

Physics-Based State-of-Charge Modelling and Operational Optimization for a Battery-Electric Shunting Locomotive Digital Twin

Tsvetomir Gotsov ^{*}, Venelin Todorov ^{†‡}

^{*}Express Service OOD, Ruse, Bulgaria

[†]Institute of Information and Communication Technologies, Bulgarian Academy of Sciences

25A Acad. G. Bonchev Str., 1113 Sofia, Bulgaria

[‡] Centre of Excellence in Informatics and Information and Communication Technologies,

Acad. G. Bonchev Str. Bl. 25A, 1113 Sofia, Bulgaria

Email: tsvetomir.gotsov@gmail.com, venelin@parallel.bas.bg,

Abstract—Battery-electric shunting locomotives in non-electrified industrial environments require reliable State-of-Charge (SoC) prediction to ensure uninterrupted shift operation and avoid unexpected battery depletion. This paper presents an empirical study and digital twin framework for the ES3000 shunting locomotive, equipped with a flooded lead-acid traction battery of 2752 Ah at 80 V. Correlation analysis of real operational telemetry collected via a Teltonika GPS tracking system and the Wialon cloud platform reveals that cumulative Ampere-hour (Ah) integration achieves a Pearson correlation of $r = -0.9975$ with the onboard Battery Discharge Indicator (BDI), while instantaneous features such as terminal voltage ($r = 0.08$) and traction current ($r = 0.14$) carry negligible predictive power. A physics-based model combining Coulomb counting with Peukert high-rate capacity correction is calibrated from a single 3-hour real-world session (1079 records, BDI 100% to 54%) and achieves a prediction RMSE of 1.81%. Building on the calibrated model, we derive a recharge threshold optimization that balances autonomy reserve against recharge frequency, identifying BDI 30–40% as the optimal trigger for mid-shift charging. A route simulation mode forecasts SoC from GPS waypoints and an operational scenario, enabling pre-shift mission planning without live telemetry. The complete pipeline is implemented as an open Python application compatible with both Wialon and PLC telemetry export formats.

Index Terms—digital twin, battery locomotive, lead-acid battery, state of charge, Ampere-hour integration, Peukert effect, recharge optimization, route simulation, railway telemetry

I. INTRODUCTION

Battery-electric shunting locomotives (BESL) are increasingly deployed in ports, steelworks, and industrial terminals where overhead electrification is impractical. Unlike road electric vehicles, a BESL operates within a confined area of 1–5 km radius, travels at most 30–40 km per shift, and is unavailable during the 4–8 hours required to recharge its traction battery. Predicting when the battery will reach a

critically low SoC is therefore a direct safety and availability concern.

The Digital Twin (DT) paradigm [1] offers a natural framework for this problem: a continuously updated virtual replica of the energy system that can forecast SoC under different shunting missions and issue early recharge warnings before the locomotive is stranded mid-operation. Survey work by Ghaboura et al. [2] confirms that DT for Railway (DTR) has expanded rapidly in maintenance and optimization domains, yet battery-electric locomotive energy management remains an open gap in the literature.

The specific challenge for BESL SoC prediction combines several unknowns: a flooded lead-acid battery with pronounced Peukert capacity loss at high traction currents; unknown and variable payload mass per movement; the absence of an onboard track gradient sensor; and a BDI sensor that updates in discrete 1% steps at irregular intervals. Prior attempts to use instantaneous features (voltage, current, speed) without time integration failed to produce useful predictions on the available data, as confirmed by the correlation analysis in this paper.

This paper makes the following contributions:

- An empirical correlation study on real ES3000 telemetry identifying cumulative Ah integration as the dominant predictor of BDI ($r = -0.9975$) and quantifying the failure of instantaneous features;
- A calibrated Peukert-corrected physical model achieving RMSE = 1.81% on a 3-hour real session without machine learning;
- A recharge threshold optimization balancing autonomy reserve and recharge frequency, with BDI 30–40% identified as the optimal trigger;
- A route simulation mode predicting SoC from GPS waypoints and an operational scenario, requiring no live telemetry.

