Transport and Communications Science Journal

HYBRID PHYSICS-MACHINE LEARNING DIGITAL TWIN FOR HIGH-RESOLUTION TRACTION SUBSTATION POWER ESTIMATION

Tran Van Khoi[1,*], An Thi Hoai Thu Anh[1]

[1]University of Transport and Communications, No 3 Cau Giay Street, Hanoi, Vietnam

ARTICLE INFO

TYPE: Research Article
Received: 15/03/2026
Revised: 02/06/2026
Accepted: 03/09/2026
Published online: 15/09/2026
https://doi.org/10.47869/tcsj.77.7.8

[*] *Corresponding author*
Email: tvkhoi.ktd@utc.edu.vn

Abstract. Real-time estimation of traction power demand plays an important role in the operation and energy management of urban railway systems. This paper proposes a hybrid digital twin framework that integrates physics-based modeling with machine learning techniques to reconstruct real-time traction power profiles at traction substations using SCADA measurements of power and energy recorded at intervals of 3-5 minutes. The proposed digital twin framework consists of three main components. First, a physics-based load-flow model of the DC traction network is employed to ensure physical consistency in representing the electrical energy system of the railway line. Second, a residual learning module to identify the SCADA measurement instants within the physical model cycle, thereby enabling the prediction of substation power evolution in real time until the next measurement interval. Third, a post-prediction power correction module is implemented using an adaptive correction coefficient in order to minimize discrepancies between the model output and the actual system behavior. The proposed approach is validated using real operational data collected over several representative days from the Cat Linh - Ha Dong urban railway line. The verification results show that the proposed digital twin framework is capable of reconstructing high-fidelity real-time substation power profiles that closely follow the dynamic patterns observed in SCADA measurements. The framework successfully enforces strict energy conservation, driving the cumulative daily energy deviation to near zero, while maintaining robust generalization capability in tracking highly non-linear peak power demands throughout the entire operational duration across individual substations.

Keywords: Digital Twin, Traction Power System, Constrained Reconstruction, Model Calibration, Physics-informed

1. INTRODUCTION

Accurately determining traction power and energy consumption is crucial for urban metro management, as the traction system represents the largest infrastructure electricity consumer. These data enable operators to evaluate energy efficiency, optimize train timetables, maximize regenerative braking utilization, and forecast electrical loads. Furthermore, analyzing traction characteristics provides a foundation for assessing substation capacity, planning grid upgrades, and detecting operational anomalies. Given that traction energy accounts for the majority of total metro consumption, precise monitoring ultimately enhances operational efficiency, reduces costs, and ensures system reliability.

In practice, traction substations are monitored through Supervisory Control and Data Acquisition (SCADA) platforms. While these systems provide reliable supervisory information, their measurements are typically logged at relatively coarse time intervals-commonly every 3 to 5 minutes-due to communication bandwidth and system robustness constraints [1-4]. Such low-frequency sampling fundamentally limits system observability. Traction power demand varies rapidly as trains accelerate, coast, brake regeneratively, and interact electrically along the line. However, many internal states governing these dynamics are not directly measured [5]. The challenge thus extends beyond conventional forecasting: it requires reconstructing hidden system behaviors from sparse and aggregated observations.

Currently, the use of fixed empirical parameters combined with conventional algorithms to construct predictive train-parameter models for estimating train energy consumption is widely adopted [6]. Traditional traction power modeling methods are primarily based on physics-based formulations, which couple DC traction network load-flow calculations with train motion dynamics under different operational scenarios [7-9]. These approaches have proven valuable for system design and scenario-based analysis because they adhere to electrical laws and energy conservation principles. However, since many parameters related to the system load vary continuously in real operation, the estimation results obtained from such methods often exhibit significant deviations from actual values and are unable to accurately reflect the true operational state of the system [10].

In the broader context of engineering time-series data analysis, recent advanced deep learning frameworks have shifted toward high-dimensional feature extraction and residual sequence learning. For instance, Le-Xuan et al.[11] proposed a hybrid architecture integrating 1D-CNN, LSTM, and ResNet modules, demonstrating that the combination of convolutional layers for local spatial-frequency extraction and recurrent blocks for temporal dependencies significantly improves model robustness under complex engineering signals. Furthermore, to capture both localized transient variations and multi-scale temporal context, specialized encoder-decoder residual networks, such as the ResUNet4T model developed by Le-Xuan et al. [12], have been successfully utilized to decode technical sensor sequences from large-scale infrastructures. While these state-of-the-art architectures (e.g., CNN-LSTM-ResNet or U-Net-based time-series models) offer exceptional capability in purely data-driven anomalies or signal reconstruction, they fundamentally demand massive, continuous, high-resolution historical training datasets. In urban rail transit, however, historical SCADA logs are strictly constrained by sparse 3-5 minute snapshot sampling, which severely limits the direct implementability of such deep end-to-end networks.

In recent years, the development of Digital Twin (DT**)** technology through the integration of information technology, artificial intelligence, and physics-based modeling has enabled the digitalization and creation of highly accurate virtual replicas of physical systems. Such digital replicas are capable of reproducing the physical processes of a system in real time without requiring continuous real-time measurements from the physical infrastructure. By synchronizing a virtual representation with measurements obtained from the real system, Digital Twin frameworks aim to enhance situational awareness and improve the accuracy of system estimation [13, 14]. In railway applications, most DT implementations have concentrated on asset management, condition monitoring, and predictive maintenance tasks, while relatively limited research has addressed DT applications in energy management or real-time optimal control of railway power systems [6, 15-21]. For instance, Wang et al. [6] proposed a digital twin model of a metro traction power supply system for operational safety assessment in urban railways. Their results showed that the train power estimation error along a track section was reduced by 11.9% compared with conventional pure modeling approaches, while the passenger load forecasting model achieved a mean absolute percentage error (MAPE) of approximately 11.72-14.65%. Similarly, Guo et al.[20] proposed a digital twin modeling framework for urban rail transit power supply systems aimed at simulating the recovery and reuse process of regenerative braking energy within the traction power network. In addition, Li and Liu [21] developed a digital twin model of a traction power supply system to analyze energy flow processes when integrating photovoltaic (PV) generation into the traction network. Overall, the aforementioned studies generally construct digital twins based on predefined system attributes and attempt to encompass a wide range of possible operating conditions. However, such approaches still struggle to capture the stochastic characteristics inherent in real-world system operation. In practice, factors such as rapidly varying load factors, deviations between real operating conditions and idealized assumptions, and the absence of explicit mechanisms for systematic calibration make it difficult for the digital twin to remain accurately synchronized with the actual physical system.

Motivated by these limitations, this study aims to propose a digital twin framework for urban railway traction power supply systems with an explicit verification and calibration mechanism using power and energy measurements obtained from the SCADA system at 5-minute intervals. The proposed digital twin architecture adopts a physics-informed machine learning approach.

