

Physics-constrained multi-task energy consumption prediction for the safe operation of embodied mobile robots

Zhilong Lv¹, Shiqi (Shawn) Ou^{1,2,*}, Shaoqin He³, Yi Jiang⁴, Hao Jing^{1,2}, Chen Zhang⁵

¹ School of Future Technology, South China University of Technology, Guangzhou, Guangdong 511442, China

² Guangdong Artificial Intelligence and Digital Economy Laboratory (Guangzhou), Guangzhou, Guangdong 510335, China

³ Guangzhou EHot Intelligent Technology Co., Ltd., Guangzhou, Guangdong 511442, China

⁴ MOUME Technology (Suzhou) Co., Ltd., Suzhou, Jiangsu 215163, China

⁵ College of Smart Energy, Shanghai Jiao Tong University, Shanghai 200240, China

* Corresponding author: sou@scut.edu.cn (S.O.)

CITATION:

Lv, Z., Ou, S. S., He, S., Jiang, Y., Jing, H., & Zhang, C. (2027). Physics-constrained multi-task energy consumption prediction for the safe operation of embodied mobile robots. *Applied Energy*, 427, 129151. <https://doi.org/10.1016/j.apenergy.2026.129151>

This work is supported by the National Key Research and Development Program of China (2024YFE0115800), the Introduced Innovative R&D Team of Guangdong (2023ZT10L145), the Guangdong Basic and Applied Basic Research Foundation (2025A1515140131), and the Fundamental Research Funds for the Central Universities. All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of sponsors.

Abstract

Reliable task-level energy assessment is essential for sustained mission execution and safe operation of battery-powered mobile robots under dynamic conditions. State of charge (SOC) and fixed-capacity estimates provide limited information about the effective energy reserve under changing electrical, thermal, and locomotion conditions. This study proposes a two-level indicator system comprising task-available energy (TAE) and a predictive relative risk margin. TAE characterizes the energy reserve available for a specific task under current operating conditions, extending conventional SOC-based assessment toward task-conditioned energy availability. An 85-day operational dataset covering five gait modes is segmented into task fragments. The proposed physics-constrained multi-task model combines categorical embeddings, stacked long short-term memory encoding, temporal attention, and current-state fusion to jointly predict TAE, future energy consumption, future minimum voltage, and future maximum battery temperature. On the independent test set, normalized root-mean-square error (NRMSE) values of 1.34%, 1.60%, 0.68%, and 0.59% are obtained for the four targets, respectively. Compared with the strongest energy-related baselines, the NRMSE values for TAE and future energy are reduced by 72.4% and 65.6%, respectively. The predictive relative risk margin integrates energy, voltage, and thermal constraints, achieving a coefficient of determination of 0.993, a risk-level agreement of 94.01%, and a dominant-risk-factor accuracy of 93.30%. This work reframes battery management as a task-aware predictive capability that integrates energy state assessment, operational risk evaluation, and autonomous decision-making, enabling safer and more reliable operation of mobile robots in dynamic environments.

Keywords: Energy prediction; Mobile robot; Operational decision support; Physics-constrained multi-task learning; Task available energy

1. Introduction

The rapid convergence of artificial intelligence, embodied intelligence, advanced sensing, and robotic hardware is accelerating the development of autonomous robots for complex real-world environments [1],[2]. These systems increasingly integrate perception, learning, and action beyond structured operating settings. Reliable autonomy, adaptive mobility, and sustained operation are consequently becoming increasingly important for deployment in dynamic field environments [3]. Legged mobile robots, particularly quadruped platforms, are increasingly considered for inspection, exploration, logistics, and emergency response because their mobility enables access to rough and unstructured environments that remain difficult for conventional mobile platforms [4]. However, the operational endurance of battery-powered mobile robots remains constrained by limited onboard energy, while locomotion, perception, computation, gait transitions, and terrain adaptation impose strongly time-varying electrical loads. State of charge (SOC) provides a capacity-related estimate, whereas the deliverable energy under actual operating conditions is also governed by discharge rate, terminal voltage, temperature, battery aging, and coupled electro-thermal behavior [5]-[10]. These characteristics make SOC alone insufficient for representing the battery capability available to support an ongoing robot task. This operating variability creates a need for task-conditioned assessment of energy availability, voltage response, and thermal behavior to support operational-margin evaluation and reliable mission execution.

Recent advances in quadruped robotics have substantially expanded locomotion capability across complex and changing environments. Perceptive locomotion, agile motor skills, challenging-terrain adaptation, multi-expert control, and high-speed locomotion have improved robustness across diverse operating conditions [11]-[15]. Research on deformable-terrain traversal, agile parkour, deep tracking control, and high-speed navigation further extends robot operation under rapidly changing terrain and motion conditions [16]-[19]. Locomotion versatility is also enhanced through morphology-dynamics adaptation, end-to-end policy learning, pretrained motor representations, viability-driven gait transitions, and bio-inspired gait strategies [20]-[24]. Related studies have further improved robotic performance through energy-efficient actuation, passive mechanical coupling, and adaptive multimodal mobility [25]-[27]. These advances continuously broaden the range and intensity of robot operation, while also exposing the onboard battery to rapidly varying energy demand, voltage loading, and thermal conditions across different tasks.

Energy prediction and battery management have received increasing attention in mobile-robot research. Physics-based, data-driven, and hybrid models have been developed to characterize robot energy consumption [28],[29]. Mission-energy prediction has been further improved through online identification of terrain-dependent power-model parameters, enabling energy estimation to adapt to changing ground conditions, although the principal objective remains traversal-energy estimation [30]. A multi-parameter predictive model incorporated SOC, travel distance, turning frequency, payload, velocity, and battery temperature to estimate battery discharge for mission planning, while future voltage and thermal responses remain outside the predictive formulation [31]. The influence of terrain interaction has also been examined through motion power losses associated with wheel slip and changing operating conditions, revealing substantial variations in locomotion energy demand without extending the analysis to battery-level available energy [32]. Hybrid power-consumption modeling further accounts for unknown and time-varying payloads, improving adaptation to dynamic mechanical demands while retaining power consumption as the primary prediction quantity [33]. At the system level, mechanical and computational energy have been jointly considered to coordinate locomotion velocity and onboard processing resources [34]. This approach extends energy management beyond locomotion alone, although future battery energy reserve, voltage response, and thermal accumulation are not jointly represented.