Venelin Todorov was partially supported by the Centre of Excellence in Informatics and ICT under the Grant No BG16RFPR002-1.014-0018, financed by the Research, Innovation and Digitalization for Smart Transformation Programme 2021-2027 and co-financed by the European Union and by BNSF under Project KP-06-N82/4 “Artificial intelligence for investigation and modeling of real processes”.

II. RELATED WORK

A. Digital Twins for Railway

Ghaboura et al. [2] survey DTR applications across maintenance (34%), optimization (28%), and inspection, identifying IoT, AI, and BIM as enabling technologies. Rolling-stock DT work includes predictive maintenance of switch machines [8], track geometry prediction [9], and physics-based design of zero-emission heavy-haul locomotives [3]. None of these works address real-time SoC tracking for battery-electric shunting units under operational conditions.

B. SoC Estimation for Traction Batteries

SoC estimation has been extensively studied for lithium-ion EV batteries using LSTM networks [7] and ensemble methods [6]. A DT augmented with terrain elevation features is proposed in [4] for lithium cells, reporting NRMSE = 0.024. For lead-acid batteries, Coulomb counting with Peukert correction remains the industrial standard [11] and the reference against which data-driven methods are benchmarked. The present study demonstrates that for the confined station environment of a BESL, a well-calibrated two-parameter physical model matches or exceeds ML approaches that lack sufficient training data diversity.

C. Energy Management Optimization

Cipek et al. [5] demonstrate that gradient-aware energy management reduces consumption in hybrid locomotives. Coordinated scheduling studies [10] report 21.6% energy savings from SoC-aware route planning. The recharge threshold optimization proposed here is a complementary operational-level contribution: given a calibrated SoC model, the paper determines when to send the locomotive for recharge to maximize shift availability.

III. SYSTEM DESCRIPTION AND DATASET

A. ES3000 Locomotive

The ES3000, shown in Fig. 1, is a battery-electric shunting locomotive whose parameters are listed in Table I. Its flooded lead-acid traction battery requires periodic water top-up and exhibits significant Peukert capacity loss at the high currents used for wagon coupling and gradient starts.

TABLE I
ES3000 LOCOMOTIVE PARAMETERS

Parameter	Value
Unladen mass M_0	48 000 kg
Maximum speed	8 km/h
Traction motors	3 × 30 kW
Battery type	Flooded lead-acid
Battery configuration	2 × 1 376 Ah / 80 V
Total capacity C_{bat}	2 752 Ah
Nominal energy	≈ 220 kWh
Max discharge current	1 500 A
Telemetry	Teltonika FMB, 10-s interval



Fig. 1. ES3000 battery-electric shunting locomotive at an industrial site. A Teltonika FMB GPS tracker (roof-mounted) streams telemetry to the Wialon platform at 10-second intervals.

B. Telemetry Pipeline

Each locomotive is equipped with a Teltonika FMB GPS tracker that transmits to the Wialon IoT cloud platform. CSV exports provide: timestamp, GPS (lat, lon), speed v , terminal voltage U , total current I , BDI [%], odometer [km], altitude [m], traction command, brake command, and temperatures. A second PLC-based 4G system logs at 30-second intervals and additionally reports tachometer speed and motor temperatures; however its BDI field shows negligible variation ($< 1\%$) over short sessions and requires full-day discharge records to be useful for modelling.

C. Dataset

The primary dataset consists of 1 079 Wialon records from locomotive ES3018, covering a 3-hour operational session with BDI declining from 100% to 54% (Table II). Fig. 2 shows the complete session: speed, current, and BDI over time.

IV. PHYSICAL MODEL

A. Coulomb Counting

The BDI sensor implements Coulomb counting: it integrates the charge extracted from the battery and maps cumulative Ah to a percentage of nominal capacity. In discrete form at $\Delta t = 10$ s:

$$Ah_k = \sum_{j=1}^k I_j \frac{\Delta t}{3600}, \quad (1)$$

$$\widehat{BDI}_k = BDI_0 - \frac{Ah_k}{C_{\text{bat}}} \times 100\%. \quad (2)$$

At the nominal C/5 rate ($I_{\text{ref}} = 550$ A), $C_{\text{bat}} = 2752$ Ah gives a theoretical depletion rate of $100/2752 = 0.0363$ %/Ah.