Within the proposed methodology, the physical model is constructed based on the dynamic characteristics of train motion to estimate the train power demand under a given load factor at each operating position. By combining train operating schedules with specific headway configurations, a traction power supply network model is established. A DC load-flow algorithm is then applied to determine the corresponding power demand of traction substations under the given train operation scenario and load factor. Subsequently, a machine learning-based module is introduced to align the simulated results with measurements obtained from the SCADA system. This module identifies the inference intervals based on the previously established physical model, generating estimated power profiles within each measurement cycle between two SCADA observations. The inferred results are then refined through the learning process to reduce estimation discrepancies. Finally, an energy-based correction module is applied to adjust the inferred results according to the ratio between the measured energy consumption from the SCADA system and the energy estimated by the model. This step ensures consistency between the reconstructed power profiles and the observed cumulative energy before producing the final digital twin outputs that reconstruct the traction substation power in near real time. The proposed digital twin model of the traction power supply system is implemented and validated in MATLAB using the design data of the Cat Linh - Ha Dong metro line urban railway. Validation is conducted using real measurement data from several representative operating days throughout the year in order to evaluate and calibrate the proposed method.

2. HYBRID DIGITAL TWIN ARCHITECTURE

The proposed Digital Twin architecture integrates a physics-based traction power model with a machine learning-driven correction mechanism to reconstruct second-level power profiles from SCADA data sampled at 5-minute intervals. This hybrid strategy preserves the physical consistency of the traction power system while enhancing estimation accuracy through data-driven adaptation to real operational conditions.

The architecture is structured into three primary functional layers:

+ Physical Layer: This layer consists of a DC traction power network model that simulates voltage distribution and power flow by solving the load-flow problem. It provides baseline power trajectories grounded in electrical laws and energy conservation principles.

+ Residual Learning Layer: This layer models the discrepancy between the results of physical simulation and the measurements obtained from the SCADA system. Machine learning algorithms are employed to learn the systematic deviations of the physical model and generate a residual correction term for the estimated results under real operational conditions. Through this process, the inference points for the temporal variation of power are determined and continuously propagated in real time until the next SCADA measurement becomes available. As a result, a preliminary digital twin reconstruction of the traction substation power profiles over time is obtained.

+ Adaptive Gain Calibration Layer: This layer performs a further calibration of the preliminary digital twin results based on the ratio between the actual energy consumption measured from the SCADA system and the energy calculated from the preliminary reconstructed results within the interval between two measurements. This procedure produces the final calibrated digital twin output.

The final hybrid Digital Twin power estimate is expressed as:

$$P_{hybrid}(t) = g(t)\,(P_{physical}(t) + f_\theta(X_t)) \tag{1}$$

Where $P_{physical}(t)$ denotes the power computed from the DC traction network model via load-flow analysis; $f_\theta(X_t)$ represents the machine learning-based residual estimator with trained parameters θ and input feature set X_t; and $g(t)$ is the adaptive time-varying gain designed to ensure energy consistency between the model output and SCADA measurements.

This three-layer architecture clearly separates physical simulation from data-driven correction. Such modular decomposition enables a balanced trade-off among engineering interpretability, operational robustness, and adaptability to real-world operating variability.

3. TRACTION POWER MODEL

3.1. Train power model

The traction power demand of a train is derived from its dynamic motion over an interstation segment, characterized by the evolution of speed, acceleration, and traction force throughout a complete running cycle.

At any instant t, the mechanical power delivered at the wheel-rail interface is expressed as:

$$P_{mech}(t) = F_{trac}(t)\, v(t) \tag{2}$$

where $v(t)$ is the instantaneous train velocity and $F_{\text{trac}}(t)$ denotes the net traction force.

The traction force is determined from Newton's second law and includes both inertial and resistive components:

$$F_{trac}(t) = ma(t) + F_{res}(t) \tag{3}$$

with m representing the train mass and $a(t)$ the longitudinal acceleration.

The total running resistance $F_{\text{res}}(t)$ accounts for rolling, aerodynamic, and gradient effects and is modeled as:

$$F_{res}(t) = A + Bv(t) + Cv^2(t) + mgsin\alpha \tag{4}$$

where A, B, and C are resistance coefficients, g is gravitational acceleration, and α is the track gradient angle.

The corresponding electrical power at the DC supply interface depends on the operating mode.

$$P_{train}(t) = \begin{cases} \dfrac{P_{mech}(t)}{\eta_{trac}}, & \text{traction} \\ -\eta_{regen} \mid P_{mech}(t) \mid, & \text{braking} \end{cases} \tag{5}$$

where η_{trac} and η_{regen} denote traction and braking efficiencies, respectively. During traction, electrical input power equals mechanical power adjusted by propulsion efficiency, whereas in regenerative braking, part of the mechanical energy is converted back to the network.

To ensure that the Digital Twin remains computationally efficient for real-time applications, a nominal train power trajectory $P_{\text{raw}}(t)$ is generated offline using a dynamic programming-based train performance simulator. This precomputed profile incorporates infrastructure characteristics (station spacing, substation layout, gradients, and curve radii), rolling stock parameters, scheduled travel time, and operational constraints such as maximum acceleration and speed limits.

During online operation, the nominal profile is scaled to reflect varying passenger loading and operational conditions through a load adjustment coefficient λ. Auxiliary consumption is added explicitly, yielding:

$$P_{train}(t) = \lambda\, P_{raw}(t) + P_{aux} \tag{6}$$

where P_{aux} represents onboard auxiliary demands, including HVAC, lighting, and control systems.

This formulation preserves the physical basis of train dynamics while enabling computationally efficient adaptation within the Digital Twin framework.

3.2. Traction power network model

The DC traction supply system is formulated as an equivalent node-branch electrical network. In this representation, traction substations are modeled as DC voltage source nodes with nominal voltage V_{sub}, while trains are treated as time-dependent power injection (or absorption) nodes. The contact line and return rail are described as distributed resistive links connecting adjacent nodes.

For each electrical section between nodes i and j, the longitudinal resistance is defined as:

$$R_{ij} = r_{line}\, l_{ij} \tag{7}$$

where r_{line} denotes the per-unit-length resistance and l_{ij} is the physical distance between the two nodes.

The entire network is assembled into a nodal admittance matrix G, whose elements are expressed as:

$$G_{ij} = \begin{cases} -\frac{1}{R_{ij}}, & i \neq j \\ \sum_{k \neq i} \frac{1}{R_{ik}}, & i = j \end{cases} \tag{8}$$

In the proposed DC load-flow formulation, the network nodes are classified into two distinct categories.

(i) Slack nodes (traction substations): The nodal voltage is prescribed and maintained at a fixed value, while the corresponding power injection is treated as a dependent variable determined through the load-flow solution.

(ii) PQ nodes (trains): The active power demand is specified based on the train operating condition, whereas the nodal voltage is considered an unknown state variable to be solved within the iterative procedure.

3.3. Substation power calculation

The DC power flow equations are solved iteratively using the Newton-Raphson algorithm.

At node i, the injected electrical power is computed as:

$$P_i(V) = V_i \sum_{j=1}^{n} G_{ij}\, V_j \tag{9}$$

where P_i is the specified nodal power (train demand or substation injection), V_i is the nodal voltage, and G_{ij} represents the admittance matrix element.

