritize thermal safety during aggressive braking/charging cycles is a nonlinear, boundary-enforcing reward function. To evaluate the action in state A_t taken in state S_t , the reward R_{t+1} is computed based on the resulting system response at step $t + 1$:

$$R_{t+1} = -\alpha(I_{\text{limit}}(t+1))^2 - \beta(|I_{\text{raw}}(t+1)| - 100)\forall \cdot |I_{\text{raw}}(t+1)| > 100 \text{ A} \quad (14)$$

where $\forall$ is an indicator function that equals 1 if the resulting current magnitude exceeds the 100 A safety threshold, and 0 otherwise. The parameter α (0.015) directly corresponds to the internal resistance R_{in} of the battery, thereby penalizing instantaneous Joule heating. The parameter $\beta = 3.2$ represents a strict, fixed penalty slope that aggressively suppresses currents exceeding the safe operational boundary of 100 A to protect battery health.

Figure 10 shows a general DQN agent strategy to prevent high internal thermal stresses of the battery. The curve illustrates a nonlinear, dual-stage penalization profile:

1. The Optimization Zone (0 A to -100 A): Between 0 A and -100 A, the reward function exhibits a smooth, progressive quadratic slope governed by the α parameter. In this region, the agent faces a minor penalty for standard current flow, which corresponds directly to the battery's internal resistance $R_{\text{in}} = 0.015 \Omega$ and the resulting Joule heating losses. This smooth gradient prevents the agent from aggressively throttling

the current unnecessarily, ensuring the vehicle harvests sufficient regenerative braking energy to maintain the transit schedule.

2. The Boundary-Enforcing Penalty Zone (Amperage > -100 A): The critical operational contribution is the severe, sharp mathematical drop (inflection point) that occurs precisely at the -100 A safety threshold. Once the current magnitude crosses this boundary, the reward immediately drops from approximately -125 points down to -190 points. Beyond this inflection point, the function transitions into an aggressive, steep penalty curve driven by the $\beta = 3.2$ constraint, plunging down to -340 points at a peak current of -150 A. This steep barrier forces the Q-network to heavily dismiss actions that violate the 100 A boundary.

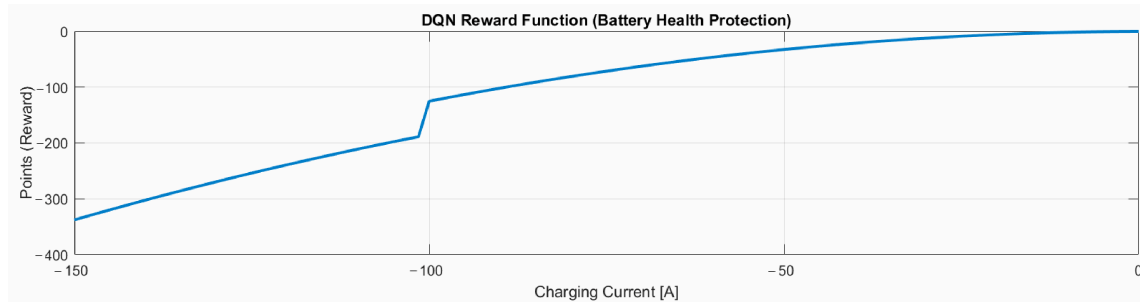


Figure 10. DQN reward function.

The state space vector governing the DQN agent is composed of four continuous variables continuously streamed from the physical and cyber layers: SoH, SoC, Tamb (ambient temperature), and the dynamically computed Predictive Index of Thermal Stress (Istres). The underlying deep Q-network is designed with nine output neurons, where each neuron maps directly to a discrete operational current-derating action state. The action space A is defined as:

$$A = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9\} = \{40, 50, 60, 70, 80, 90, 100, 110, 120\} A$$

These discrete currents dictate the maximum allowable current limits enforced by the control layer, categorized into three operational zones:

Strong derating regime-actions a_1 – a_6 (40–90 A)—This subset corresponds to the thermally safe α zone. The agent aggressively restricts the raw battery input currents within a (40–90 A) window. While electrochemical safety is prioritized, the immediate reward remains small because the restricted current limits prevent the vehicle from fully exploiting its kinetic regenerative braking capacity.

Inflection point—action a_7 (100 A)—This state represents the boundary limit. It permits a maximum current injection of 100 A immediately prior to entering the high-risk, mathematically penalized β zone (see Figure 10).

High-rate energy capture regime-actions a_8 and a_9 (110 A; ≥ 120 A)—The agent unthrottles the current constraints to allow maximum charging throughput or intense regenerative braking capture. This zone is subject to severe mathematical penalty thresholds if executed under elevated battery thermal stress conditions.

The DQN architecture is defined as a multi-layer perceptron with four inputs, two hidden layers (60 and 30 neurons, respectively) and nine outputs. The model was trained using a batch size of 64, target network update of 1000 epochs, and the Adam optimizer. The convergence criterion was bounded by a mean action-value variation $\Delta Q_{\text{mean}} < 0.01$ across 50 consecutive episodes ranging from -5 to -340 points. Replay buffer capacity

1×10^5 , exploration policy ϵ -greedy, total training horizon: 2000 episodes. Action-value approximations are computed via the standard Bellman optimality equation:

$$Q(s, a) = R_t + \gamma \max_{a'} Q(s', a') \quad (15)$$

where $\gamma = 0.95$ represents the temporal discount factor. The policy network evaluates this nonlinear action-value function to select the specific current-clipping action that maximizes the expected cumulative reward for the active state-space vector.

To guarantee fleet resilience across continental climate extremes, the cyber–physical framework must be adaptive to sub-zero winter environments ($<0^\circ\text{C}$) and ultra-high summer settings ($>40^\circ\text{C}$). Operating in these regimes without control adaptations can lead to structural performance degradation, such as premature HVAC thermal saturation in summer or accelerated electrochemical degradation in winter. To prevent this, the integration layer utilizes a dynamic parametric tuning strategy.

Sub-Zero Winter Adaptation Strategy (Lithium Plating Mitigation)—Under sub-zero conditions, the cell's internal resistance (R_{in}) increases nonlinearly, and the solid-state diffusion of lithium ions into the graphite anode slows down dramatically. Executing high-rate regenerative braking surges in this state forces the anode potential below 0 V vs. Li/Li^+ , inducing severe lithium plating and catastrophic capacity fade. To mitigate this risk, the DQN reward function cannot maintain a static inflection point at 100 A. Instead, the framework introduces a temperature-dependent scaling operator for the inflection point (ξ_{switch}). When the ambient temperature drops below 0°C , the control logic automatically shifts the hard penalty inflection threshold down to 50 A. This dynamic adjustment triggers the severe penalty β -zone much earlier, forcing the policy network to execute aggressive, pre-emptive current-clipping. Crucially, this threshold modification is co-dependent on the continuous cell state; if the LSTM digital twin predicts that internal Joule heating (Q_{gen}) has successfully raised the core temperature (T_{core}) back into a safe operational window ($>15^\circ\text{C}$), the DQN dynamically scales the inflection boundary back toward 100 A, seamlessly expanding regenerative energy harvesting capacity in real time.

Under ultra-high summer temperatures, the primary challenge transitions to the thermal saturation of the cooling apparatus, as the coefficient of performance (COP) of the chiller compressor drops significantly. To compensate for the diminished heat rejection capability of the liquid loop, the DQN reward function dynamically increases the magnitude of the β penalty parameter (e.g., from $\beta = 3.2$ to $\beta = 5.5$). By mathematically amplifying the cost of heat generation (Q_{gen}), the agent becomes highly sensitive to thermal accumulation, executing proactive current derating far before the cells reach their critical degradation threshold.

7.3. DQN Control Strategy Validation—Results and Discussion

The trained DQN policy was deployed to dynamically optimize the battery current constraints across two distinct operational extremes: (1) high-rate transient regenerative braking sequences, and (2) stationary ultra-fast charging events at the terminal station.

The baseline dynamic response under highly transient acceleration and deceleration (braking) profiles is illustrated in Figure 11. Figure 11 tracks the empirical vehicle velocity fluctuations (upper panel), the corresponding battery terminal current variations (middle panel, where negative values represent acceleration loads, and positive values signify regenerative braking energy capture), and the resulting Predictive Index of Thermal Stress (lower panel). The DQN agent remains inactive during standard driving phases, allowing the factory traction system to operate unhindered. However, during braking intervals, the DQN actively intervenes to safeguard the pack based on the real-time state-space inputs.