TABLE II
DATASET SUMMARY — ES3018, WIALON, 01 APRIL 2026

Parameter	Value
Records	1 079
Sampling interval	10 s
Session duration	3.0 h
GPS coverage	100%
BDI start / end	100% / 54%
Total BDI change	46%
Total distance	9.16 km
Total Ah consumed	1 907 Ah
Energy consumed	≈153 kWh
Mean current (moving)	880 A
Median current (moving)	672 A
Mean speed (moving)	6.3 km/h
Moving fraction	87.4%
Altitude range	518–548 m

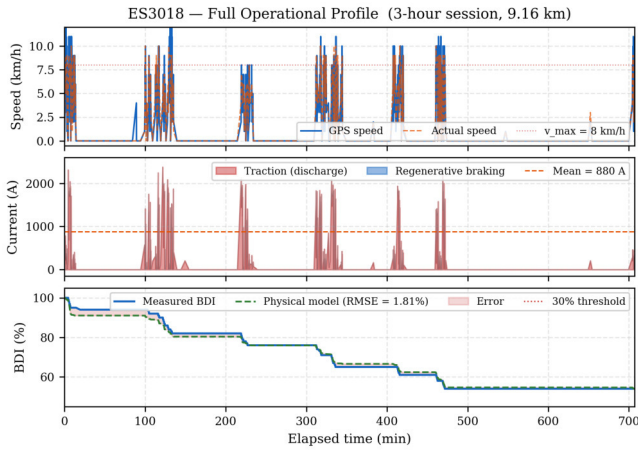


Fig. 2. Full 3-hour operational session of ES3018 (1 079 records, 9.16 km). *Top*: GPS and actual speed; dashed line marks $v_{\max} = 8$ km/h. *Middle*: total current; red = traction, grey = standing auxiliary loads, blue = regenerative braking; dashed line = mean traction current (880 A). *Bottom*: measured BDI (blue) and Peukert-corrected physical model (green, dashed); dotted line = 30% recharge threshold.

B. Peukert Correction

At high discharge currents the effective capacity of a flooded lead-acid battery is reduced [11]:

$$C_{\text{eff}}(I) = C_{\text{bat}} \left(\frac{I_{\text{ref}}}{I} \right)^{n-1}, \quad (3)$$

where n is the Peukert exponent. The corrected Ah increment is:

$$\Delta \text{Ah}_k^{\text{pk}} = \frac{I_k^n}{I_{\text{ref}}^{n-1}} \frac{\Delta t}{3600}. \quad (4)$$

C. Model Calibration from Operational Data

From the ES3018 session:

$$k_{\text{BDI}}^{\text{meas}} = \frac{46\%}{1\,907 \text{ Ah}} = 0.0241 \text{ \%/Ah}. \quad (5)$$

The Peukert factor:

Peukert Effect — Flooded Lead-Acid Battery (ES3000)

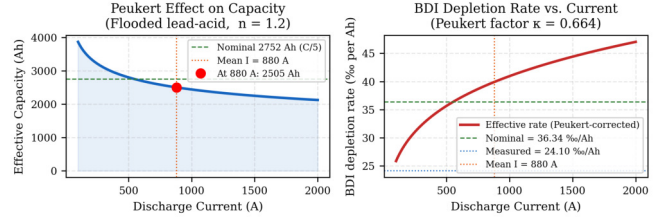


Fig. 3. *Left*: effective battery capacity vs. discharge current for $n = 1.20$. At the mean traction current of 880 A the effective capacity is reduced to 2 291 Ah (83% of nominal). *Right*: BDI depletion rate vs. current; dotted line = measured rate of 0.0241 %/Ah, consistent with Peukert factor $\kappa = 0.664$.