The power imbalance at node i is defined by:

$$\Delta P_i = P_i^{set} - P_i(V) \tag{10}$$

The iterative solution proceeds as follows:

Step 1 - Initialization:

$$V^{(0)} = V_{nom} \tag{11}$$

Step 2 - Iterative update

At iteration k, compute the nodal powers $P^{(k)}$ based on $V^{(k)}$, and evaluate the mismatch:

$$\Delta P^{(k)} = P^{set} - P^{(k)} \tag{12}$$

The Jacobian matrix is given by:

$$J_{ij} = \frac{\partial P_i}{\partial V_j} = \begin{cases} 2V_i G_{ii} + \sum_{m \neq i} V_m\, G_{im}, & i = j \\ V_i G_{ij}, & i \neq j \end{cases} \tag{13}$$

Step 3 - Voltage correction:

$$V^{(k+1)} = V^{(k)} + J^{-1} \Delta P^{(k)} \tag{14}$$

Step 4 - Convergence check:

$$max \mid \Delta P^{(k)} \mid < \varepsilon \tag{15}$$

Once convergence is achieved, the total power supplied by substations is obtained by summing the nodal injections at all slack nodes:

$$P_{sub} = \sum_{i \in \text{sub}} V_i \sum_j G_{ij}\, V_j \tag{16}$$

To construct a database of representative operating conditions, multiple simulation scenarios are conducted under varying traffic intensities. The number of trains ranges from 1 to 9, trains operate with uniform headways, and passenger load is fixed at 100% of rated capacity.

For each scenario, the resulting substation power trajectories are recorded as:

$$\boldsymbol{P}_{sim}^{(n)}(t) = [P_1(t), P_2(t), \dots, P_M(t)] \tag{17}$$

where n denotes the number of trains and M is the total number of substations.

These offline-generated time series constitute a library of physically coherent operating patterns, which later serve as reference profiles for SCADA-based matching within the Digital Twin framework.

4. DATA-DRIVEN ERROR COMPENSATION

4.1. Problem formulation

Although the DC load-flow-based Digital Twin provides a physically consistent baseline estimation of substation power, systematic discrepancies arise due to modeling simplifications, stochastic passenger variation, timetable perturbations, and unmodeled operational disturbances. Let $P_{\text{phy}}(t)$ denote the substation power obtained from the physics-based model at second-level resolution, and $P_{\text{SCADA}}(t_k)$ represent the measured power sampled at discrete 5-minute instants t_k. The reconstruction problem can be formulated as estimating a continuous high-resolution trajectory $\hat{P}(t)$ such that it remains physically coherent with traction network constraints, while also minimizing the discrepancy with SCADA measurements at the sampling instants and preserving cumulative energy consistency within each sampling interval.

To achieve these objectives, a structured three-stage compensation mechanism is proposed: (i) Residual estimation at SCADA instants; (ii) Temporal reconstruction within intervals; (iii) Energy-consistent adaptive calibration. The structural diagram of the model is presented in Figure 1.

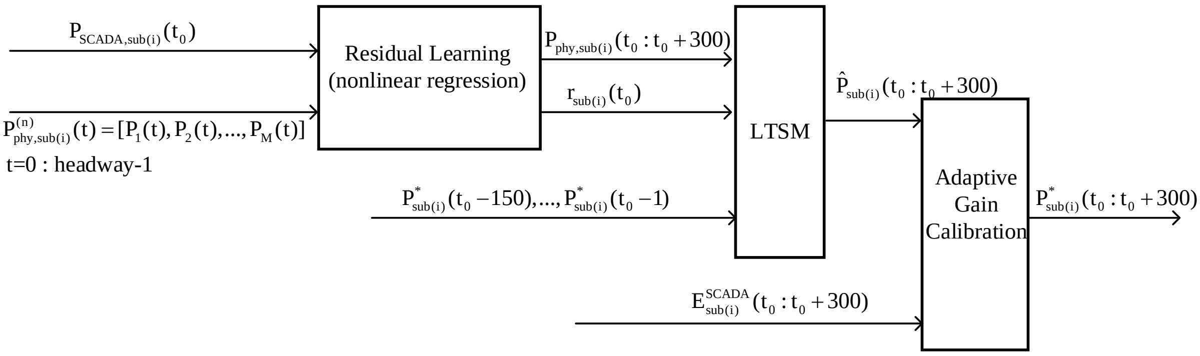

Figure 1. SCADA Data-Based Model for Estimating Traction Substation Power.

4.2. Residual learning at measurement instants

Instead of directly predicting absolute power values, the methodology adopts a residual modeling strategy. The residual at each SCADA instant is defined as:

$$r(t_k) = P_{\text{SCADA}}(t_k) - P_{\text{phy}}(t_k) \tag{18}$$

This decomposition ensures that machine learning only models systematic mismatch rather than the full physical process, thereby preserving interpretability and preventing violation of electrical constraints.

The residual is approximated by a nonlinear regression function:

$$\hat{r}(t_k) = f_\theta(x_k) \tag{19}$$

where x_k is a feature vector that includes the physics-based estimated power, the number of operating trains, headway information, encoded time-of-day indicators, and operational state flags indicating whether the system is in an active or idle period.

After performing the regression process, the achieved optimal results ensure that the total deviation across all substations is minimized, provided that the maximum deviation at each substation does not exceed a specified threshold (the limit threshold established in this study is 500 kW). Concurrently, the SCADA measurement instants are identified and synchronized with the corresponding time steps in the simulation cycle of the physics-based model.

$$t_0 = \min \sum_{i=1}^{M} \hat{r}(t_k) \tag{20}$$

This synchronization allows for the preliminary determination of the baseline power profile on a per-second basis for the subsequent 300 seconds. The physics-based baseline power profile is determined starting from the initial time t_0 within the simulation cycle until the end of the cycle period, and can wrap around back to the beginning of the cycle if necessary until the full duration of t_0 *+300* seconds is satisfied.

$$\hat{P} = P_{phy}(t_0 : t_0 + 300) \tag{21}$$

4.3. LSTM-based temporal reconstruction

The primary objective of this stage is to reconstruct the high-resolution power trajectory during the 300-second intervals between consecutive SCADA measurement scans. Rather than relying on conventional mathematical interpolation methods which typically fail to capture the highly non-linear train dynamics such as acceleration, coasting, and regenerative braking, this study proposes a hybrid Deep Learning architecture utilizing a Stacked Long Short-Term Memory (LSTM) network. Instead of functioning as a purely autoregressive predictor, the network establishes a hybrid sequence-to-sequence mapping by concurrently integrating three distinct data streams: short-term historical context, future physical baseline guidelines, and the instantaneous system residual at the boundaries.

At each SCADA measurement instant *t*, a comprehensive hybrid input feature vector *X* is formulated via sequence concatenation as follows:

$$X(t) = \left[P^*(t - 150 : t - 1), P_{phy}(t : t + 300), \hat{r}(t)\right] \tag{22}$$

Where $P^*(t - 150 : t - 1)$ represents the historical 150-second values; $P_{phy}(t : t + 300)$ denotes the 300-second future baseline power profile derived from the physics-based model; $\hat{r}(t)$ is the absolute power residual measured at the traction substation at instant t, providing a boundary anchor to calibrate the structural bias of the model.