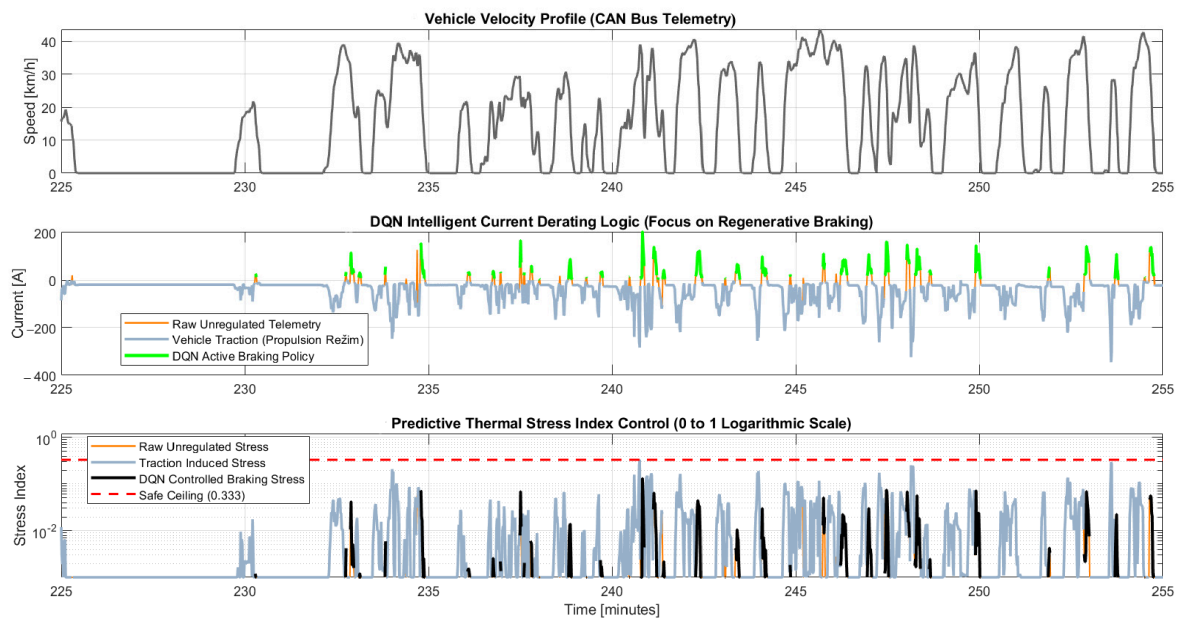


Figure 11. The correlation between the measured vehicle speed (**upper panel**), measured current (**I**—**middle panel**) during driving (negative value—grey line) and braking (positive value—green line), and the calculated battery thermal index stress (**lower panel**, driving stress—grey line; braking stress—black line).

Figure 12 displays the responsive DQN action trajectories relative to these transient inputs. To establish a baseline safety buffer before triggering the heavy mathematical penalties of the β zone (defined in Figure 10), the nominal current threshold was mapped at +100 A, indicated by the horizontal red boundary line in the upper panel of Figure 12. The middle panel tracks the discrete DQN action choices executed in real time, while the lower panel displays the corresponding instantaneous reward allocated to the control loop.



Figure 12. DQN control actions and reward (points) versus the inputs shown in Figure 11 (the vehicle speed, current I , and I_{stress}) under the ambient temperature of 35 °C.

Rather than enforcing a rigid, deterministic current cap for every transient spike exceeding +100 A, the DQN agent dynamically assesses the broader thermodynamic state of the cell matrix. For instance, at approximately $t = 241$ min (Figure 12), the regenerative braking input current surges past +200 A. Instead of aggressively clipping this energy stream, the DQN agent selects action state 9, successfully permitting this high-rate energy harvesting phase to proceed. This intelligent policy selection is fully justified because the underlying Predictive Index of Thermal Stress remains strictly below the safe operating ceiling of 0.33 (Figure 11, lower panel), supported by a robust SOH of 94.6%, an SOC of 75.5%, and nominal ambient temperature of 35 °C. Consequently, this action is rewarded favorably by the optimization function since it maximizes regenerative efficiency without violating electrochemical safety boundaries. This specific event demonstrates that the reinforcement learning policy successfully avoids overly conservative current caps, validating its capacity to enhance the vehicle's net energy efficiency while maintaining strict, nonlinear thermal containment.

To provide explicit temporal context for the dynamic profiles illustrated in Figures 11 and 12, reference is made to the complete operational timeline in Figure 9. Specifically, the 30 min interval from 225 min to 255 min was extracted to isolate and evaluate the tracking performance of the DQN control strategy. The safe threshold ceiling of the thermal stress index, established at a value of 0.33 (indicated by the dashed horizontal line in the lower panel of Figure 11), must be interpreted within a broader multi-physical context. This dynamic boundary is dependent on the instantaneous core temperature of the battery cells, which is driven by the internal heat generation rate Q_{gen} —a function of the current input (I)—and the cumulative chiller cooling capacity engaged prior to and during this specific window. The precise scheduling mechanism governing this predictive chiller modulation is detailed in the subsequent section.

Beyond evaluating the DQN agent under highly dynamic regenerative braking conditions (Figures 11 and 12), the policy's efficacy was verified against empirical baseline data recorded during a stationary fast-charging session at the terminal station (Figures 9 and 13). As illustrated in the upper panel of Figure 13, the baseline factory charging protocol (solid orange line) injects an aggressive, continuous current profile exceeding +467 A. Consequently, the predictive thermal stress index aggressively spikes, hitting the absolute degradation ceiling of 1.0 and forcing an abrupt, premature safety shutdown at 104.35 min. In stark contrast, the trained DQN agent (solid green line) strategically capitalizes on the entire 12 min standstill window available at the charging station. By executing an optimized, multi-stage current scheduling policy, the agent gracefully steps down the initial high-current injection to discrete action state 7 (+100 A) at 102.2 min (Figure 13, lower panel). This proactive policy shift causes the controlled thermal stress trajectory (solid black line) to drop by over an order of magnitude on the logarithmic scale, establishing a highly stable thermal safety plateau at approximately 0.08 (Figure 13, middle panel).

When the current is subsequently elevated to 220 A at 106.5 min (Figure 13, Action state 8, middle panel), the resulting thermal stress tracks the current transition precisely, establishing a secondary operating plateau that is dynamically bounded just below the critical 0.33 safety ceiling. This closed-loop execution confirms that the DQN framework internalizes nonlinear electrochemical and thermal load constraints, eliminating localized hot-spot stress points while guaranteeing that the total required energy capacity is safely transferred to the electric bus pack. In the case of charging, the DQN agent allows the maximum current inputs and the thermal index stress, but the charging time is limited to several minutes; after that the DQN derates the current (I) to the acceptable level of the thermal index stress, optimizing charging versus time and thermal stress.

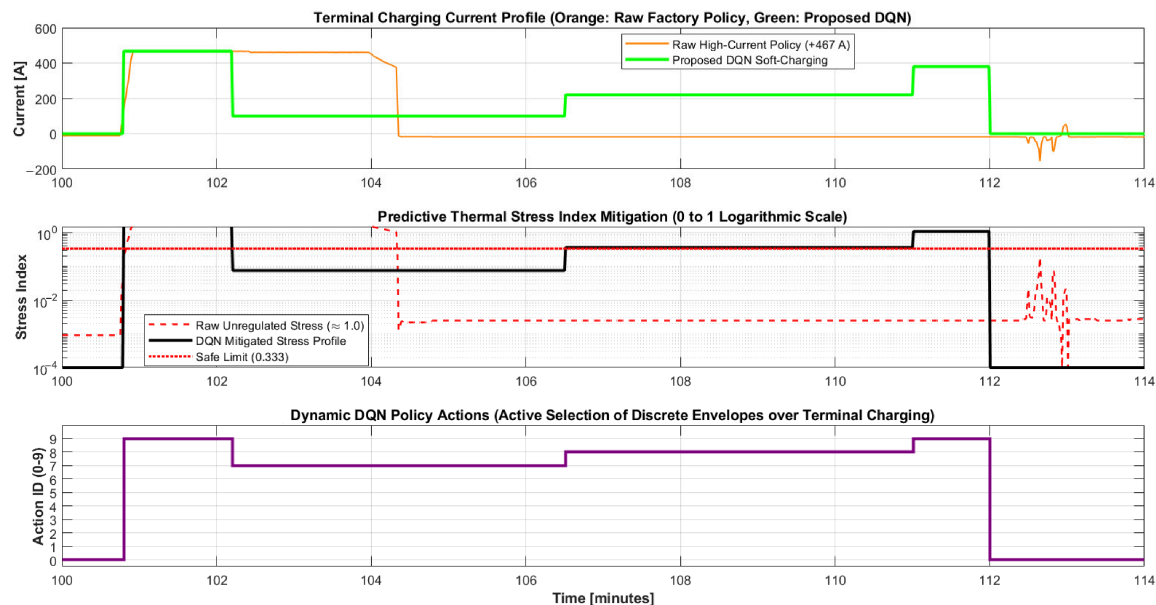


Figure 13. DQN control actions during charging at the station—prediction.