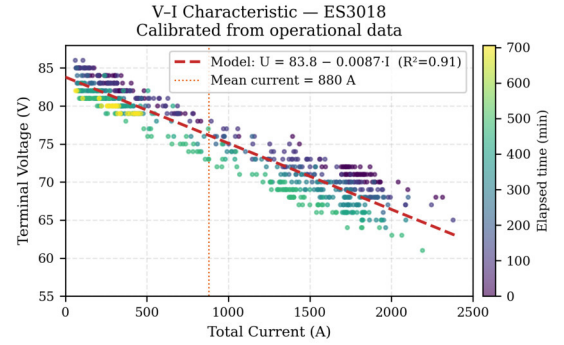


Fig. 4. V-I characteristic of the ES3018 traction battery. Scatter coloured by elapsed time; dashed line = calibrated model $U = 83.8 - 0.0087I$ ($R^2 = 0.91$). At the mean traction current (880 A) terminal voltage is 76.2 V; at maximum current (1 500 A) it reaches 70.8 V.

$$\kappa = \frac{k_{\text{BDI}}^{\text{meas}}}{k_{\text{BDI}}^{\text{nom}}} = \frac{0.0241}{0.0363} = 0.664, \quad (6)$$

corresponding to $n = 1.20$, which is the standard value for flooded lead-acid batteries [11] and confirms the physical consistency of the calibration. Fig. 3 illustrates the effect of $n = 1.20$ on effective capacity and BDI depletion rate.

D. Terminal Voltage Model

Linear regression on traction-current records ($I > 50$ A) yields:

$$U(I) = 83.8 - 0.0087I \text{ [V]}, \quad R^2 = 0.91. \quad (7)$$

Fig. 4 shows measured data overlaid with the fitted model and identifies two operating points relevant to mission planning.

Fig. 5 shows the evolution of current and voltage over the full session, with the calibrated voltage model overlaid.

V. CORRELATION ANALYSIS

Before model training or parameter fitting, we compute Pearson correlations between BDI and candidate features on the full 1 079-record dataset (Table III). Fig. 6 visualises the BDI-cumulative Ah relationship.

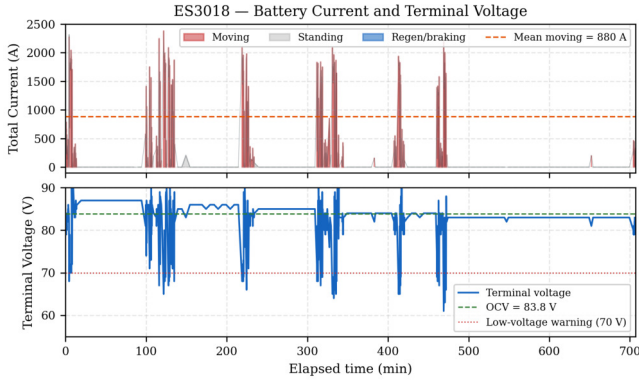


Fig. 5. Battery current (*top*) and terminal voltage (*bottom*) over the 3-hour ES3018 session. Current: red = traction, grey = standing auxiliary loads, blue = regenerative braking. Voltage: measured (blue) and model $U = 83.8 - 0.0087 I$ (red dashed).

TABLE III
PEARSON CORRELATION WITH BDI — ES3018 SESSION

Feature	r	Interpretation
Cumul. Ah (simple)	-0.9975	Primary driver
Cumul. Ah (Peukert)	-0.9975	Primary driver
Terminal voltage U	+0.0812	Weak — not used
Total current I	+0.1379	Weak — indirect
Speed v	+0.0766	Negligible

The result explains the failure of prior instantaneous-feature approaches: voltage ($r = 0.08$) and current ($r = 0.14$) alone carry no information about the accumulated energy state. Any useful BDI predictor must integrate current over time.

VI. RESULTS

A. Physical Model Accuracy

Fig. 7 compares the calibrated physical model against the measured BDI over the full ES3018 session. The model achieves:

- RMSE = 1.81%
- Maximum deviation < 3%
- No machine learning required — only two parameters: k_{BDI} and n

Table IV benchmarks the physical model against simpler approaches.

TABLE IV
SOC PREDICTION ACCURACY — ES3018 SESSION

Method	RMSE [%]	Notes
Instantaneous features (v, I, U)	> 10	No integration
Simple Coulomb counting	2.84	No Peukert
Peukert-corrected model (2)	1.81	Calibrated
Gradient Boosting (cum. Ah)	future work	Multi-file needed

Table V lists operational statistics derived from the calibrated model.