To effectively process these multi-source inputs while preventing overfitting and maintaining real-time computation capabilities, a Deep Stacked LSTM architecture consisting of two hidden layers with 32 hidden units per layer is deployed. The finalized generalized mathematical mapping for Stage 2 is denoted as follows:

$$\hat{P}(t : t + 300) = LSTM\left(P^*(t - 150 : t - 1), P_{phy}(t : t + 300), \hat{r}(t)\right) \tag{23}$$

The resulting output vector provides a smoothed, high-resolution (1-second) power profile. This profile inherently preserves train kinematics owing to the internal memory retention of the stacked hidden layers, while strictly adhering to the global physical trends dictated by the future baseline guidance.

4.4. Energy-constrained adaptive gain calibration

Even after residual correction, cumulative energy mismatch may persist. To enforce energy conservation over each SCADA interval, an adaptive gain mechanism is introduced. For interval k, the gain is determined according to:

$$G(t) = \frac{E_{SCADA}(t_{5min})}{\int_{t_0}^{t_0+300} \hat{P}(t)\, dt} \tag{24}$$

where $E_{\mathrm{SCADA}}(t_{5\mathrm{min}})$ is the measured energy during the 5-minute interval, and the denominator represents the energy computed from the physical model over the same duration.

Rather than maintaining G_k as a stepwise constant, the gain is allowed to vary smoothly between consecutive intervals through interpolation:

$$G(t) = G_k + \alpha(t)(G_{k+1} - G_k) \tag{25}$$

where $\alpha(t) \in [0,1]$ is a time-dependent interpolation factor.

The final reconstructed power is:

$$P^*(t) = G(t)\hat{P}(t) \tag{26}$$

By embedding this adaptive calibration mechanism, the hybrid Physical-ML Digital Twin achieves a balanced compromise between physical stability and data-driven responsiveness under fluctuating operating conditions.

5. SIMULATION RESULTS

The performance of the proposed framework was validated using a simulation-based case study implemented in the MATLAB environment. The underlying railway system model follows the configuration reported in [22]. The investigated metro corridor extends over 12.6 km and includes 12 passenger stations: Cat Linh, La Thanh, Thai Ha, Lang, Thuong Dinh, Vanh Dai 3, Phung Khoang, Van Quan, Ha Dong, La Khe, Van Khe, and Yen Nghia. Power is delivered through six traction substations positioned along the route at Cat Linh, Lang, Vanh Dai 3, Van Quan, La Khe, and Yen Nghia. Key electrical specifications, network topology, and rolling stock parameters are compiled in Table 1. Operational timing parameters are provided in Table 2, including interstation running times and end-to-end journey durations [23]. In the simulation setup, the departure instant at the origin station is set to t=0 s, and a uniform dwell time of 30 s is assumed at every intermediate station. Under these assumptions, the total travel time from the first to the last station is 1292 s, while the reverse-direction journey requires 1285 seconds.

Based on the principal electrical parameters and network topology summarized in Table 1, together with the train running times along the corridor presented in Table 2, the instantaneous traction power demand of each train is determined. In order to evaluate the power according to Equation (1), the traction force and velocity profiles are obtained using an empirical approach consistent with the methodology described in [24]. For a predefined number of trains operating, a dispatch schedule is generated to define their entry times and headways along the line. Subsequently, the DC load-flow algorithm outlined in Section 3.2 is executed to compute the power supplied by each traction substation at successive time instants within a headway cycle.

Figure 2 depicts the temporal evolution of substation power under two representative operating conditions: (i) six trains in service with a 600 s headway, and (ii) nine trains with a reduced headway of 360 s. These scenarios illustrate how variations in traffic density influence the dynamic power demand of the traction power system.

The same procedure is repeated for operating scenarios involving between one and nine trains. The corresponding simulation outputs are stored in a structured database in the form P_{sim}, where each entry represents a specific traffic density condition. This database serves as the reference input for the Digital Twin under operating plans with comparable headways. If the timetable or headway configuration changes, the simulation process must be re-executed and the database updated to preserve model fidelity.

Table 1. The main electrical and topological characteristics.

Parameters	**Value**
Line length	12.6 km
Station location	0; 0.93; 1.81; 2.89; 4.14; 5.15; 6.45; 7.75; 9.08; 10.18; 11.62; 12.60
Substation location	0; 2.89; 5.15; 7.75; 10.18; 12.60
Nominal voltage	750 VDC
Train mass	145 ton
Auxiliary power	200 kW
Resistance per unit of contact line and track	0.065 Ω/km
Resistance of feeder	0.01 Ω

Table 2. The journey time of the train along the railway line.

Departure station	**Arrival station**	**Depart time (sec)**	**Arrival time (sec)**	**Journey time (sec)**	**Rest time (sec)**
1	2	0	88	88	30
2	3	118	196	78	30
3	4	226	317	91	30
4	5	347	450	103	30
5	6	480	559	79	30
6	7	589	693	104	30
7	8	723	809	86	30
8	9	839	936	97	30
9	10	966	1050	84	30
10	11	1080	1181	101	30
11	12	1211	1292	81	30
12	11	1292	1372	80	30
11	10	1402	1503	101	30
10	9	1533	1617	84	30

9	8	1647	1744	97	30
8	7	1774	1859	85	30
7	6	1889	1995	106	30
6	5	2025	2103	78	30
5	4	2133	2237	104	30
4	3	2267	2355	88	30
3	2	2385	2464	79	30
2	1	2494	2577	83	30