Beyond system level energy coordination, battery management research has increasingly focused on improving the estimation of battery states under dynamic operating conditions. Neural network-based state of energy (SOE) estimation has been developed to capture the nonlinear relationship between measurable battery signals and available energy [35]. Optimization-assisted adaptive filtering has also improved the accuracy and robustness of SOC estimation under varying operating conditions [36]. Joint SOC and SOE estimation has also been explored through adaptive dual filtering and core factor feedback correction, enabling charge and energy states to be estimated within a coordinated framework [37]. These developments provide a more informative description of battery condition and energy availability, while their primary focus remains battery state estimation rather than the task dependent energy reserve and future electro thermal behavior required for mobile robot operation. In addition, task-centric battery management has connected robot characteristics, mission requirements, and battery states, establishing an important link between battery management and robot tasks [38]. However, effective battery energy reserve, future energy

demand, voltage response, and thermal behavior have rarely been jointly characterized under task-varying robot operation, limiting their direct use for integrated operational assessment and decision support.

In summary, previous studies reveal a persistent disconnect among battery energy-state assessment, robot energy-demand prediction, and task-level operational decision support. Battery-oriented studies have progressively extended conventional state-of-charge estimation toward available-energy and remaining-discharge-energy assessment under varying electrical and thermal conditions. In contrast, robot-oriented studies primarily characterize power consumption and mission energy demand associated with terrain, payload, locomotion, and computational activities. Task-centric battery management further establishes the importance of linking battery states with robot operating requirements. However, a quantitatively validated pathway that connects robot task conditions with effective battery energy reserve, future electro-thermal responses, and interpretable operational decisions remains insufficiently developed. In particular, available energy, future energy demand, voltage decline, and thermal accumulation are rarely represented within a unified predictive formulation, and their joint influence on task-level operational risk remains largely unexplored. To address this gap, this study develops a task-conditioned energy assessment and operational decision-support framework for mobile robots, as illustrated in **Fig. 1**. Continuous operating logs are segmented into task fragments that preserve the temporal relationships among battery, motion, thermal, and operating states. Task-available energy (TAE) is introduced to characterize the effective energy reserve under the current operating condition. A physics-constrained multi-task model (PCMT) then jointly predicts TAE, future energy consumption, minimum voltage, and maximum battery temperature. These predictions are integrated into a predictive relative risk margin for interpretable operational assessment and decision support. The main contributions of this study are summarized as follows:

- (1) A task-fragment-based energy and electro-thermal characterization method is developed for continuous mobile robot operating logs. Functional stages, including charging, standby, locomotion, task execution, and return, are identified, and task-level features are constructed to describe energy consumption, voltage variation, thermal response, and motion conditions.
- (2) This study proposes both TAE and a predictive relative risk margin as a two-level indicator system for task-conditioned energy assessment and operational risk evaluation. TAE achieves a normalized root mean square error (NRMSE) of 1.34%, a coefficient of determination (R^2) of 0.995, and a mean absolute percentage error (MAPE) of 1.12%.

- (3) The proposed physics-constrained multi-task model (PCMT) jointly predicts TAE, future energy consumption, minimum voltage, and maximum battery temperature through nonnegative physical reconstruction and reliability-aware supervision. Relative to the strongest baselines, TAE NRMSE decreases from 4.85% to 1.34%, while future-energy NRMSE decreases from 4.65% to 1.60%.
- (4) A three-level safety-oriented decision-support strategy converts predicted energy, voltage, and thermal margins into relative operational-risk levels and corresponding recommendations. On the test set, risk-level agreement reaches 94.01%, optimistic misclassification is limited to 5.61%, and dominant-risk-factor accuracy reaches 93.30%.

The remainder of this study is organized as follows. Section 2 describes the experimental setup and dataset, including the mobile robot platform, data acquisition system, operating conditions, and dataset characteristics. Section 3 presents the proposed methodology. Section 4 presents the prediction results and performance evaluation. Section 5 concludes this study by summarizing the main findings, discussing the current limitations, and outlining directions for future research.

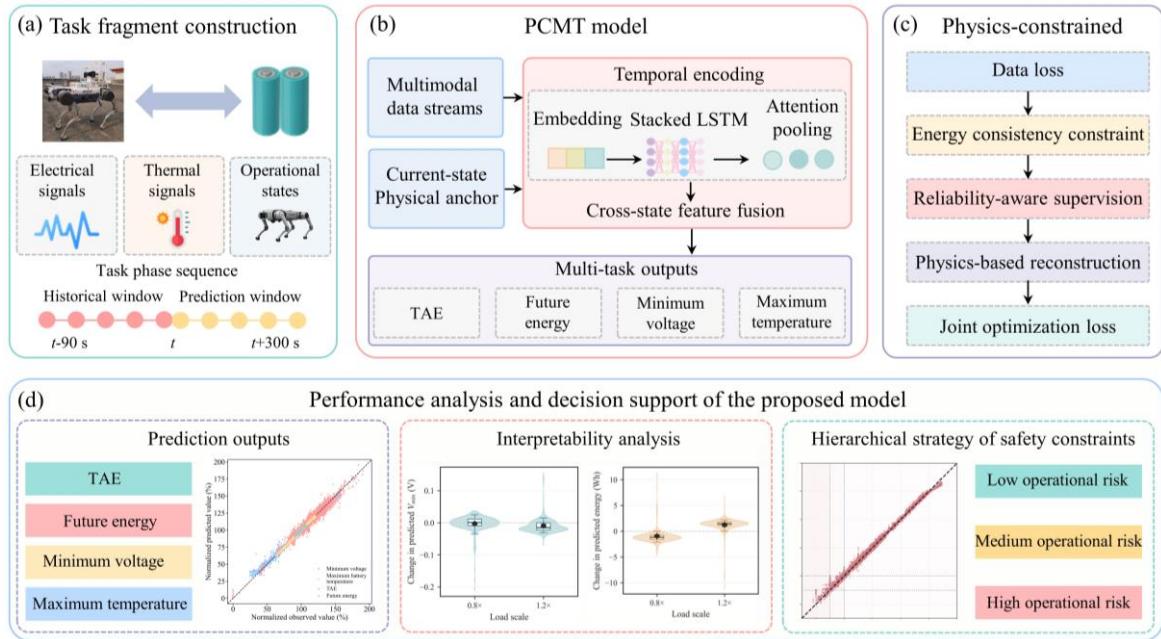


Fig. 1. Overall framework of the proposed method. (a) Task-fragment construction from electrical, thermal, and operational signals using historical and prediction windows. (b) PCMT architecture integrating multimodal sequences and current-state physical anchors through embedding, stacked LSTM, attention pooling, and cross-state feature fusion to jointly predict TAE, future energy consumption, minimum voltage, and maximum temperature. (c) Physics-constrained learning combining data loss, energy-consistency

regularization, reliability-aware supervision, physics-based reconstruction, and joint optimization. (d) Model evaluation through prediction-performance analysis, interpretability analysis, and hierarchical operational-risk assessment for decision support.