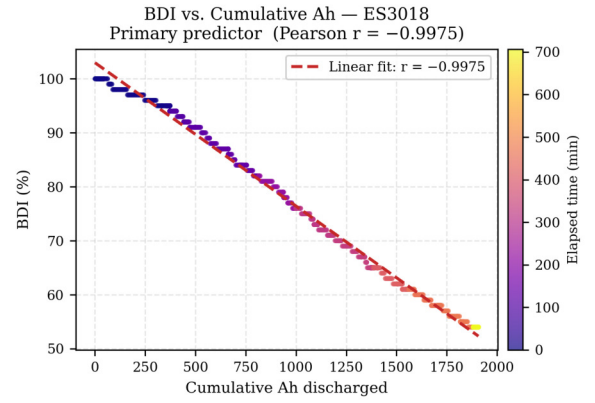


Fig. 6. BDI vs. cumulative Ah for the ES3018 session (Pearson $r = -0.9975$). Scatter coloured by elapsed time shows the monotonic discharge trajectory. The near-linear relationship confirms that Coulomb counting is the primary physical mechanism underlying the BDI. Deviations from linearity are attributed to the 1% discrete update resolution of the onboard indicator and Peukert losses at high currents.

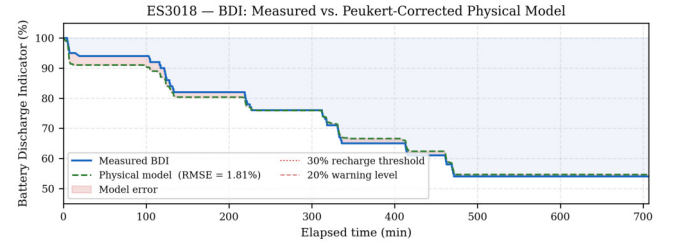


Fig. 7. Measured BDI (blue) vs. Peukert-corrected physical model (green dashed) for the 3-hour ES3018 session (RMSE = 1.81%, 1079 records). Shaded area = prediction error. Dotted lines mark the 30% recharge threshold and the 20% warning level. The step-wise appearance of the measured BDI reflects its 1%-resolution discrete update.

B. Recharge Threshold Optimization

A key operational question is: *at what BDI level should the locomotive be sent for recharge?* An aggressive threshold risks stranding the locomotive mid-shift; a conservative threshold wastes operational time on unnecessary recharges.

The remaining autonomy at threshold τ [%] is:

$$T_{\text{rem}}(\tau) = \frac{\tau / k_{BDI}}{I_{\text{mean}}} \times 60 \quad [\text{min}], \quad (8)$$

with $k_{BDI} = 0.0241 \text{ \%}/\text{Ah}$ and $I_{\text{mean}} = 880 \text{ A}$.

The effective 8-hour shift consumption at the measured duty cycle (87.4% moving, 12.6% standing):

$$I_{\text{eff}} = 0.874 \times 880 + 0.126 \times 30 \approx 773 \text{ A}, \quad (9)$$

$$\Delta BDI_{\text{shift}} = I_{\text{eff}} \times 8 \times k_{BDI} \approx 149\%. \quad (10)$$

Since $\Delta BDI_{\text{shift}} > 100\%$, a single charge cannot sustain a full 8-hour shift at the observed utilization, confirming that one mid-shift recharge is operationally necessary.

TABLE V
CALIBRATED OPERATIONAL STATISTICS — ES3018

Metric	Value
BDI depletion rate k_{BDI}	0.0241 %/Ah
Theoretical rate (C/5)	0.0363 %/Ah
Peukert factor κ	0.664
Peukert exponent n	1.20
Energy per km	16.7 kWh/km
Ah per km	208 Ah/km
BDI consumed per km	5.0 %/km
Full-charge autonomy	≈ 20 km

Recharge Threshold Optimization — ES3000

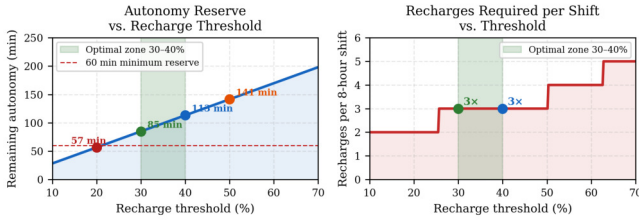


Fig. 8. Recharge threshold optimization for ES3000. *Left*: remaining autonomy (min) vs. recharge threshold (%); annotated values at key thresholds. *Right*: required recharges per 8-hour shift vs. threshold. Green band marks the recommended optimal zone of 30–40%, providing 85–113 min of reserve with one mid-shift recharge.