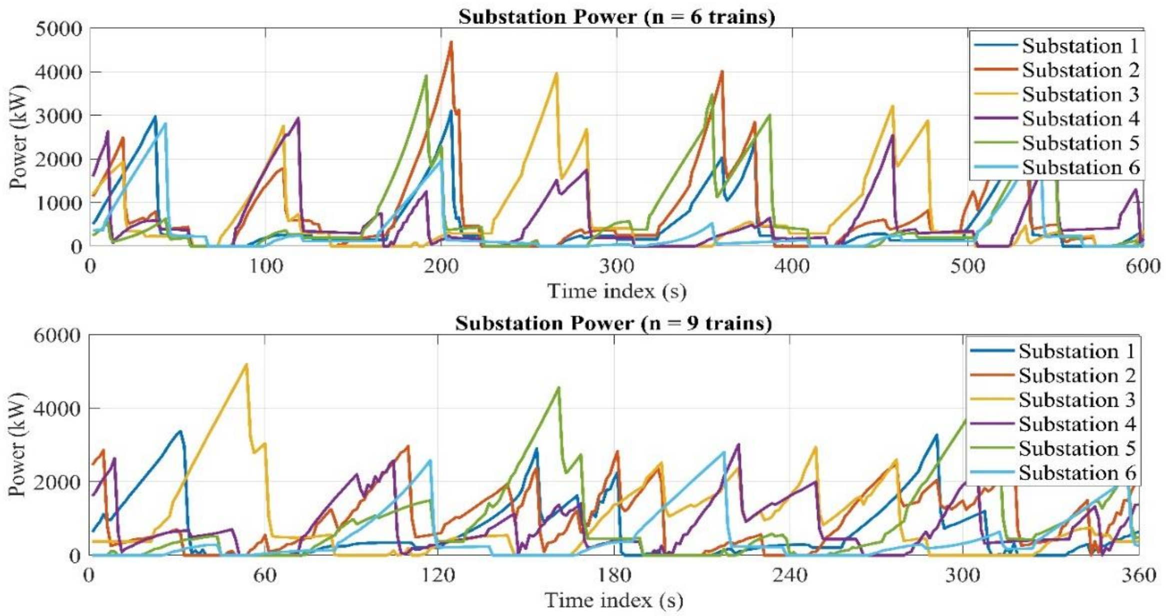

Figure 2. Substation power of the simulation model under operation with 6 trains and 9 trains.

Figure 3 presents the matching results between simulated substation power profiles and SCADA measurements for a representative day on September 5, 2025. At each measurement instant, a vector comprising the power values of all six substations is compared against the simulated profiles, considering the scheduled train dispatch times. The matching process identifies the simulated scenario that minimizes the aggregated deviation across substations, thereby determining the inferred substation power trajectory for the corresponding 300-second interval. From the graphical comparison, it can be observed that the reconstructed power profiles consistently converge toward the measured values. This behavior indicates that the developed model captures the fundamental load characteristics and physical properties of the traction power system. The remaining discrepancies primarily reflect stochastic operational factors, such as variations in passenger boarding and alighting at individual stations, which influences train mass and traction demand. In addition, real-world speed profiles deviate from the idealized patterns assumed in simulation, further contributing to the observed residual differences.

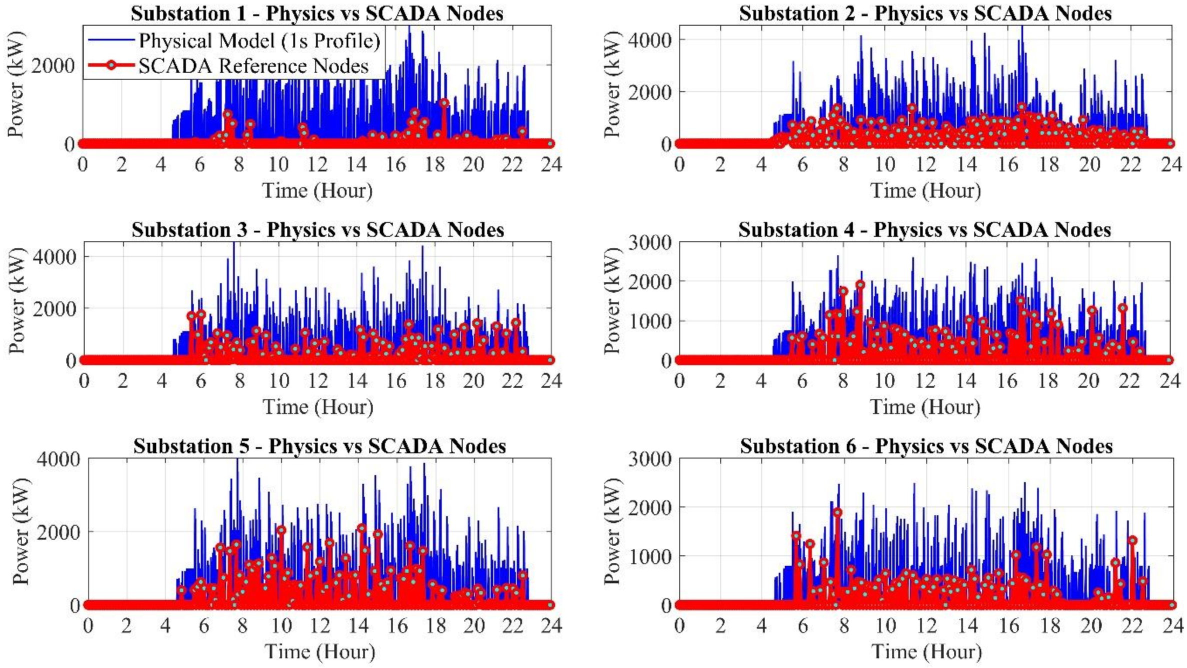

Figure 3. Comparison of SCADA and the physical model power profiles on September 5, 2025.

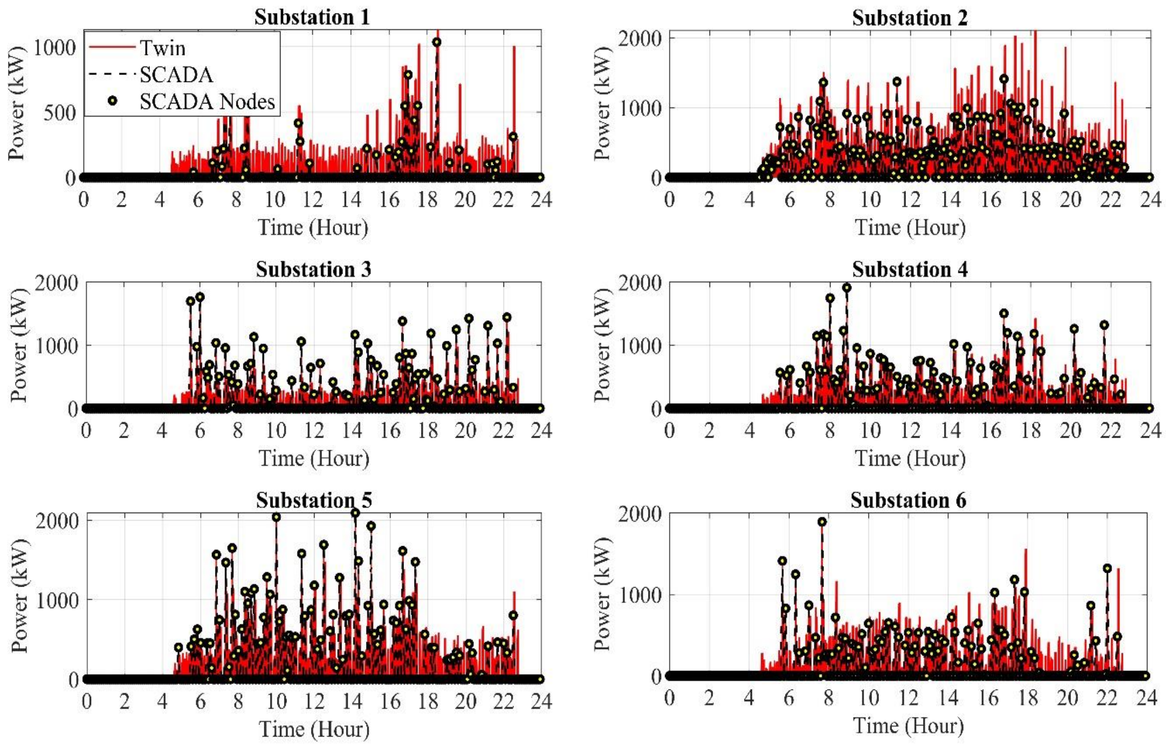

Figure 4. Comparison of SCADA and TWIN power on September 5, 2025.