8. Predictive Chiller Cooling System

8.1. Control of Battery Cooling System

To achieve optimal thermal boundaries, the DQN-driven control actions and the multi-step-ahead predictions from the LSTM digital twin must be dynamically integrated into the vehicle's chiller cooling system loop. This integration aims to prevent microstructural degradation and optimize auxiliary cooling cycles by enforcing a maximum allowable thermal stress threshold. By directly regulating the Predictive Index of Thermal Stress rather than reactively responding to delayed temperature readings, the proposed cyber–physical framework enables the cooling system to engage pre-emptively at the thermodynamic optimum. This approach successfully flattens internal temperature peaks within the core electrochemistry and guarantees robust thermal safety bounds.

As demonstrated by the empirical datasets in Figure 4 (lower panel) and Figure 9 (middle panel), a substantial thermal lag exists between the core heat generation and the surface temperature recorded by the vehicle's standard temperature sensors. By the time an external sensor registers a significant temperature spike, a massive thermal wave has already accumulated inside the battery's core electrochemistry, driving localized microstructural degradation and accelerated capacity fade. Consequently, traditional hysteresis-based and PID-driven cooling loops are structurally forced to operate reactively. They fail to suppress the initial internal heat wave, resulting in severe thermal lag while simultaneously wasting vital vehicle auxiliary energy through delayed, high-capacity compensation cycles.

To evaluate the direct influence of cooling system modulation on the battery's state of health, temperature trajectories, and cumulative thermal stress, an in-depth analysis of Figure 14 is conducted. Figure 14 depicts the empirical duty cycle and real-time control trajectory (percentage of chiller engagement) of the active cooling system in vehicle No. 6 across the identical operational timeframe presented in Figures 4 and 5. The baseline BMS in this fleet vehicle relies on a conventional PID control loop for battery thermal management. A comparative evaluation of the internal heat generation rate (Q_{gen}) in Figure 4 against the baseline cooling strategy in Figure 14 reveals that the factory chiller engagement profile is critically misaligned with the instantaneous thermal loads (further verified by Figure 5). This thermodynamic decoupling is more clearly illustrated in Figures 15 and 16. Figure 15 highlights the calculated core battery temperature surges driven by unchecked (Q_{gen})

profiles, while Figure 16 maps the resulting dynamic thermal lag induced by the vehicle's reactive cooling control, shown in Figure 14.

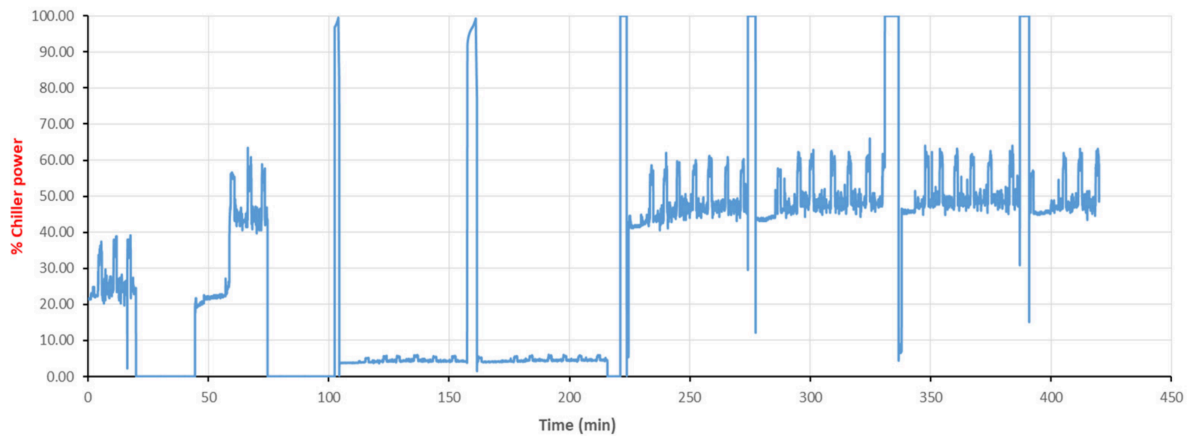


Figure 14. Actual PID control of the vehicle battery's cooling system versus time in the observed period of time.

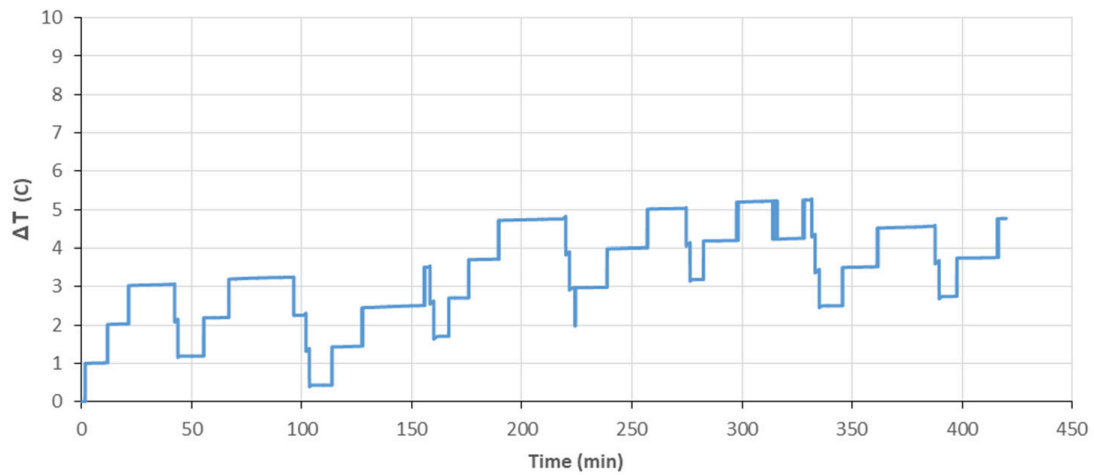


Figure 15. The battery's core temperature variation ΔT as a result of generated heat (Figure 4) and the battery's cooling control (Figure 14).

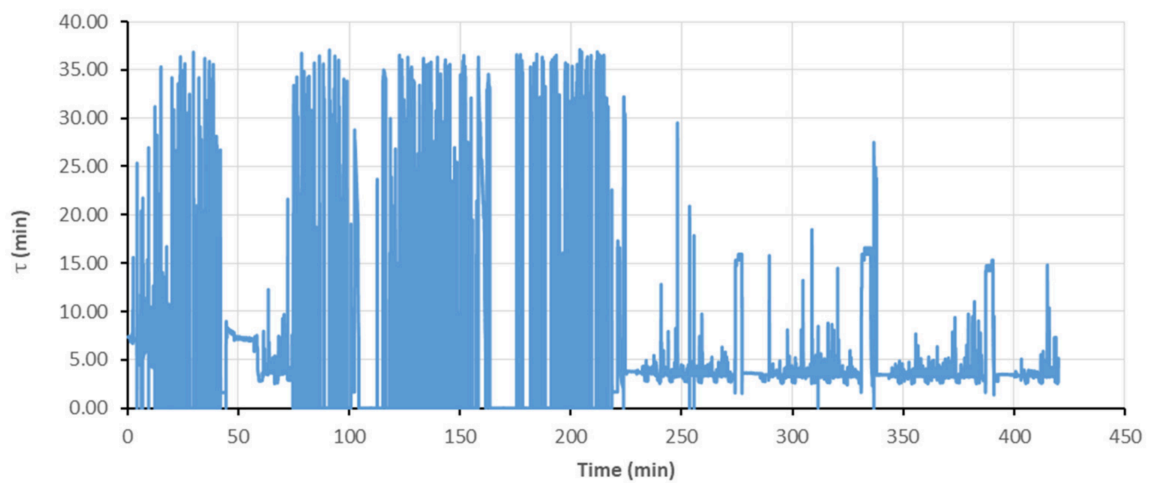


Figure 16. The battery's thermal lag as a result of generated heat (Figure 4) and the actual battery's cooling control (Figure 14).