2. Experimental setup and dataset description

2.1. Experimental setup

The experimental platform used in this study is a quadruped robot equipped with onboard perception, navigation, communication, and state-monitoring modules, as shown in **Fig. 2**. The robot contains four electrically actuated legs and an upper mounting structure for the installation of sensing and auxiliary devices. Perception units are installed on the front and upper sections of the robot to support environmental observation, localization, and autonomous navigation. The onboard monitoring system continuously records the electrical, motion, thermal, and operational states of the robot during task execution. The experiments are conducted in practical operating environments, including an outdoor paved test area with sufficient space for locomotion and navigation tasks. The representative experimental site shown in **Fig. 2** consists of a relatively level concrete surface in an open rooftop environment. Additional operating scenarios are included to cover rough-terrain traversal, slope and anti-slip locomotion, and stair-climbing tasks. These environments provide different locomotion demands and actuator loading conditions, thereby increasing the diversity of the collected operating data.

During data acquisition, the robot is operated under normal task procedures and executes navigation and locomotion commands according to the assigned mission. The operating states are automatically recorded by the onboard monitoring and navigation systems and are stored as timestamped daily log files. The collected signals include battery voltage and current, charging status, locomotion speed, position, orientation, gait command, terrain mode, joint temperature, body temperature, and task-related status information. All variables are aligned according to their timestamps to preserve the temporal relationships among the electrical, mechanical, thermal, and control states. Data are collected over 85 operating days. Both stationary and locomotion periods are retained, together with the transitions among charging, standing, walking, terrain traversal, and stair-climbing states.

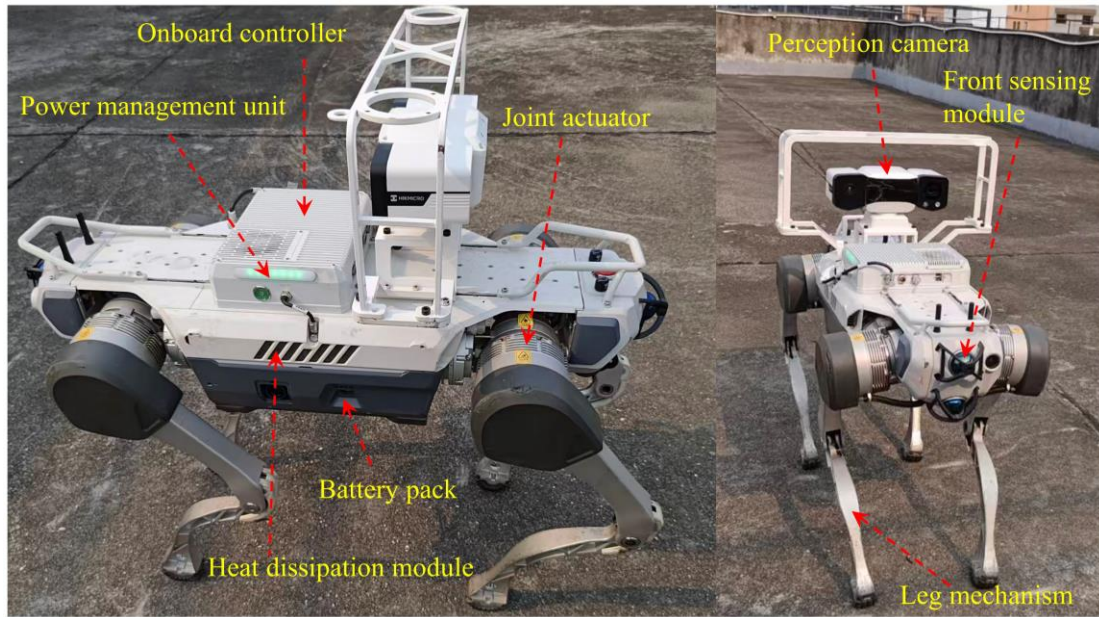


Fig. 2. Quadruped robot used for experimental data collection.

2.2. Data description

The dataset used in this study is constructed from onboard monitoring and navigation logs recorded during routine operation of a quadruped robot in practical task environments. Data from a total of 85 operating days are included. The raw logs are organized by operating date and arranged chronologically according to the recorded timestamps. Each observation represents a multivariate state snapshot of the robot at a specific sampling instant. The recorded variables describe the electrical, motion, thermal, and operational states of the robot, including battery voltage, battery current, SOC, locomotion speed, position, orientation, joint temperature, body temperature, gait command, terrain mode, charging status, and task status. The main characteristics and overall ranges of the key variables are summarized in **Table 1**.

The dataset contains stationary periods, locomotion periods, charging processes, and transitions among different operating states. Five representative gait modes are covered, including walking, rough terrain, slope, stair sensing, and continuous stairs. The distributions of the principal operating variables under these gait modes are presented in **Fig. 3**. Specifically, **Fig. 3(a)** to **(f)** show the distributions of battery voltage, discharge current magnitude, SOC, body temperature, maximum joint temperature, and battery temperature, respectively. The gait-based grouping allows the electrical load, battery condition, and thermal response of the robot to be compared under different locomotion conditions. Together, the statistical results indicate that the dataset covers broad battery states, electrical loads, and thermal conditions under diverse practical tasks.