Figure 4 illustrates the reconstructed substation power profiles generated by the Digital Twin framework incorporating the proposed machine learning algorithm over a 24-hour period on September 5, 2025. The reconstruction is based on the operating states identified during the profile-matching process shown in Figure 3, together with the measured cumulative energy recorded for each SCADA sampling interval. Overall, the proposed model successfully reproduces the temporal evolution of load demand across all six traction substations. The characteristic daily operating pattern is clearly reflected: high-demand intervals coincide with peak service hours, whereas low or near-zero power levels correspond accurately to non-operational periods. The occurrence, frequency, and temporal distribution of power peaks exhibit strong consistency between reconstructed outputs and measured data, indicating that the model effectively captures the dynamic load behavior induced by train movements. Although minor local deviations in magnitude are observed at certain time instants primarily attributable to simplified physical assumptions and the omission of some stochastic operational factors that the reconstructed curves maintain a high degree of qualitative agreement with SCADA measurements. These results demonstrate that Digital Twin provides a reliable representation of the overall load behavior of the traction power system, thereby establishing a solid foundation for applications such as energy monitoring, operational performance assessment, and decision support in urban railway power system management.

Figure 5 presents a comparison of cumulative energy between SCADA measurements and the values estimated by the Digital Twin model. By applying an energy-ratio-based correction mechanism that compensates for the discrepancy between simulated and measured consumption, the model is able to closely track actual operating conditions regardless of whether SCADA sampling instants coincide with characteristic operating events. As a result, the total daily energy deviation is reduced to nearly zero for all traction substations. This confirms that the proposed calibration strategy effectively enforces the principle of energy conservation, which is essential for operational planning and energy management applications.

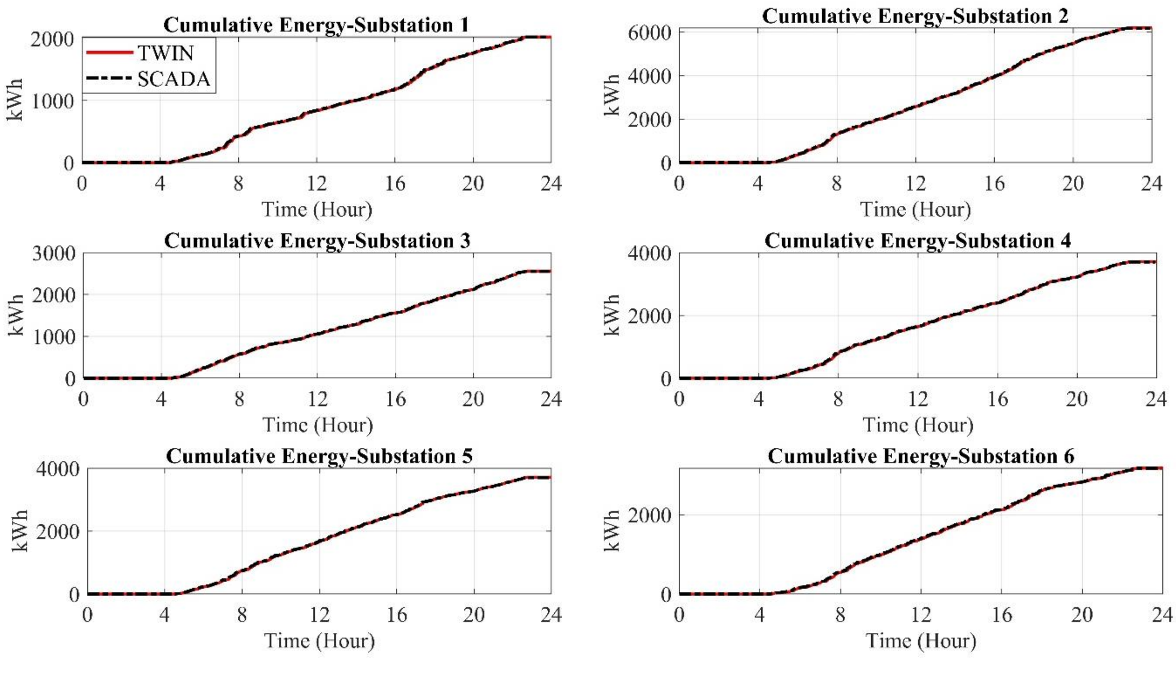

Figure 5. Comparison of cumulative substation energy between SCADA and TWIN on 5/9/2025.

Figure 6 illustrates the real-time (1-second resolution) power discrepancy between the TWIN and the physical model throughout the operation day of September 5, 2025. Inherently, a deviation between the TWIN and the idealized physical model persists; however, this discrepancy fluctuates around the 0 kW baseline, demonstrating the high tracking fidelity of the TWIN relative to the physical profile. The magnitude of this error is predominantly governed by load variations and operational train scenario mismatches in practice. For instance, the difference during acceleration between an empty train (zero passenger load and minimal auxiliary power) and a fully loaded train (maximum passenger capacity and peak auxiliary load) can exceed 1000 kW. Under high headway conditions (such as 9-10 operational trains on the Cat Linh - Ha Dong line), a single traction substation concurrently supplies power to 2 to 4 trains within its designated section. Consequently, the idealized model yields noticeable deviations compared to actual operational circumstances. Regarding the profile for September 5, the idealized physical model exhibits significantly higher power demands than the TWIN, which can be attributed to the relatively low passenger load demand on this specific day. Furthermore, peak discrepancies are densely concentrated during specific intervals, namely 5:00-7:00, 8:00-8:30, 16:30-17:30, and 20:00-20:30. These timeframes correspond to the periods when trains are being dispatched onto or withdrawn from the main line according to the official train regulation orders. Since these transient phases involving variations in the number of active trains cannot be precisely formulated within a deterministic physical model, the occurrence of larger errors during these intervals is reasonable and anticipated.