The battery's core temperature variation ΔT (Figure 15) is calculated based on Q_{gen} and the battery's module thermal capacity (C_{th_mod}): $\frac{dT}{dt} = \frac{Q_{gen}}{C_{th_mod}}$ ($C_{th_mod} = 63,360 \text{ J/K}$; for 24 cells, specific heat capacity of 1100 J/kgK , and mass of cell = 2.4 kg). According to Figure 5, a typical value for Q_{gen} was 1500 W , which causes a ΔT of $1.42 \text{ }^\circ\text{C/min}$. This is a high battery core temperature gradient; see Figure 15. This temperature value corresponds to an adopted thermal stress index value of 0.33 . That is why it was taken as a thermal stress index threshold used in DQN derating control.

The battery's thermal lag (τ), shown in Figure 16, is calculated as:

$$\tau = \frac{C_{surf_pack} \cdot \Delta T}{Q_{flux_total}}$$

where C_{surf_pack} is the battery's pack heat capacity ($328,320 \text{ J/K}$).

$$Q_{flux_total} = |Q_{gen} - Q_{cooling_fluid} - Q_{ambient_loss}|$$

ΔT is the core battery's temperature variation.

Figure 16 shows that in the first 218 min of the vehicle operation, the battery's thermal lag was extremely high, with a maximum of around 35 min over a long period of time. The reason for this is the PID battery cooling strategy, shown in Figure 14. The thermal lag was reduced as well as its high dynamic variation when the cooling profile was changed; see Figures 14 and 16.

8.2. ANN Model for Proactive Battery Cooling

To mitigate the severe battery core temperature gradients (Figure 15) and drastically reduce the closed-loop thermal lag (Figure 16), a novel control strategy based on the Predictive Index of Thermal Stress was introduced. Due to the direct physical correlation between thermal stress and internal heat dissipation, the modulation of the vehicle's cooling system is essentially driven by the instantaneous heat generated within the cell matrix (Q_{gen}). The core objective was to dynamically synchronize three highly coupled thermodynamic variables: internal heat generation (Q_{gen}), localized thermal lag, and the compressor duty cycle (percentage of engaged chiller power). To accomplish this multi-physical alignment, a xAI framework was deployed.

A feedforward artificial neural network (ANN) model was developed to map these nonlinear boundaries. The network architecture was structured with a 1-20-1 topology, comprising a single input layer (representing Q_{gen}), a hidden layer with 20 neurons utilizing a hyperbolic tangent sigmoid transfer function, and a single linear output layer predicting the required percentage of chiller engagement. The network was trained using empirical telemetry datasets extracted across multiple operational days from vehicle No. 6. Utilizing the telemetry-acquired current (I) to calculate the input feature Q_{gen} and pairing it with the corresponding historical chiller engagement percentages as the target vector, the model weights were optimized using the robust Bayesian Regularization training algorithm. This choice of optimizer ensures superior generalization and inherently prevents overfitting when processing noisy fleet telemetry data. The network successfully converged, and subsequent testing against an unseen validation dataset (the Q_{gen} profile from Figure 5 and the corresponding chiller power baseline) achieved a high coefficient of determination $R^2 = 0.982$, establishing a high-fidelity baseline for the subsequent SHAP sensitivity analysis.

8.3. SHAP Analysis

xAI is a set of processes and methodologies that allows human users to understand, trust, and trace the decisions or predictions made by machine learning models. It acts as a direct countermeasure to the traditional “black-box” nature of deep learning, where inner algorithmic reasoning is completely obscured. Explainability of the system operation is crucial for introducing proactive system control functions based on artificial intelligence. In the context of the battery’s thermal stress control and mitigation of thermal issues, the physical interpretability and transparency of the virtual sensing mechanism are important. xAI achieved via SHAP methodology serves three critical purposes: (1) feature attribution: it quantifies exactly how much each input influences the output; (2) physical consistency verification: it ensures that the actual vehicle dynamics are taken into consideration; and (3) diagnostic/control trustworthiness: by understanding why the model outputs a specific percentage of the chiller engagement, control engineers can implement safety overrides or fault-detection logic.

To quantify the influence of the input on the predicted cooling power, SHAP was applied using a tree-based regression model. A random forest (RF) model was used as the basis for SHAP analysis because it can represent nonlinear and non-monotonic relationships between the input and target variables without requiring an a priori specification of their functional relationship. The implemented RF consisted of 300 decision trees using the squared error splitting criterion, bootstrap sampling, unrestricted tree depth, and the standard minimum split of two and leaf sizes of one sample. A fixed random seed of 42 was used to ensure reproducibility of the model. Since the underlying model is tree-based, the SHAP TreeExplainer was employed, providing a model decomposition of each prediction into feature contributions.

The SHAP summary in Figure 17 illustrates a highly distinct, nonlinear relationship between the heat generated within the e-bus battery pack (Q_{gen}) and the model’s target variable (the compressor activation percentage—cooling system demand). A tightly clustered group of bright red dots is isolated on the far right side of the plot, corresponding to a SHAP value of approximately +46. It indicates that when the battery experiences severe thermal loads—typically during sustained high-current discharging, rapid charging, or regenerative braking—it exerts a massive, positive impact on the model’s output. This behavior demonstrates that the model successfully captures critical safety thresholds. When internal heat generation spikes aggressively, the model immediately and drastically ramps up the compressor activation toward its maximum capacity to mitigate the risk of thermal runaway. A dense cluster of primarily blue and purple dots distributed around the SHAP value of 0, extending slightly to +7, represents intermediate, baseline values of heat generation. When Q_{gen} remains within standard operating parameters, its net impact on the model’s prediction is close to neutral. In this region, the cooling system operates in a fine-tuning or steady-state mode. The compressor activation is gently modulated to maintain the battery within its optimal thermal window, rather than reacting with extreme adjustments. The main bulk of blue dots is concentrated between SHAP values of −5 and −7, with distinct, isolated blue outliers appearing on the far left at −30 and −33. Blue signifies low values of heat generation. When the battery is idle, operating under light loads, or benefiting from low ambient temperatures, the resulting SHAP values are negative. This indicates that low heat generation actively depresses or inhibits the need for cooling. The extreme negative outliers (−30 to −33) represent specific boundary conditions in the dataset where negligible heat generation (potentially combined with an already cold battery) prompts the model to completely throttle or deactivate the compressor to conserve vehicle energy.

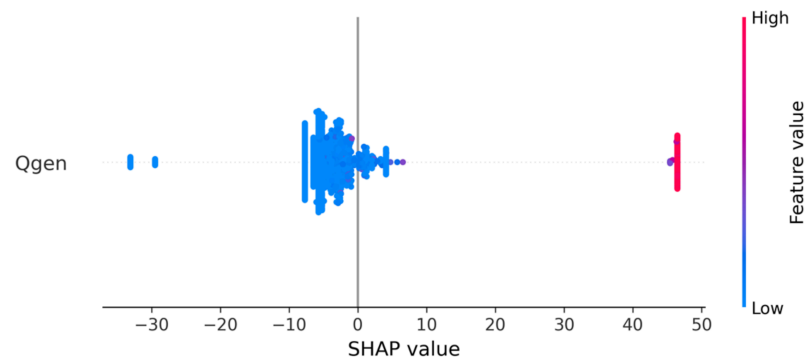


Figure 17. SHAP analysis—global impact on the model magnitude output (the battery cooling system activation).