The optimization objective balances autonomy reserve (risk of stranding) against recharge frequency (availability loss):

$$\tau^* = \arg \min_{\tau} \alpha P_{\text{strand}}(\tau) + (1 - \alpha) N_{\text{charge}}(\tau), \quad (11)$$

where $\alpha \in [0, 1]$ is an operator-defined risk weight. Table VI quantifies the trade-off, and Fig. 8 shows the full optimization curves.

TABLE VI
AUTONOMY RESERVE VS. RECHARGE THRESHOLD

τ	Remaining Ah	Autonomy	Risk level
20%	830 Ah	57 min	High stranding risk
30%	1 245 Ah	85 min	Moderate — optimal
40%	1 660 Ah	113 min	Safe — optimal
50%	2 075 Ah	141 min	Conservative

For the ES3000 with $k_{\text{BDI}} = 0.0241$ %/Ah and $I_{\text{mean}} = 880$ A, the knee of the trade-off curve lies at $\tau^* = 30\text{--}40\%$, providing 85–113 minutes of reserve while requiring exactly one mid-shift recharge.

VII. ROUTE SIMULATION

For pre-shift mission planning the pipeline accepts: GPS waypoints of the planned route, an operational scenario (speed, payload, and mode per segment: *move* / *stand* / *brake*), and initial BDI. The simulation proceeds as:

- 1) Interpolate route at 10-m steps (Haversine formula);
- 2) Query elevation via Open-Elevation API (locally cached JSON to avoid repeated calls on the same route);

- 3) Compute gradient per segment: $i_k = (h_k - h_{k-1}) / d_k \times 1000$ [% $^\circ$];
- 4) Estimate traction current $\hat{I}_k = I_{\text{ref}}^{\text{move}} \cdot f_v(v_k) \cdot f_{\text{load}}(M_T) \cdot f_{\text{grad}}(i_k, M_T)$;
- 5) Integrate Peukert-corrected Ah (4);
- 6) Compute BDI trajectory from (2).

The traction model is calibrated from the ES3018 session: median moving current 672 A, scaled by speed and gradient; standing auxiliary load 30 A; payload scaling factors from Table II in Section IV. No live telemetry is required: input is GPS waypoints and an operational scenario only.

VIII. DISCUSSION

Why cumulative Ah dominates. The BDI is fundamentally an Ah-counting device. Any feature capturing this integral dominates over instantaneous signals, explaining why voltage ($r = 0.08$) and current ($r = 0.14$) fail as standalone predictors. The $r = -0.9975$ correlation confirms that Coulomb counting is the correct physical basis, and that Gradient Boosting or other ML methods can add value only by correcting the model residuals (Peukert non-linearity, temperature effects, ageing). With one training session these residuals cannot be robustly learned; additional multi-month fleet data are required.

BDI resolution. The onboard indicator updates in discrete 1% steps. Over the 3-hour session only 46 updates are recorded. This quantisation introduces an irreducible minimum RMSE of approximately 0.5% for any prediction method.

Gradient contribution. The 30-m altitude variation of the study route (518–548 m) produces measurable gradient effects on traction current. At the mean GPS step distance of 8.5 m, gradient resolution after smoothing is approximately $\pm 1\%$, which is sufficient for the simulation model but too noisy for per-step ML features without API elevation enrichment.

Single-session limitation. Results are calibrated on one 3-hour session of one locomotive. The Peukert exponent $n = 1.20$ is consistent with published values for flooded lead-acid batteries, providing confidence in generalizability; however cross-machine and cross-season validation requires the full fleet dataset.