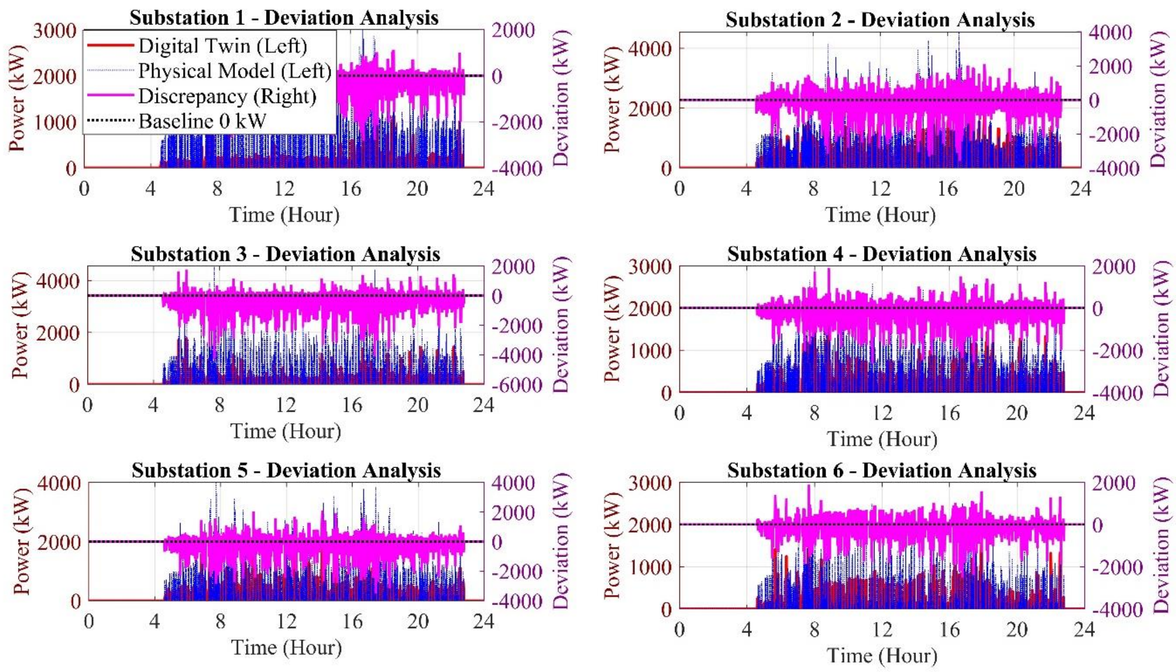

Figure 6. Power error between the TWIN and the physical model of the substations.

Figure 7 illustrates the power discrepancies between the TWIN and the SCADA benchmarks specifically at the corresponding discrete SCADA measurement nodes. With the primary objective of calibrating the errors to ensure exact boundary synchronization at both ends of each 300-second block between the TWIN and the SCADA data, the machine learning model has successfully addressed this requirement. As demonstrated in the figure, the discrepancy values at these validation nodes remain exactly zero throughout the entire operational duration across all six traction substations.

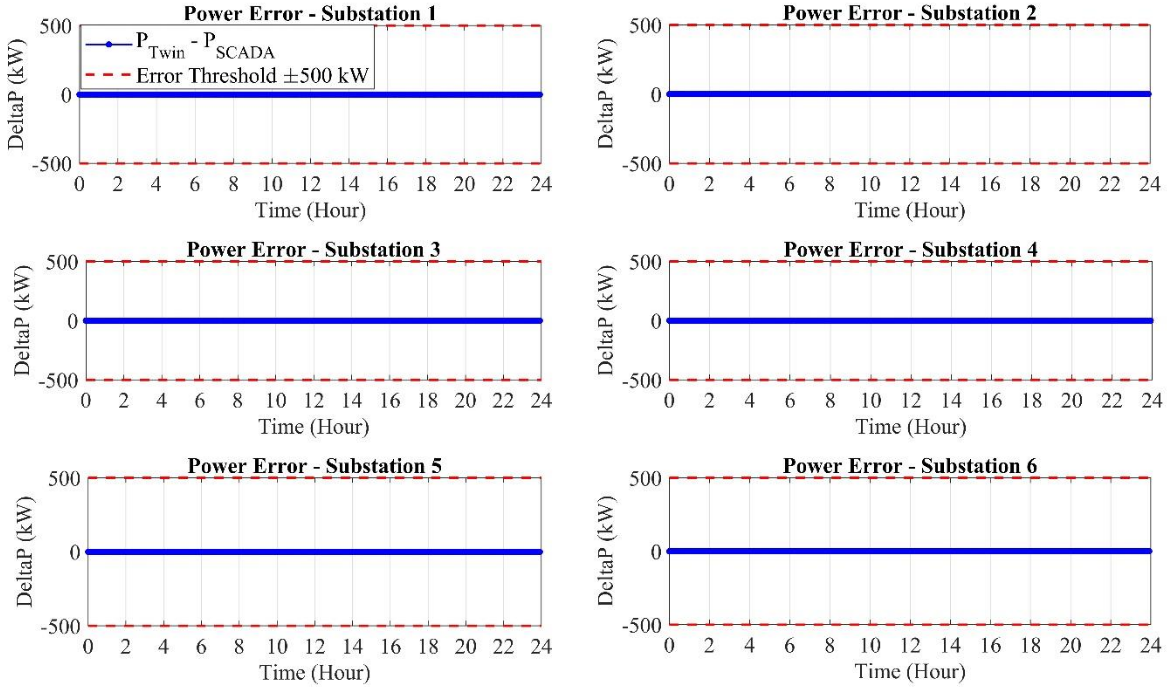

Figure 7. Power error between the TWIN and the SCADA of the substations.

The obtained results indicate that the proposed Digital Twin framework is capable of effectively reconstructing the traction substation power demand while preserving long-term energy consistency during system operation. Although short-term estimation errors may arise under highly dynamic operating conditions, the approach still provides a practical and dependable solution for energy-oriented metro system operation.

Compared with Digital Twin modeling approaches that rely purely on physical models, which must account for the full range of system state variables under all possible operating scenarios resulting in highly complex models and substantial computational requirements while the proposed method only needs to represent a set of representative scenarios with fundamental parameters. As a result, the overall model structure remains relatively simple, and the computational burden is significantly reduced. In addition, the proposed framework is capable of adaptive calibration to accommodate variations encountered in real-world operation and stochastic parameter changes, whereas purely physics-based models generally lack the ability to adjust to such variations once formulated. Compared with fully data-driven modeling approaches, the proposed method also offers several distinct advantages. First, it preserves physical interpretability, ensuring that the model output remains consistent with the underlying physical processes governing system behavior. Second, it requires only a relatively modest amount of training data, since the machine learning component is mainly used to correct residual errors rather than to predict the entire system behavior. Third, the proposed framework can be integrated directly into a real-time SCADA environment without requiring extensive additional sensing infrastructure. Furthermore, the block-wise calibration strategy establishes a pragmatic balance between modeling fidelity and computational efficiency, enabling the Digital Twin to maintain sufficient accuracy while remaining operationally feasible for real-world deployment scenarios.

6. CONCLUSION

This paper presents a hybrid Digital Twin framework that integrates physical knowledge with machine learning techniques to reconstruct high-resolution traction substation power profiles from low-frequency SCADA data. The proposed architecture consists of three core components: a physical model of the energy processes within the traction power supply system, a residual learning module that determines the power inference cycle of substations based on SCADA measurements while correcting discrepancies between simulation results and corresponding real operational measurements, and an energy-matching calibration module designed to reduce estimation errors caused by non-representative measurement points. Validation using real operational data demonstrates that the proposed Digital Twin model is capable of accurately capturing the variation trends of traction load demand, while maintaining the accuracy of cumulative energy and reducing short-term estimation errors under conditions of limited system observability. The proposed approach achieves a balanced integration of physical interpretability and data-driven adaptability, thereby providing a reliable foundation for energy management, monitoring, and operational optimization in urban railway traction power supply systems.