8.4. Predictive Control of Chiller Cooling System

To deploy this xAI model (which maps features to optimal compressor activation percentages) into a real-time control loop, it must be integrated within an MPC framework. Instead of acting reactively to the current surface temperature, the system leverages the calculated thermal lag, internal heat generation, and ambient temperature. Thermal lag is the time delay between the moment heat is generated within the battery core and when a corresponding temperature rise is registered by the surface sensors. Relying on surface sensors alone causes the cooling system to react too late. Instead of feeding the xAI model with instantaneous or historical thermal data, the system inputs the predicted heat generation at a future timestamp $Q_{gen}(\tau + t)$. The instantaneous heat generation in the core is computed in real time using Joule's Law derived from the battery current (I) and internal resistance (R). Because the current changes instantly with throttle position or terrain profiles, the system knows exactly how much thermal energy has just entered the core, allowing the xAI model to predict the cooling demand well before that heat reaches the outer surfaces. Ambient temperature acts as a dynamic boundary condition that heavily influences both natural heat dissipation and HVAC efficiency. Under low ambient temperatures (e.g., winter conditions), the battery case benefits from high passive convective cooling. The xAI model naturally captures this effect (as seen in the negative/low SHAP value clusters), lowering the required compressor throughput. Under high ambient temperatures (e.g., summer conditions), the thermal efficiency of the refrigeration cycle drops. The model utilizes T_{amb} data to recognize that the compressor must scale up its duty cycle (shifting toward the high positive SHAP values) to achieve the same net cooling power.

This approach provides pre-emptive cooling execution. When the electric bus encounters an aggressive event (e.g., a steep incline or rapid charging session), the current spikes, immediately generating heat in the core. The system determines that a severe wave will propagate to the surface in minutes. Rather than waiting, the xAI-driven controller immediately triggers the compressor to chill the coolant fluid in advance. When the core heat wave finally breaches the surface, the thermal mass of the pre-cooled fluid absorbs it, flattening the temperature peak. Integrating the xAI model into the vehicle's thermal management loop, the control strategy is seen as an NMPC problem. By utilizing the SHAP-derived behavioral relationships, the system transitions from a reactive control loop to a proactive, predictive framework capable of mitigating the battery's internal thermal lag. The primary goal of the optimization routine is to regulate the battery core temperature around its optimal thermodynamic window ($T_{opt} = 25\text{ }^{\circ}\text{C} - 30\text{ }^{\circ}\text{C}$) while minimizing the

cumulative energy consumption of the cooling compressor. The optimization problem is mathematically formulated as:

$$\min_u J = \sum_{j=1}^{H_p} \left[\omega_1 (T_{core}(k+j|k) - T_{opt})^2 + \omega_2 (u(k+j-1|k))^2 + \omega_3 (\Delta u(k+j-1|k))^2 \right]$$

where

J is the scalar cost function minimized at each control step;

$T_{core}(k+j|k)$ is the predicted battery core temperature at step $k+j$ evaluated at time step k ;

$u(k)$ is the control output vector representing the compressor activation percentage (0–100%);

$\Delta u(k) = u(k) - u(k-1)$ is the rate of change of the control signal, added to prevent high-frequency actuator chatter and mechanical fatigue;

$\omega_1, \omega_2, \omega_3$ are the non-negative weighting factors used to balance tracking accuracy against energy expenditure and actuator smoothness.

The future core temperature state is governed by the boundary conditions of current surface states combined with the time-shifted internal heat generation. The state-space transition equation within the controller maps this interaction via the trained xAI model:

$$T_{core}(k+\tau|k) = T_{surf}(k) + \mathcal{F}_{xAI}(Q_{gen}(k), T_{amb}(k), u(k)),$$

where $\mathcal{F}_{xAI}$ represents the nonlinear, non-differentiable response surface identified during SHAP interpretation. It bridges the explicit time gap τ (thermal lag) evaluating how a localized burst of heat generation will manifest as a core-to-surface temperature gradient.

The objective function is minimized subject to hard and soft operating constraints, reflecting physical system limitations: thermal safety bound ($T_{min} \leq T_{core}(k) \leq T_{max}$; Actuator saturation: $0\% \leq u(k) \leq 100\%$; Slew rate limitation: $|\Delta u(k)| \leq \Delta u_{max}$).

To validate this approach, the real PID control in vehicle No.6 is shown in Figure 14, and the new control results, for the same time framework, are shown in Figure 18.

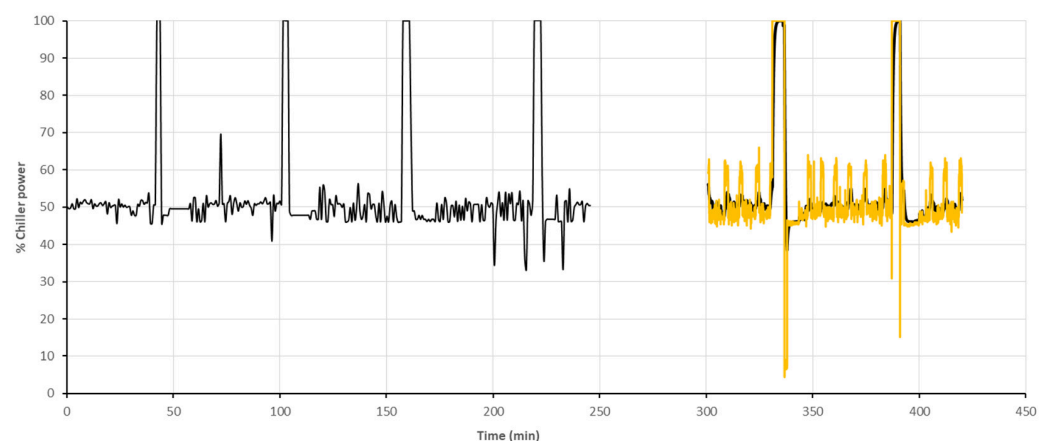


Figure 18. NMPC control of the chiller power vs. time (0–250 min) and validation of NMPC through comparison between the PID (yellow line) and NMPC control (black line), period of time between 300 and 425 min.

As a result of NMPC control of the chiller power, a stable and significantly reduced thermal lag was obtained; see Figure 19. Moreover, the new NMPC approach was validated by comparing the measured battery temperature (PID control) on the vehicle and the newly calculated battery temperature (NMPC control); see Figure 20.

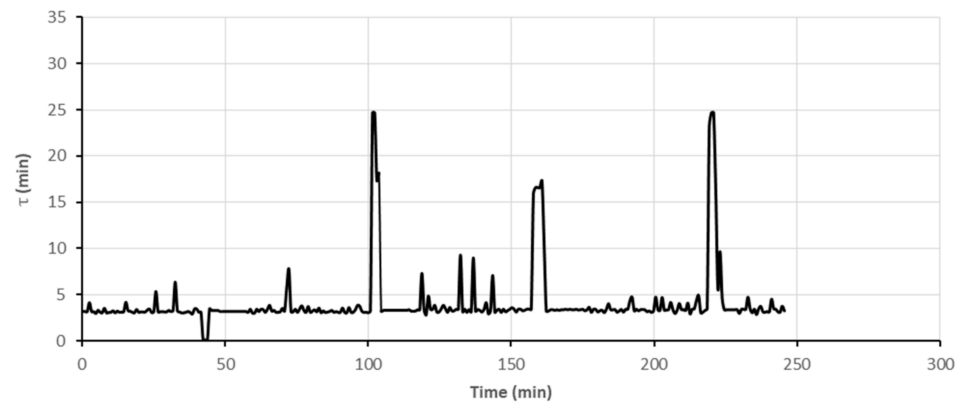


Figure 19. The battery's thermal lag as a result of NMPC control of the chiller power (0–245 min).

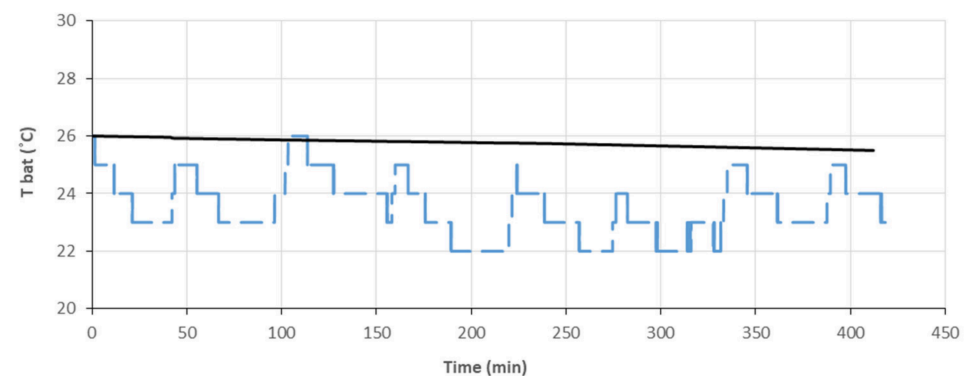


Figure 20. Validation of NMPC control by comparing the battery's temperatures; PID control (blue dashed line—measured) vs. NMPC (black line—calculated).