IX. CONCLUSION

This paper presented a physics-based digital twin pipeline for SoC prediction and operational optimization of the ES3000 battery-electric shunting locomotive equipped with a flooded lead-acid traction battery. The central empirical finding is that cumulative Ah integration achieves a Pearson correlation of $r = -0.9975$ with the onboard BDI, far exceeding any instantaneous feature. A Peukert-corrected physical model calibrated from a single 3-hour real session achieves RMSE = 1.81% without machine learning, and self-consistently recovers the standard Peukert exponent $n = 1.20$ for flooded lead-acid cells. A recharge threshold optimization identifies BDI 30–40% as the recommended trigger for mid-shift charging, providing 85–113 minutes of autonomy reserve. A route

simulation mode enables pre-shift SoC forecasting from GPS waypoints and an operational scenario alone.

Future work will extend the dataset to multi-month logs from the full ES3000 fleet, enabling Gradient Boosting correction of Peukert residuals, temperature-aware capacity modelling, and cross-machine generalization validation.

REFERENCES

- [1] A. El Saddik, "Digital twins: The convergence of multimedia technologies," *IEEE MultiMedia*, vol. 25, no. 2, pp. 87–92, 2018, doi: 10.1109/MMUL.2018.023121167.
- [2] S. Ghaboura, R. Ferdousi, F. Laamarti, C. Yang, and A. El Saddik, "Digital twin for railway: A comprehensive survey," *IEEE Access*, vol. 11, pp. 120237–120257, 2023, doi: 10.1109/ACCESS.2023.3327042.
- [3] M. Spiryagin, E. Bernal, S. Ahmad, Q. Wu, C. Cole, and T. McSweeney, "Digital twin framework and platform for zero-emission heavy haul locomotive design and development," *Digital Twins and Applications*, vol. 1, no. 2, pp. 182–197, 2024, doi: 10.1049/dgt2.12017.
- [4] R. Issa, M. M. Badr, O. Shalash, A. A. Othman, E. Hamdan, M. S. Hamad, A. S. Abdel-Khalik, S. Ahmed, and S. M. Imam, "A data-driven digital twin of electric vehicle Li-ion battery state-of-charge estimation enabled by driving behavior application programming interfaces," *Batteries*, vol. 9, no. 10, Art. no. 521, 2023, doi: 10.3390/batteries9100521.
- [5] M. Cipek, D. Pavković, and Z. Kljaić, "Optimized energy management control of a hybrid electric locomotive," *Machines*, vol. 11, no. 6, Art. no. 589, 2023, doi: 10.3390/machines11060589.
- [6] M. A. Hannan, M. S. Hossain Lipu, A. Hussain, P. J. Ker, T. M. I. Mahlia, M. Mansor, A. Ayob, M. H. Saad, and Z. Y. Dong, "Toward enhanced state of charge estimation of lithium-ion batteries using optimized machine learning techniques," *Scientific Reports*, vol. 10, Art. no. 4687, 2020, doi: 10.1038/s41598-020-61464-7.
- [7] J. N. Njoku, E. C. Nkoro, R. M. Medina, C. I. Nwakanma, J. M. Lee, and D.-S. Kim, "Leveraging digital twin technology for battery management: A case study review," *IEEE Access*, vol. 13, pp. 21382–21412, 2025, doi: 10.1109/ACCESS.2025.3531833.
- [8] J. Yang, Y. Sun, Y. Cao, and X. Hu, "Predictive maintenance for switch machine based on digital twins," *Information*, vol. 12, no. 11, Art. no. 485, 2021, doi: 10.3390/info12110485.
- [9] J. Sresakoolchai and S. Kaewunruen, "Track geometry prediction using three-dimensional recurrent neural network-based models cross-functionally co-simulated with BIM," *Sensors*, vol. 23, no. 1, Art. no. 391, 2023, doi: 10.3390/s23010391.
- [10] M. T. Khalid, M. Yousif, S. A. A. Kazmi, M. U. Iftikhar, S. Alghamdi, and M. Waseem, "Coordinated scheduling and multi-energy storage planning for electric train operations," *Results in Engineering*, vol. 27, Art. no. 106751, 2025, doi: 10.1016/j.rineng.2025.106751.
- [11] W. Peukert, "Über die Abhängigkeit der Kapazität von der Entladestromstärke bei Bleiakкумуляtoren," *Elektrotechnische Zeitschrift*, vol. 18, no. 20, pp. 287–288, 1897.