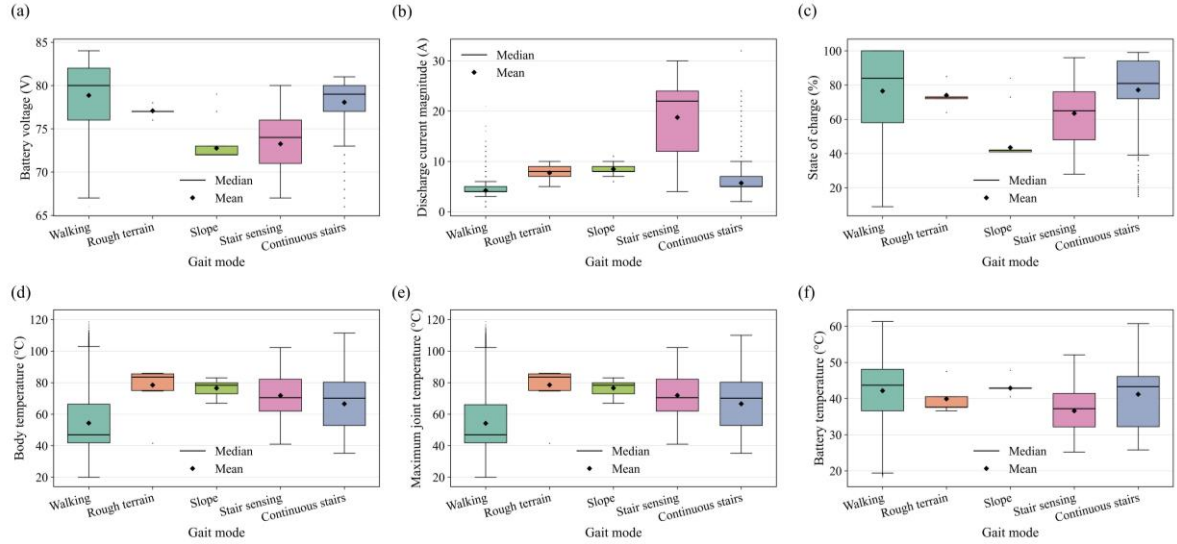


Fig. 3. Distributions of key operating variables under different gait modes. (a) battery voltage, (b) discharge current magnitude, (c) SOC, (d) body temperature, (e) maximum joint temperature, and (f) battery temperature. Center lines and diamonds indicate the median and mean, respectively.

Table 1. Main characteristics of the quadruped robot dataset.

Category	Description	Unit	Statistical summary
Dataset coverage	Operating period	day	85
Electrical state	Battery voltage	V	66 to 84
Electrical state	Battery current	A	-33 to 6
Battery state	SOC	%	9 to 100
Motion state	Locomotion speed	m/s	0 to 2.39
Thermal state	Battery temperature	°C	18 to 61.6
Thermal state	Maximum joint temperature	°C	18.7 to 119.79

3. Methodology

3.1. Data preprocessing and task fragment construction

Continuous operating logs are processed to establish synchronized descriptions of battery, motion, thermal, and task states. The recorded variables include SOC, voltage, current, battery temperature, body temperature, velocity, odometry, spatial position, heading, gait, pavement type, control mode, charging state, docking state, and return state. Charging, standby, locomotion, task execution, and return stages are identified from the recorded operating and motion signals. These stages are used to characterize differences in energy consumption and electro-thermal responses under different robot functions. Invalid timestamps,

duplicate records, and samples with missing essential variables are removed. Each continuous log is divided when the time interval between adjacent measurements exceeds 5 s. The remaining segments are resampled at 1 Hz. Numerical variables are reconstructed through time interpolation, while categorical variables are filled according to the nearest valid operating state. SOC is converted to a fraction within [0,1], and discharge power is denoted by Eq. (1).

$$P_{\text{dis}}(t) = V(t)I_{\text{dis}}(t) \quad (1)$$

where $P_{\text{dis}}(t)$ is the discharge power, $V(t)$ is the terminal voltage, and $I_{\text{dis}}(t)$ is the nonnegative discharge current. A sliding-window strategy is applied to construct task fragments. Each sample contains a historical observation window of 90 s and a prediction horizon of 300 s. Adjacent samples are generated at intervals of 10 s. The task fragment and the corresponding extrema targets are denoted by Eq. (2).

$$\begin{aligned} S_i &= \{\mathbf{X}_{t_i-89:t_i}, \mathbf{x}_{t_i}, \mathbf{y}_{t_i+1:t_i+300}\}, \\ V_{\min,H}(t_i) &= \min \left\{ V(t_i), \min_{\tau \in (t_i, t_i+H]} V(\tau) \right\}, \\ T_{\max,H}(t_i) &= \max \left\{ T(t_i), \max_{\tau \in (t_i, t_i+H]} T(\tau) \right\}. \end{aligned} \quad (2)$$

where S_i is the i -th task fragment, $\mathbf{X}_{t_i-89:t_i}$ is the historical feature sequence, $\mathbf{x}_{t_i}$ is the current-state vector, and $\mathbf{y}_{t_i+1:t_i+300}$ represents the future supervision information used for target construction. $V_{\min,H}(t_i)$ and $T_{\max,H}(t_i)$ denote the minimum battery voltage and maximum battery temperature over the prediction horizon H , respectively. $V(t_i)$ and $T(t_i)$ are the battery voltage and battery temperature at the current prediction point, while τ denotes the time within the prediction interval. The current voltage and battery temperature are explicitly included as the initial states of the extrema evaluation interval, whereas the subsequent future measurements are used only for label construction and are excluded from the model inputs.

The constructed task fragments are divided into training, validation, and test sets using approximate proportions of 70%, 15%, and 15%, respectively. The partition is performed according to complete source files or recording dates rather than by randomly dividing overlapping windows. This grouped partition prevents task fragments constructed from the same raw operating period from appearing in different subsets. When temporal blocks from the same continuous record are used, an isolation interval of at least 390 s is maintained between adjacent subsets. This interval corresponds to the combined duration of the historical window of 90 s and the prediction horizon of 300 s. The resulting date ranges, fragment counts, and retained

operating modes for each subset are summarized in **Table 2**. Because complete recording dates are preserved during partitioning, the final subset proportions differ slightly from the nominal 70%, 15%, and 15% allocation. The calibration set used to determine the data-driven reference boundaries is drawn from the validation subset and therefore does not constitute an additional partition. All calibration-derived boundaries are fixed before test-set evaluation, and no test labels are used for calibration.