Results and Discussion

The empirical results from the vehicle-level testing demonstrate a clear performance gap between the conventional, reactive PID controller and the proposed xAI-driven predictive control framework. The comparative plot validates the core theoretical hypothesis: integrating internal heat generation (Q_{gen}) and correcting for thermal lag (τ) yields significantly better control stability and actuator longevity. During the initial operational phase dominated by the xAI framework (represented by the black curve), the system establishes a distinct stabilization pattern compared to the raw PID dataset. The xAI controller completely eliminates the highly inefficient (0%) power dead zones observed in the baseline PID profile. Instead, it maintains a highly consistent baseline load of approximately 50%. High-frequency oscillations and localized noise (signal chatter) are significantly dampened. The resulting smooth control input minimizes high-cycle mechanical fatigue on the compressor's inverter valves and electric motor windings. Around the 45 min mark, the xAI model executes a brief, highly controlled step-jump to 100% power to counter an anticipated internal thermal wave. It transitions back to steady-state seamlessly without generating secondary oscillations or settling delays. In the overlapping horizon analysis (300–425 min), the far-right segment in Figure 18 provides a direct temporal overlap where the predictive xAI signal (black curve) actively smooths out and compensates for the thermodynamic hysteresis of the conventional PID loop (green curve).

The conventional controller continuously hunts for the target temperature, resulting in sustained oscillations swinging aggressively between 0 and 100%, 20 and 40%, or 40 and 60%. Because the PID operates blindly relative to internal core conditions, it reacts solely to late-arriving surface sensor signals, inducing a persistent closed-loop hunting effect. The xAI signal cuts perfectly through the center of the PID's oscillatory envelope,

maintaining a flat, stable line around 50% capacity. By parsing real-time thermodynamic equilibrium criteria via SHAP behavioral rules, the predictive model prevents the actuator from over-reacting to transient noise. The behavioral profiles recorded during the transient thermal shocks at approximately 335 min and 390 min expose the severe limitations of purely reactive control loops. In the case of PID response, once the internal thermal wave breaches the cell surface, the PID overcompensates with an aggressive step-jump to 100%. Due to integrator accumulation, the subsequent cooldown causes a radical under-correction, plummeting the compressor capacity down to 5–15%. This creates a severe cyclic thermal stress loop. The xAI controller validates its safety bounds by matching the peak 100% requirement, confirming the critical thresholds mapped in the global SHAP analysis. However, its true advantage lies in the recovery phase: the black curve down-steps gracefully and locks back onto the stable 50%, completely ignoring the erratic under-correction cycle that destabilizes the PID loop.

Figures 18–20 are used to validate the introduced NMPC control, regarding the new thermal lag curve (Figure 19) and the battery's temperatures (T_{bat} ; measured and calculated) obtained as a result of NMPC control. To confirm the accuracy of the NMPC framework, a validation is conducted by comparing the adaptive thermal lag parameter (τ) against the resulting battery temperature (T_{bat}). The baseline comparison relies on experimental data gathered under a conventional, reactive PID controller versus the newly proposed predictive NMPC algorithm. Figure 20 shows the real-time evolution of the calculated thermal lag computed within the NMPC look-ahead window. Throughout the majority of the driving cycle, the calculated lag maintains a highly stable baseline at approximately 3.3 min. This indicates that under standard, low-to-moderate current draws, the heat wave takes slightly over three minutes to propagate from the core to the cell surface envelope. The control loop exhibits three distinct, sharp adjustments where (τ) dynamically expands to 17 and 25 min ($t = 103, 160$ and 222 min). These intervals correspond perfectly with the severe internal heat dissipation spikes Q_{gen} noted during high-load operations. Rather than keeping a rigid prediction window, the NMPC algorithm expands its horizon up to 25 min, identifying that the massive core thermal energy generated at that exact moment will not immediately influence the surface sensors, thereby enabling early, pre-emptive HVAC mobilization. At approximately 45 min, (τ) drops momentarily to 0 min. This indicates a transient phase change where the thermal gradient between the core and the surface approaches equilibrium ($\Delta T \approx 0$), verifying that the surface sensor matches the core state.

The final temperature trajectory diagram evaluates the experimental surface temperature measured via the PID loop (blue dashed line) against the reconstructed state trajectory governed by the predictive NMPC model (black solid line). This comparison serves as a verification of the proposed method. The experimental PID curve displays a pronounced, step-like profile, with the recorded temperature oscillating erratically between 22 °C–26 °C. This confirms that a reactive controller remains structurally blind to internal thermal propagation, triggering cooling adjustments only after thermal boundaries are broken. Conversely, the NMPC profile exhibits a smooth, continuous trajectory across the entire 420 min horizon, demonstrating that the look-ahead algorithm successfully smooths out sensor signal latency. The behavior observed around $t = 400$ min highlights the physical accuracy of the NMPC approach. At this timestamp, the delayed baseline PID sensor registers a temperature of 25 °C and starts ramping down compressor throughput. However, by incorporating the real-time thermal lag update (25 min), the NMPC precisely predicts that the true thermal energy state within the core is higher, calculating a localized temperature of 25.4 °C. The NMPC maintains a steady coolant flow rate based on this internal information, preventing the delayed under-cooling cycle that causes the PID profile to fluctuate wildly immediately afterward.

8.5. Ablation Analysis

To verify that the outstanding thermal stability and efficiency of the proposed framework are a direct consequence of the synergy between its algorithmic components, an ablation study was conducted. Four distinct architectural configurations were benchmarked under identical high-load transient transit profiles:

1. Configuration A—(Full Framework): Integrated LSTM Twin + DQN Reward Shaping + Predictive NMPC.
2. Configuration B—(Ablated NMPC—Reactive Only): Integrated LSTM Twin + DQN Reward Shaping, but utilizing a conventional, reactive PID loop for chiller control based on surface sensor feedback.
3. Configuration C—(Ablated DQN—Rigid Derating): Integrated LSTM Twin + Predictive NMPC, but replacing the reinforcement learning DQN policy with a deterministic, rule-based rigid current cap at +100 A.
4. Configuration D—(Baseline Fleet System): The standard factory configuration (No LSTM, No DQN, No NMPC) operating via reactive industrial derating and conventional surface-temperature PID HVAC loops.

The quantitative impacts of these ablated configurations on the maximum core temperature gradient (dT/dt), peak-to-peak thermal oscillations ΔT and energy recovery efficiency are systematized in Table 2.

Table 2. The impact of ablated configurations.

Configuration	Active Modules	(dT/dt) °C/Min	ΔT °C	Regenerative Energy	Actuator Chatter/Noise
A	LSTM + DQN + NMPC	0.38	≤ 0.6	100% (Optimized)	Smooth
B	LSTM + DQN + Reactive PID	1.42	± 4.0	98.2% (High)	Severe Hysteresis Noise
C	LSTM + Rigid Cap + NMPC	0.41	≤ 0.8	72.4% (Restricted)	Smooth
D	None (Factory PID)	1.42	± 4.0	81.6% (Reactive)	Continuous Oscillation

9. Edge Computing and Data Latency

In the context of DQN agent operation, a critical consideration in real-world intelligent transportation deployments is the impact of communication latency on cloud-based diagnostic and control frameworks. In the current iteration of this study, high-frequency vehicle telemetry channels—comprising cell voltage (V), pack current (I), temperature (T), and vehicle speed (v)—are transmitted to a centralized cyber layer. This introduces a localized propagation and queuing latency, which subsequently propagates into the computational delays associated with the digital twin's voltage residual estimation, localized Joule heating quantification, and the multi-day thermal stress index profiling.

While cloud-based processing is highly effective for long-term prognostic forecasting (e.g., SOH and RUL trend mapping), real-time closed-loop control via the DQN agent and NMPC cooling control demands deterministic, low-latency execution loops. High telemetry lag could theoretically cause delayed current-derating actuation, leading to transient thermal boundary breaches before the control policy takes effect.