REFERENCES

[1]. Dimitrios Pliatsios, Panagiotis Sarigiannidis, Thomas Lagkas, Antonios G. Sarigiannidis, A Survey on SCADA Systems: Secure Protocols, Incidents, Threats and Tactics, IEEE Communications Surveys & Tutorials, 22 (2020) 1942-1976. https://doi.org/10.1109/COMST.2020.2987688

[2]. Shize Huang, Lingyu Yang, Ji Xue, Kai Yu, Urban Rail Transit Power Monitoring System Techniques Based on Synchronous Phasor Measurement Unit, in Proceedings of the 4th International Conference on Electrical and Information Technologies for Rail Transportation (EITRT), 2019, Springer Singapore, 251-268.

[3]. Marcel Nicola, Claudiu-Ionel Nicola, Florin Teișanu, Constantin Chelan, Automation System of Railway Level Crossings and SCADA Integration, International Journal on Engineering Applications, 7 (2019). https://doi.org/10.15866/irea.v7i3.17261

[4]. Zheng Pan, Liang Che, Chunming Tu, Pseudo-measurement-based state estimation for railway power supply systems with renewable energy resources, IET Generation, Transmission & Distribution, 18 (2024) 871-880. https://doi.org/10.1049/gtd2.13120

[5]. Shaofeng Lu, Stuart Hillmansen, Tin Kin Ho, Clive Roberts, Single-Train Trajectory Optimization, IEEE Transactions on Intelligent Transportation Systems, 14 (2013) 743-750. https://doi.org/10.1109/TITS.2012.2234118

[6]. Ling Wang, Xiang Chen, Feng Ding, Digital twin modeling and intelligent optimization for rail operation safety assessment, International Journal for Simulation and Multidisciplinary Design Optimization, 15 (2024) 1-10. https://doi.org/10.1051/smdo/2024002

[7]. Hammad Alnuman, Daniel Gladwin, Martin Foster, Electrical Modelling of a DC Railway System with Multiple Trains, Energies, 11, 3211, 2018. https://doi.org/10.3390/en11113211

[8]. Hwanhee Cho, Jaewon Kim, Hosung Jung, Hyungchul Kim, Simultaneous DC Railway Power System Analysis Method Using Model-Based TPS, Applied Sciences, 12, 6929, 2022. https://doi.org/10.3390/app12146929

[9]. Xinyang Yu, Xin Wang, Yuxin Qin, Urban Rail System Modeling and Simulation Based on Dynamic Train Density, Electronics, 13, 853, 2024. https://doi.org/10.3390/electronics13050853

[10]. Yulong Che, Xiaoru Wang, Leijiao Ge, Hongjian Lin, Xiaoqin Lyu, Hongsheng Su, Hao Wang, A review of research on traction load models and modeling methods for electrified railways, Renewable and Sustainable Energy Reviews, 219 (2025). https://doi.org/10.1016/j.rser.2025.115869

[11]. Thang Le-Xuan, Thanh Bui-Tien, Hoa Tran-Ngoc, A novel approach model design for signal data using 1DCNN combing with LSTM and ResNet for damaged detection problem, Structures, 59, 105784, 2024. https://doi.org/10.1016/j.istruc.2023.105784

[12]. Thang Le-Xuan, Thanh Nguyen-Chi, Thanh Bui-Tien, Hoa Tran-Ngoc, ResUNet4T: A potential deep learning model for damage detection based on a numerical case study of a large-scale bridge using time-series data, Engineering Structures, 340, 120668, 2025. https://doi.org/10.1016/j.engstruct.2025.120668

[13]. Michael Grieves, John Vickers, Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems, in Transdisciplinary Perspectives on Complex Systems, 2017, Springer, 85-113. https://doi.org/10.1007/978-3-319-38756-74

[14]. Yuqian Lu, Chao Liu, Kevin I-Kai Wang, Huiyue Huang, Xun Xu, Digital Twin-driven smart manufacturing: Connotation, reference model, applications and research issues, Robotics and Computer-Integrated Manufacturing, 61, 101837, 2020. https://doi.org/10.1016/j.rcim.2019.101837

[15]. Sara Ghaboura, Rahatara Ferdousi, Fedwa Laamarti, Chunsheng Yang, Abdulmotaleb El Saddik, Digital Twin for Railway: A Comprehensive Survey, IEEE Access, 11 (2023) 120237 - 120257. https://doi.org/10.1109/ACCESS.2023.3327042

[16]. Sylwia Werbińska-Wojciechowska, Robert Giel, Klaudia Winiarska, Digital Twin Approach for Operation and Maintenance of Transportation System-Systematic Review, Sensors, 24, 6069, 2024. https://doi.org/10.3390/s24186069

[17]. Dharmendra Kushwaha, Ankit Kumar, S. P. Harsha, Advancements and applications of digital twin in the railway industry: a literature review, International Journal of Rail Transportation, 13 (2025) 865-890. https://doi.org/10.1080/23248378.2024.2434834

[18]. Evelin Krmac, Boban Djordjevic, Digital Twins for Railway Sector: Current State and Future Directions, IEEE Access, 12 (2024) 108597 - 108615. https://doi.org/10.1109/ACCESS.2024.3439471

[19]. Zi-Yang Zhang, Du Shang, Shuai Su, Digital twin in railway industry: a bibliometric analysis and systematic review, Digital Twin, 2 (2025) 1-25. https://doi.org/10.1080/27525783.2025.2533858

[20]. Jian Guo, Xiaobo Wu, Hong Liang, Junfeng Hu, Baoming Liu, Digital-twin based Power Supply System Modeling and Analysis for Urban Rail Transportation, in IEEE International Conference on Energy Internet (ICEI), 2020. https://doi.org/10.1109/ICEI49372.2020.00022

[21]. Xin Li, Yingzhi Liu, A Hybrid Data and Mechanism Model-driven Digital Twin Modelling Approach for Novel Traction Power Systems, Urban Rail Transit, 11 (2025) 221-233. https://doi.org/10.1007/s40864-024-00236-2

[22]. Tran Van Khoi, An Thi Hoai Thu Anh, Dang Viet Phuc, An optimizing method of inverter location and capacity on the urban rail power supply system, Transport and Communications Science Journal, 72 (2021) 536-551. https://doi.org/10.47869/tcsj.72.5.3

[23]. Tran Van Khoi, An Thi Hoai Thu Anh, Optimization of supercapacitor energy storage systems and solar power systems integrated into urban railway lines to lessen CO2 emissions, Archives of Electrical Engineering, 74 (2025) 583-602. https://doi.org/10.24425/aee.2025.154985

[24]. Tran Van Khoi, An Thi Hoai Thu Anh, Dang Viet Phuc, Optimizing the urban train speed to minimize the energy consumption and comfort, Transport and Communications Science Journal, 72 (2021) 317-329. https://doi.org/10.47869/tcsj.72.3.7