The model inputs are divided into categorical historical features, continuous historical features, and current-state features. The categorical features include gait, pavement type, task phase, operating mode, and control mode. The continuous historical input comprises 21 variables describing the electrical, thermal, motion, spatial, and operational states of the robot. These variables include the SOC fraction S_{soc} , battery voltage V_{bat} , battery current I_{bat} , discharge power P_{dis} , battery temperature T_{bat} , maximum body temperature T_{body} , robot speed S_{spd} , odometry increment D_{odo} , spatial coordinates x , y , and z , heading components H_{sin} and H_{cos} , heading variation H_{Δ} , charging indicator B_{chg} , and docking indicator B_{dock} . Short-term dynamic information is further represented by the 30 s voltage variation $V_{\Delta 30}$, 30 s battery-temperature variation $T_{\Delta 30}$, 60 s mean discharge power $P_{\mu 60}$, 60 s standard deviation of discharge power $P_{\sigma 60}$, and 60 s rolling minimum voltage $V_{\text{min}60}$. The current-state vector contains the latest SOC, voltage, current, battery temperature, mean discharge power, speed, heading components, charging state, and docking state at the prediction time. The Pearson correlation matrix of the 21 continuous historical features is shown in **Fig. 4**. Strong correlations are observed among SOC, battery voltage, and recent minimum voltage, as well as among discharge power and its rolling statistics. Charging and docking indicators also exhibit similar correlation patterns because docking is closely associated with charging operations.

Table 2. Temporal partitioning, fragment counts, and operating-mode distribution across dataset subsets.

Subset	Date range	Fragments	Walking	Rough terrain	Slope	Stair sensing	Continuous stairs
Training	2026/01/11– 2026/05/04	41529	41170	2	0	57	300
Validation*	2026/05/06– 2026/05/09	9812	9788	0	0	0	24
Calibration*	2026/05/06– 2026/05/09	9812	9788	0	0	0	24
Test	2026/05/10– 2026/05/19	9837	9781	0	8	0	48
Total	2026/01/11– 2026/05/19	61178	60739	2	8	57	372

*The calibration set corresponds to the validation subset used for determining the data-driven reference boundaries and is not counted as an additional dataset partition.

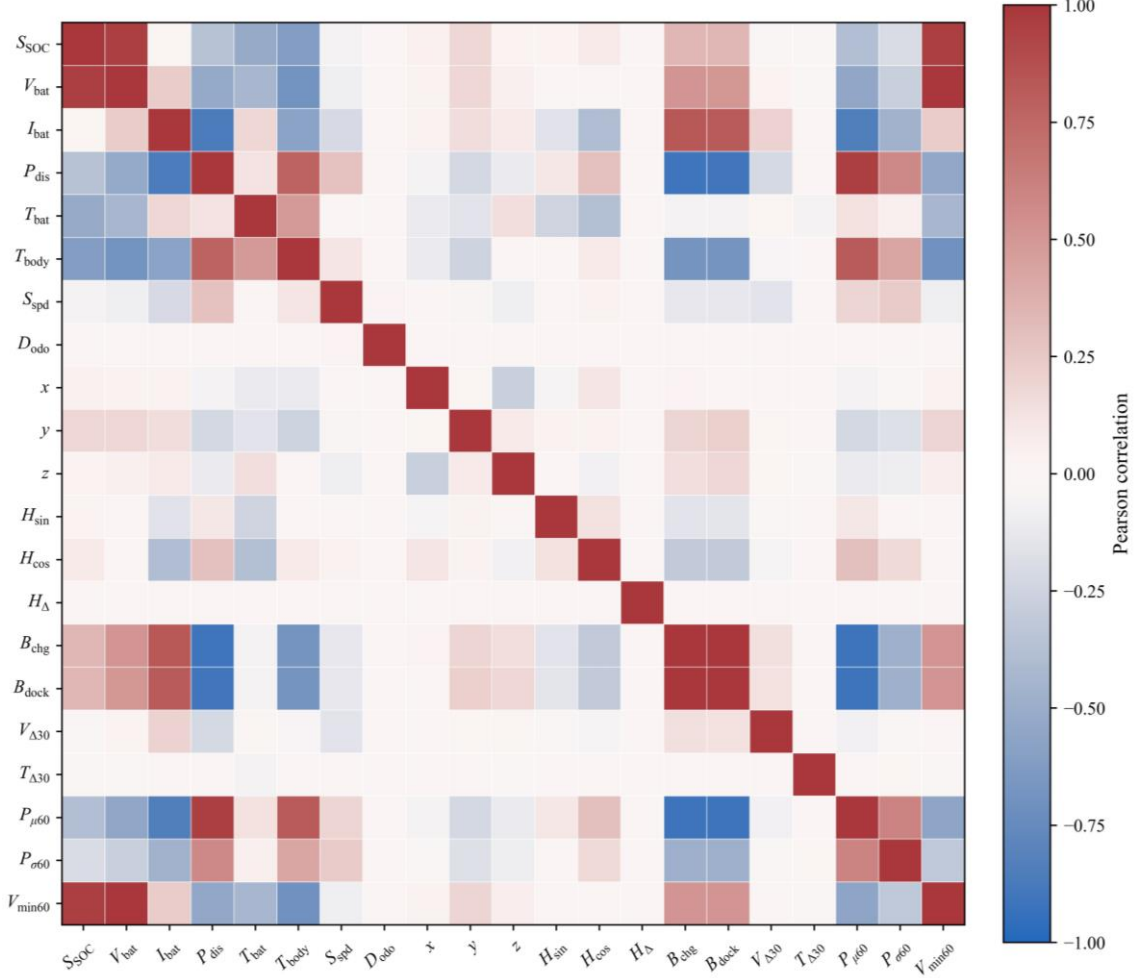


Fig. 4. Pearson correlation matrix of the input features.