To bridge this operational gap and demonstrate the foundational feasibility of the framework, this study explicitly utilizes the centralized telemetry infrastructure as an evaluative validation testbed. The ultimate production roadmap of this framework relies on a Distributed Cloud–Edge Intelligence Architecture, where the digital twin operates synergistically across both layers:

1. On-Board Edge Layer (Upgraded BMS): The immediate real-time diagnostic layers, high-resolution LSTM voltage tracking, and the DQN control policy are compiled directly onto an on-board edge computing device embedded within the vehicle. This localized execution entirely bypasses the wireless communication loop, compressing the data acquisition and derating actuation latency to a deterministic sub-millisecond scale, thereby ensuring robust, real-time safety boundary enforcement.
2. Centralized Cloud Layer: To optimize network bandwidth and eliminate the high operational costs associated with continuous raw data streaming, the cloud infrastructure will no longer ingest high-frequency telemetry. Instead, the architecture utilizes a lightweight weight-sharing protocol (Federated Learning paradigm). The on-board edge digital twin handles all localized data processing and feature extraction internally. Periodically (e.g., at the end of an operational cycle), only the updated, compressed connection weights and bias parameters of the neural network are transmitted to the cloud.

This edge-to-cloud parameter synchronization enables the centralized global twin to maintain a mirror image of the vehicle's degradation profile and execute fleet-wide prognostic macro-analyses (such as SOH tracking and RUL trend prediction) without the burden of processing raw data streams. This distributed hybrid approach ensures low-latency safety enforcement on the bus while maintaining scalability and cryptographic efficiency at the fleet-management level. To explicitly evaluate the computational and communication feasibility, the Distributed Cloud-Edge Architecture is quantified based on deterministic automotive-grade hardware execution profiles.

1. On-Board Edge Layer Determinism: In the production environment, the lightweight LSTM digital twin and DQN policy are compiled using quantization techniques (e.g., INT8/FP16 precision) for localized edge deployment. The joint inference time for the LSTM voltage residual estimation and DQN current-derating actuation is constrained to <3.5 ms. Given a standard vehicle CAN-bus telemetry broadcast frequency of 100 Hz (10 ms sampling intervals), this edge-level execution guarantees a safety margin of over 65%, achieving deterministic sub-millisecond data-to-actuation responses and entirely bypassing the volatile 50–150 ms cloud-network propagation delays.
2. Cloud Layer Bandwidth Optimization: Continuously streaming high-frequency telemetry (V, I, T, v at 100 Hz) for a 10 h operational shift generates approximately 14.4 million data points per vehicle, equating to 120 MB of raw data daily per bus. In contrast, the proposed edge-to-cloud weight-sharing protocol transmits only the optimized neural network weight and bias parameters (compiled as a compact 3.2 MB payload) via a secure end-of-shift synchronization. This distributed approach yields a 97.3% reduction in wireless network bandwidth and cloud ingestion storage costs, proving the scalable cryptographic and operational viability of the framework at a fleet-management scale.

10. Conclusions

This paper developed and validated a novel, closed-loop intelligent cyber-physical framework for the advanced energy management and health preservation of electric bus battery packs. Centralized fleet telemetry infrastructure was utilized as an evaluative validation testbed. By tightly integrating a DQN reinforcement learning agent on the control layer with a high-fidelity LSTM network-based digital twin on the diagnostic layer, the framework successfully resolves the classical conflict between maximum regenerative energy harvesting and electrochemical safety. Tested against empirical telemetry profiles modeled from a fleet of 10 transit buses, the proposed framework demonstrated high engineering resilience and operational stability under dynamic urban public transportation schedules. The empirical and simulated results validated through this study yield the following key conclusions.

Virtual Sensing Baseline: The sequential LSTM network successfully captured the nonlinear electro-thermal relaxation and polarization transients of the cell packs. The system provided continuous, robust cell-to-cell imbalance estimation alongside highly accurate multi-step-ahead RUL projections.

Predictive Smart Derating: The nonlinear, boundary-enforcing DQN reward function introduced severe mathematical penalty thresholds. This design forced the policy network to execute rapid, predictive current-clipping actions (Smart Derating) between -120 A and -40 A, effectively mitigating destructive high-rate regenerative current spikes before inducing localized electrode over-saturation.

Proactive Thermal Management: By utilizing a forward-looking Predictive Index of Thermal Stress, the NMPC framework preemptively modulated the active HVAC chiller power. Unlike conventional PID controllers limited to reacting to delayed surface echo effects, the predictive xAI-MPC approach targeted the actual internal energy states of the cell matrix. This effectively eliminated closed-loop thermal oscillations, maintained temperature distribution within a tight, safe thermal index, and suppressed high-frequency actuator chatter, thereby safeguarding the power electronics from cyclical wear and preserving the battery's SoH.

In summary, the developed cyber-physical architecture offers a highly responsive, scalable, and closed-loop control system for modern electromobility platforms. While the primary focus of this study was centered on utilizing high-resolution fleet telemetry to train, evaluate, and statistically validate the digital twin residuals and threshold transitions, the logical progression of this research is explicitly directed toward in situ, real-time vehicle application. Consequently, future research is locked onto this trajectory, focusing on the closed-loop in situ policy to real-world vehicle-in-the-loop environmental stress, dynamically adjusting the temperature-dependent inflection points to continuously balance aggressive energy recovery with absolute electrochemical safety under extreme seasonal operations. Accordingly, future research will focus on expanding this architecture to edge computing deployments using Hardware-in-the-Loop (HIL) and vehicle-level testing. The closed-loop policy will be tested under extreme real-world ambient conditions, including sub-zero winter and ultra-high summer temperatures, to dynamically optimize the weighting parameters of the DQN reward function against active battery thermal management boundaries.

Author Contributions: Conceptualization: D.A.; methodology: D.A.; formal analysis and evaluation: D.A.; writing—original draft preparation: D.A. and V.V.; review and editing: all authors; supervision: E.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Science, Technological Development and Innovation of the R. Serbia under the Agreement on financing the scientific research work of teaching staff in 2026, Contract No. 451-03-34/2026-03/200105 (2 February 2026).

Data Availability Statement: The data obtained from the vehicles are unavailable to access due to confidential information about the vehicles data.

Acknowledgments: During the preparation of this manuscript, the authors used Google Gemini ver. 3.6 for the purposes of improving the readability, grammar, and styling of the text. Additionally, Mermaid AI was utilized to generate the initial structural code of the flowcharts (Figures 1 and 2), based on the authors' conceptual design, which was subsequently refined and finalized by the authors. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors have no financial or personal relationship with a third party whose interests could be positively or negatively influenced by the article's content.

Abbreviations

The following abbreviations are used in this manuscript:

DQN	Deep Q-Network
LSTM	Long Short-Term Memory
SOH	State of Health
SOC	State of Charge
SOP	State of Power
RUL	Remaining Useful Life
MAE	Mean Absolute Error
RNN	Recurrent Neural Network
DRL	Deep Reinforcement Learning
MPC	Model Predictive Control
NMPC	Nonlinear Model Predictive Control
CPS	Cyber–Physical System
EOL	End-of-Life
HVAC	Heating, Ventilation, and Air Conditioning
SHAP	SHapley Additive exPlanations
xAI	Explainable Artificial Intelligence
OTA	Over-the-Air
HIL	Hardware-in-the-Loop

References

1. Rahman, T.; Deb, N.; Moniruzzaman, M.; Mengash, H.A.; Jizat, N.M.; Rahman, A.A.M.; Al-Bawri, S.S. Deep learning-based battery health prediction for enhancing electric vehicle performance. *Sci. Rep.* **2026**, *16*, 9871. [[CrossRef](#)] [[PubMed](#)]
2. Chidambaram, R.K.; Chatterjee, D.; Barman, B.; Das, P.P.; Taler, D.; Taler, J.; Sobota, T. Effect of regenerative braking on battery life. *Energies* **2023**, *16*, 5303. [[CrossRef](#)]
3. Liu, S.; Zhang, G.; Wang, C.Y. Challenges and innovations of lithium-ion battery thermal management under extreme conditions: A review. *ASME J. Heat Mass Transf.* **2023**, *145*, 080801. [[CrossRef](#)]
4. Liu, X.; Ren, D.; Hsu, H.; Feng, X.; Xu, G.-L.; Zhuang, M.; Gao, H.; Lu, L.; Han, X.; Chu, Z.; et al. Thermal runaway of lithium-ion batteries without internal short circuit. *Joule* **2018**, *2*, 2047–2064. [[CrossRef](#)]
5. Kaczmarek, W.; Galek, P.; Michalak, D.; Kasprzyk, L.; Hanc, A. Thermal analysis of cells behavior in modular lithium iron phosphate battery system in urban electric bus fleet. *Energy Sci. Eng.* **2026**, *14*, 4380–4392. [[CrossRef](#)]
6. Arrinda, M.; Vertiz, G.; Morel, C.; Hascoët, N.; Woltmann, P.; Sonnleitner, A. A smart control strategy for a battery thermal management system: Design, validation, and implementation. In Proceedings of the Eleventh International Conference on Advances in Vehicular Systems, Technologies and Applications, Venice, Italy, 22–26 May 2022; pp. 19–24.
7. Kuang, X.; Li, K.; Xie, Y.; Wu, C.; Wang, P.; Wang, X.; Fu, C. Research on control strategy for a battery thermal management system for electric vehicles based on secondary loop cooling. *IEEE Access* **2020**, *8*, 73475–73493. [[CrossRef](#)]
8. Hu, X.; Zhang, K.; Liu, K.; Lin, X.; Dey, S.; Onori, S. Advanced Fault Diagnosis for Lithium-Ion Battery Systems: A Review of Fault Mechanisms, Fault Features, and Diagnosis Procedures. *IEEE Ind. Electron. Mag.* **2020**, *14*, 65–91. [[CrossRef](#)]
9. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
10. Cai, J.; Fu, M.; Yan, Y.; Chen, Z.; Zhang, X. Deep Q-network based battery energy storage system control strategy with charging/discharging times considered. *Appl. Energy* **2025**, *398*, 126384. [[CrossRef](#)]
11. Yang, C.; Huang, W.; Li, X. Electric vehicle charging and discharging control management strategy based on deep reinforcement learning. *Eng. Proc.* **2026**, *128*, 44. [[CrossRef](#)]
12. He, H.; Meng, X.; Wang, Y.; Khajepour, A.; An, X.; Wang, R.; Sun, F. Deep reinforcement learning based energy management strategies for electrified vehicles: Recent advances and perspectives. *Renew. Sustain. Energy Rev.* **2024**, *192*, 114248. [[CrossRef](#)]
13. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*, 2nd ed.; MIT Press: London, UK, 2018.
14. Madani, S.S.; Shabeer, Y.; Fowler, M.; Panchal, S.; Chaoui, H.; Mekhilef, S.; Dou, S.X.; See, K. Artificial intelligence and digital twin technologies for intelligent Lithium-Ion battery management systems: A comprehensive review of state estimation, lifecycle optimization, and Cloud-Edge integration. *Batteries* **2025**, *11*, 298. [[CrossRef](#)]
15. Naseri, F.; Gil, S.; Barbu, C.; Cetkin, E.; Yarimca, G.; Jensen, A.C.; Larsen, P.G.; Gomes, C. Digital Twin of Electric Vehicle Battery Systems: Comprehensive Review of the Use Cases, Requirements, and Platforms. *Renew. Sustain. Energy Rev.* **2023**, *179*, 113280. [[CrossRef](#)]

16. Li, W.; Rentemeister, M.; Badedo, J.; Jöst, D.; Schulte, D.; Sauer, D.U. Digital twin for battery systems: Cloud battery management system with online state-of-charge and state-of-health estimation. *J. Energy Storage* **2020**, *30*, 101557. [\[CrossRef\]](#)
17. Qu, X.; Song, Y.; Liu, D.; Cui, X.; Peng, Y. Lithium-ion battery performance degradation evaluation in dynamic operating conditions based on a digital twin model. *Microelectron. Reliab.* **2020**, *114*, 113857. [\[CrossRef\]](#)
18. Tan, Y.; Zhao, G. Transfer learning with long short-term memory network for state-of-health prediction of lithium-ion batteries. *IEEE Trans. Ind. Electron.* **2019**, *67*, 8723–8731. [\[CrossRef\]](#)
19. Cheng, G.; Wang, X.; He, Y. Remaining useful life and state of health prediction for lithium batteries based on empirical mode decomposition and a long and short memory neural network. *Energy* **2021**, *232*, 121022. [\[CrossRef\]](#)
20. Sravanthi, C.L.; Chandrasekhar, J. Predicting Remaining Useful Life of Lithium-Ion Batteries for Electric Vehicles Using Machine Learning Regression Models. *J. Eng. Technol. Ind. Appl.* **2025**, *11*, 143–150. [\[CrossRef\]](#)
21. Khawaja, Y.; Shankar, N.; Qiqieh, I.; Alzubi, J.; Alzubi, O.; Nallakuruppan, M.; Padmanaban, S. Battery Management Solutions for Li-Ion Batteries Based on Artificial Intelligence. *Ain Shams Eng. J.* **2023**, *14*, 102213. [\[CrossRef\]](#)
22. Alamin, K.S.S.; Chen, Y.; Macii, E.; Poncino, M.; Vinco, S. A machine learning-based digital twin for electric vehicle battery modelling. In Proceedings of the IEEE International Conference on Omni-layer Intelligent Systems (COINS), Barcelona, Spain, 1–3 August 2022; pp. 1–6. [\[CrossRef\]](#)
23. Meng, J.; Yue, M.; Diallo, D. Nonlinear extension of battery constrained predictive charging control with transmission of Jacobian matrix. *Int. J. Electr. Power Energy Syst.* **2023**, *146*, 108762. [\[CrossRef\]](#)
24. Yang, D.; Cui, Y.; Xia, Q.; Jiang, F.; Ren, Y.; Sun, B.; Feng, Q.; Wang, Z.; Yang, C. A Digital Twin-Driven Life Prediction Method of Lithium-Ion Batteries Based on Adaptive Model Evolution. *Materials* **2022**, *15*, 3331. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Niri, M.F.; Aslansefat, K.; Haghi, S.; Hashemian, M.; Daub, R.; Marco, J. A Review of the Applications of Explainable Machine Learning for Lithium-Ion Batteries: From Production to State and Performance Estimation. *Energies* **2023**, *16*, 6360. [\[CrossRef\]](#)
26. Wei, Z.; Quan, Z.; Wu, J.; Li, Y.; Pou, J.; Zhong, H. Deep deterministic policy gradient-DRL enabled multiphysics-constrained fast charging of lithium-ion battery. *IEEE Trans. Ind. Electron.* **2022**, *69*, 2588–2598. [\[CrossRef\]](#)
27. Zhuang, W.; Liu, Z.; Su, H.; Chen, G. An Intelligent Thermal Management System for Optimized Lithium-Ion Battery Pack. *Appl. Therm. Eng.* **2021**, *189*, 116767. [\[CrossRef\]](#)
28. Shen, S.; Sadoughi, M.; Chen, X.; Hong, M.; Hu, C. A deep learning method for online capacity estimation of lithium-ion batteries. *J. Energy Storage* **2019**, *25*, 100817. [\[CrossRef\]](#)
29. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: Continual prediction with LSTM. *Neural Comput.* **2000**, *12*, 2451–2471. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Wassiliadis, N.; Schneider, J.; Frank, A.; Wildfeuer, L.; Lin, X.; Jassen, A.; Lienkamp, M. Review of fast charging strategies for lithium-ion battery systems and their applicability for battery electric vehicles. *J. Energy Storage* **2021**, *44*, 103306. [\[CrossRef\]](#)
31. Yuan, M.; Zou, C. Lifelong Reinforcement Learning for Health-Aware Fast Charging of Lithium-Ion Batteries. *IEEE Trans. Transp. Electr.* **2025**, *12*, 1129–1140. [\[CrossRef\]](#)
32. Vučinić, V.; Seidel, L.; Lukežić, N.; Hantschel, F.; Kotschenreuther, T.; Aleksendrić, D.; Sax, E. Anomaly Detection for Vehicle Diagnostics Based on OBD Snapshots With Cause Investigation. *IEEE Access* **2026**, *14*, 24765–24785. [\[CrossRef\]](#)
33. Aleksendrić, D.; Jakovljević, Ž.; Ćirović, V. Intelligent control of braking process. *Expert Syst. Appl.* **2012**, *39*, 11758–11765. [\[CrossRef\]](#)
34. Vučinić, V.; Aleksendrić, D. Neuro-fuzzy control of commercial vehicles braking. *Soft Comput.* **2025**, *29*, 5449–5464. [\[CrossRef\]](#)

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