3.2. Task available energy and predictive relative risk margin

Since conventional SOC and fixed-capacity estimates cannot fully characterize changes in effective energy availability caused by voltage, current, temperature, and task-dependent loading, this study proposes TAE to quantify the operating-condition-dependent energy reserve. For a prediction horizon H , the discharged energy and SOC reduction are presented by Eqs. (3) and (4).

$$E_H(t) = \frac{1}{3600} \int_t^{t+H} P_{\text{dis}}(\tau) d\tau \quad (3)$$

$$\Delta s_H(t) = s(t) - s(t + H) \quad (4)$$

where $E_H(t)$ is the discharged energy in Wh, $P_{\text{dis}}(\tau)$ is the discharge power in W, $s(t)$ is the SOC fraction at time t , and $\Delta s_H(t)$ is the SOC reduction over H . The energy associated with one unit of SOC and the corresponding TAE are denoted by Eqs. (5) and (6).

$$\beta_H(t) = \frac{E_H(t)}{\Delta s_H(t)} \quad (5)$$

$$TAE(t) = s(t)\beta_H(t) \quad (6)$$

where $\beta_H(t)$ is the operating-condition-dependent energy per unit SOC, and $TAE(t)$ is the estimated TAE.

Platform-specific certified protection thresholds are unavailable for the present dataset. Therefore, the operational decision analysis is formulated using data-driven reference boundaries determined exclusively from the calibration set. The warning and extreme boundaries are defined by Eq. (7).

$$\begin{aligned} E_{\text{warn}} &= Q_{0.10}(TAE_{\text{cal}}), & E_{\text{ext}} &= Q_{0.025}(TAE_{\text{cal}}), \\ V_{\text{warn}} &= Q_{0.10}(V_{\text{min},H,\text{cal}}), & V_{\text{ext}} &= Q_{0.025}(V_{\text{min},H,\text{cal}}), \\ T_{\text{warn}} &= Q_{0.90}(T_{\text{max},H,\text{cal}}), & T_{\text{ext}} &= Q_{0.975}(T_{\text{max},H,\text{cal}}) \end{aligned} \quad (7)$$

where $Q_p(\cdot)$ denotes the p -th quantile, and the subscript cal denotes the measured values obtained from the calibration set. The lower-tail quantiles of TAE and minimum voltage are used to characterize energy- and voltage-related risk, whereas the upper-tail quantiles of maximum battery temperature are used to characterize thermal risk. Only calibration samples satisfying the TAE reliability-mask condition are included when determining E_{warn} and E_{ext} . The corresponding reference scales are expressed by Eq. (8).

$$E_{\text{ref}} = E_{\text{warn}} - E_{\text{ext}}, \quad V_{\text{ref}} = V_{\text{warn}} - V_{\text{ext}}, \quad T_{\text{ref}} = T_{\text{ext}} - T_{\text{warn}} \quad (8)$$

where E_{ref} , V_{ref} , and T_{ref} are the positive intervals between the warning and extreme boundaries of TAE, minimum voltage, and maximum battery temperature, respectively. All reference boundaries and scales are fixed before test-set evaluation. The normalized energy, voltage, and temperature margin components are expressed by Eq. (9).

$$\begin{cases} M_E(t) = \frac{TAE(t) - E_{\text{warn}}}{E_{\text{ref}}}, \\ M_{V,H}(t) = \frac{T_{\text{min},H}(t) - V_{\text{warn}}}{V_{\text{ref}}}, \\ M_{T,H}(t) = \frac{T_{\text{warn}} - T_{\text{max},H}(t)}{T_{\text{ref}}}. \end{cases} \quad (9)$$

where $M_E(t)$, $M_{V,H}(t)$, and $M_{T,H}(t)$ denote the normalized energy, voltage, and temperature margins, respectively; $TAE(t)$, $\hat{V}_{\text{min},H}(t)$, and $\hat{T}_{\text{max},H}(t)$ are the corresponding PCMT predictions; and E_{warn} , V_{warn} , and T_{warn} are the data-driven warning boundaries. The unified relative risk margin is expressed by Eq. (10).

$$M_H(t) = \min\{M_E(t), M_{V,H}(t), M_{T,H}(t)\} \quad (10)$$

A positive M_H indicates that all predicted quantities remain outside the data-defined risk tails. A value between -1 and 0 indicates that at least one predicted quantity lies between its warning and extreme boundaries, while $M_H \leq -1$ indicates an extreme operating condition relative to the calibration distribution. The resulting margin characterizes relative operational risk within the observed data distribution rather than a certified hardware safety limit. The dominant risk factor is identified by Eq. (11).

$$c_H^*(t) = \arg \min\{M_E(t), M_{V,H}(t), M_{T,H}(t)\} \quad (11)$$

where c_H^* identifies energy reserve, minimum voltage, or maximum battery temperature as the dominant risk factor. For quantitative evaluation, the reference margin M_H^{ref} is calculated using the measured future labels, whereas the predicted margin M_H is calculated using the corresponding PCMT outputs. The data-driven decision-support rules are presented in **Table 3**.

Table 3. Data-driven relative risk levels and recommended operational actions.

Relative risk margin	Operating condition	Recommended action
$M_H > 0$	Normal operating range	Continue the current task
$-1 < M_H \leq 0$	Elevated operational risk	Apply energy-saving operation
$M_H \leq -1$	Extreme operating condition	Initiate return and activate protective measures

3.3. Physics-constrained multi-task prediction model

A physics-constrained multi-task model (PCMT) is developed to capture temporal dependencies in historical operating sequences and enforce physical consistency among energy, voltage, and temperature predictions. The overall architecture of PCMT is illustrated in **Fig. 5**. The model receives the categorical historical sequence $\mathbf{X}_i^{\text{cat}}$, continuous historical sequence $\mathbf{X}_i^{\text{con}}$, current-state vector $\mathbf{X}_i^{\text{cur}}$, and unnormalized physical-anchor vector $\mathbf{a}_i = [s_i, V_i, T_i]$ as inputs. Here, s_i , V_i , and T_i denote the current SOC fraction, battery voltage, and battery temperature, respectively. The historical sequences are encoded using a stacked long short-term memory (LSTM) network and a temporal attention mechanism, while the current-state vector is directly fused with the temporal representation through a skip connection. The resulting shared features are then supplied to multiple physics-constrained output heads for the joint prediction of TAE, future energy consumption, minimum voltage, and maximum battery temperature. The categorical variables are

transformed through trainable embedding layers and concatenated with the continuous variables at each historical timestamp. The input representation is presented by Eq. (12).

$$\mathbf{z}_{i,t} = \left[\text{Emb}_1(c_{i,t}^1), \dots, \text{Emb}_K(c_{i,t}^K), \mathbf{x}_{i,t}^{con} \right] \quad (12)$$

where $\mathbf{z}_{i,t}$ is the fused input at time t , $c_{i,t}^k$ is the k -th categorical variable, $\text{Emb}_k(\cdot)$ is its embedding function, and $\mathbf{x}_{i,t}^{con}$ contains the continuous historical features. The temporal sequence is encoded using a stacked LSTM, as presented by Eq. (13).

$$\mathbf{h}_{i,t} = \text{LSTM}(\mathbf{z}_{i,t}, \mathbf{h}_{i,t-1}) \quad (13)$$

where $\mathbf{h}_{i,t}$ is the hidden representation at time t . An attention mechanism is applied to assign different importance to historical timestamps. The attention feature is denoted by Eq. (14).

$$\begin{aligned} e_{i,t} &= \mathbf{w}_a^T \tanh(\mathbf{W}_a \mathbf{h}_{i,t} + \mathbf{b}_a), \\ \alpha_{i,t} &= \frac{\exp(e_{i,t})}{\sum_{\tau=1}^L \exp(e_{i,\tau})}, \\ \mathbf{h}_i^{att} &= \sum_{t=1}^L \alpha_{i,t} \mathbf{h}_{i,t} \end{aligned} \quad (14)$$

where $e_{i,t}$ is the attention score, $\alpha_{i,t}$ is the normalized attention weight, L is the historical sequence length, and $\mathbf{h}_i^{att}$ is the weighted temporal representation.

To preserve the latest operating information, the current-state vector bypasses the recurrent encoder and is directly fused with the attention feature. This current-state skip connection is presented by Eq. (15).

$$\mathbf{q}_i = \left[\mathbf{h}_i^{att}, \mathbf{x}_i^{cur} \right], \quad \mathbf{g}_i = \phi(\mathbf{q}_i) \quad (15)$$

where $\mathbf{q}_i$ is the combined present context, $\phi(\cdot)$ is the shared fully connected network, and $\mathbf{g}_i$ is the shared feature supplied to the physical prediction heads. An auxiliary branch also predicts future motion and operating summaries from $\mathbf{q}_i$. These predictions provide additional supervision during training and are excluded from the physical prediction heads. Four physical heads are established to predict unit-SOC energy, future energy consumption, voltage drop, and temperature rise. Their outputs are constrained through the Softplus function, as presented by Eq. (16).

$$\begin{aligned}
\beta_i &= \text{Softplus}(f_\beta(\mathbf{g}_i))_{,i} \\
E_{H,i} &= \text{Softplus}(f_E(\mathbf{g}_i)), \\
\Delta V_{H,i} &= \text{Softplus}(f_V(\mathbf{g}_i)), \\
\Delta T_{H,i} &= \text{Softplus}(f_T(\mathbf{g}_i))
\end{aligned} \tag{16}$$

where β_i is the predicted energy per unit SOC, $E_{H,i}$ is the predicted future energy consumption, $\Delta V_{H,i}$ is the predicted voltage drop, and $\Delta T_{H,i}$ is the predicted temperature rise. The final physical outputs are reconstructed using the current physical anchors, as denoted by Eq. (17).

$$\begin{aligned}
TAE_i &= s_i \beta_i, \\
V_{\min,H,i} &= V_i - \Delta V_{H,i}, \\
T_{\max,H,i} &= T_i + \Delta T_{H,i}
\end{aligned} \tag{17}$$

where TAE_i is the predicted TAE, $V_{\min,H,i}$ is the predicted minimum voltage, and $T_{\max,H,i}$ is the predicted maximum battery temperature. The physical reconstruction layer converts unconstrained latent representations into physically admissible quantities before the final prediction targets are obtained. The Softplus transformation enforces nonnegative unit SOC energy, future energy consumption, voltage drop, and temperature rise. TAE is reconstructed as the product of the current SOC fraction and the predicted energy per unit SOC. The future minimum voltage is reconstructed by subtracting the predicted voltage drop from the current terminal voltage, whereas the future maximum battery temperature is reconstructed by adding the predicted temperature rise to the current battery temperature. This formulation is consistent with the extrema definitions because the current voltage and battery temperature are included as the initial states of the evaluation interval. The nonnegative heads therefore represent additional voltage depression and thermal accumulation relative to the current state, rather than imposing monotonic voltage or temperature trajectories over the prediction horizon. During voltage recovery, if all subsequent voltage values exceed the current voltage, the minimum-voltage target equals the current voltage and the corresponding voltage drop becomes zero. Similarly, during cooling, if all subsequent battery temperatures remain below the current temperature, the maximum-temperature target equals the current temperature and the corresponding temperature rise becomes zero.

Small observed SOC reductions increase the sensitivity of the TAE label to SOC quantization; therefore, a reliability mask is applied. The TAE mask is presented by Eq. (18).

$$m_i = \begin{cases} 1, & \Delta S_{H,i} > \Delta S_{\min}, \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

where m_i is the TAE reliability mask, $\Delta S_{H,i}$ is the observed SOC reduction, and $\Delta S_{\min}$ is the minimum reliable SOC reduction. Samples with $m_i = 0$ remain available for energy, voltage, temperature, and auxiliary-task learning, while their TAE and unit-SOC-energy losses are excluded. In the present implementation, $\Delta S_{\min} = 0.015$. Given the 1 percentage point SOC resolution and the strict criterion $\Delta S_H > \Delta S_{\min}$, this corresponds to an effective minimum accepted observed SOC decrease of 2 percentage points over the prediction horizon of 300 s. This threshold is introduced to reduce the amplification of SOC quantization and measurement uncertainty in the energy-to-SOC-change ratio. The mask is used only during training and evaluation and is not supplied as a model input. The supervised physical loss is presented by Eq. (19).

$$\begin{aligned} \mathcal{L}_{TAE} &= \frac{\sum_{i=1}^N m_i \text{Huber}(TAE_i - TAE_i^*)}{\max(\sum_{i=1}^N m_i, 1)} \\ \mathcal{L}_E &= \frac{1}{N} \sum_{i=1}^N (E_{H,i} - E_{H,i}^*)^2 \\ \mathcal{L}_V &= \frac{1}{N} \sum_{i=1}^N (V_{\min,H,i} - V_{\min,H,i}^*)^2 \\ \mathcal{L}_T &= \frac{1}{N} \sum_{i=1}^N (T_{\max,H,i} - T_{\max,H,i}^*)^2 \\ \mathcal{L}_{\text{phy}} &= \lambda_{TAE} \mathcal{L}_{TAE} + \lambda_E \mathcal{L}_E + \lambda_V \mathcal{L}_V + \lambda_T \mathcal{L}_T \end{aligned} \quad (19)$$

where $*$ denotes the measured reference value, the hat denotes the corresponding model prediction, and the tilde denotes standardization using statistics calculated exclusively from the training set. $\mathcal{L}_{TAE}$ is evaluated only for reliable samples satisfying $m_i = 1$, whereas $\mathcal{L}_E$, $\mathcal{L}_V$, and $\mathcal{L}_T$ are evaluated over all samples in the mini-batch. N denotes the batch size. The four task weights $\mathcal{L}_{TAE}$, $\mathcal{L}_E$, $\mathcal{L}_V$, and $\mathcal{L}_T$ control the relative contributions of the physical prediction objectives. λ terms are task weights. All target losses are calculated in standardized spaces to reduce scale imbalance. An auxiliary loss is imposed on the predicted unit-SOC energy, as presented by Eq. (20).

$$\begin{aligned} \mathcal{L}_\beta &= \frac{\sum_{i=1}^N m_i \text{Huber}(\tilde{\beta}_i - \tilde{\beta}_i^*)}{\max(\sum_{i=1}^N m_i, 1)} \\ \beta_i^* &= \frac{E_{H,i}^*}{\Delta S_{H,i}} \end{aligned} \quad (20)$$

where $\tilde{\beta}_i$ and $\tilde{\beta}_i^*$ denote the predicted and reference unit-SOC energy standardized using the training-set statistics, respectively. β_i^* is the reference unit-SOC energy calculated from the measured future energy $E_{H,i}^*$ and observed SOC reduction $\Delta s_{H,i}$; m_i is the TAE reliability mask defined in Eq. (18); N is the batch size. The denominator is lower-bounded by 1 to avoid numerical instability when no reliable TAE sample is available in a mini-batch.

The future energy predicted by the independent energy head is constrained to remain consistent with the energy reconstructed from unit-SOC energy and SOC reduction. The consistency loss is denoted by Eq. (21).

$$\begin{aligned} E_{H,i}^{\text{head}} &= \frac{E_{H,i} - \mu_E}{\sigma_E}, \quad E_{H,i}^\beta = \frac{\beta_i \Delta s_{H,i} - \mu_E}{\sigma_E}, \\ \mathcal{L}_{\text{con}} &= \frac{\sum_{i=1}^N m_i r_i \left(E_{H,i}^{\text{head}} - E_{H,i}^\beta \right)^2}{\max \left(\sum_{i=1}^N m_i r_i, 1 \right)} \end{aligned} \quad (21)$$

where $E_{H,i}$ is the future energy predicted directly by the independent energy head, and $\beta_i \Delta s_{H,i}$ is the corresponding future energy reconstructed from the predicted unit-SOC energy β_i and observed SOC reduction $\Delta s_{H,i}$. $E_{H,i}^{\text{head}}$ and $E_{H,i}^\beta$ denote these two energy estimates standardized using the mean μ_E and standard deviation σ_E of the future-energy target calculated exclusively from the training set. m_i is the TAE reliability mask defined in Eq. (18), r_i is the energy-consistency validity indicator, and N is the batch size. The denominator is lower-bounded by 1 to avoid numerical instability when no valid consistency sample is available in a mini-batch. The observed SOC reduction is used only to construct the consistency constraint during training and is excluded from the inference inputs. The auxiliary future-summary loss is presented by Eq. (22).

$$\mathcal{L}_{\text{plan}} = \mathcal{L}_{\text{H}} \left(\mathbf{P}_i^{\text{con}}, \mathbf{P}_i^{\text{con}} \right) + \mathcal{L}_{\text{BCE}} \left(p_i^{\text{dock}}, p_i^{\text{dock}} \right) \quad (22)$$

where $\mathbf{P}_i^{\text{com}}$ contains the continuous future motion summaries, p_i^{dock} is the future docking-event label, $\mathcal{L}_{\text{H}}$ is the Huber loss, and $\mathcal{L}_{\text{BCE}}$ is the weighted binary cross-entropy loss. The total multi-task loss is presented by Eq. (23).

$$\mathcal{L}_{\text{total}} = \lambda_{\text{TAE}} \mathcal{L}_{\text{TAE}} + \lambda_E \mathcal{L}_E + \lambda_V \mathcal{L}_V + \lambda_T \mathcal{L}_T + \lambda_\beta \mathcal{L}_\beta + \lambda_{\text{con}} \mathcal{L}_{\text{con}} + \lambda_{\text{plan}} \mathcal{L}_{\text{plan}} \quad (23)$$

where λ_{TAE} , λ_E , λ_V , and λ_T control the contributions of TAE, future-energy, minimum-voltage, and maximum-temperature supervision, respectively. λ_β , λ_{con} , and λ_{plan} control the contributions of unit-SOC-energy supervision, energy-consistency regularization, and auxiliary future-summary learning, respectively. This learning structure separates shared temporal representation from target-specific physical reconstruction. The shared encoder captures temporal dependencies common to energy, voltage, and thermal responses, while the individual physical heads preserve the distinct output relationships defined by the current SOC, voltage, and temperature anchors. The auxiliary learning objective further regularizes the shared representation by incorporating complementary information from future operating summaries. Consequently, online prediction relies exclusively on historical operating sequences and current-state measurements, while the physical constraints and auxiliary objectives regulate how these observations are mapped to the four prediction targets. The principal data-processing, network-architecture, loss-function, and training hyperparameters of PCMT are summarized in **Table 4**.

The hyperparameter configuration summarized in **Table 4** is chosen to balance model complexity, regularization, and stable multi-task convergence. The hidden dimensions and number of recurrent layers are selected to provide sufficient temporal representation capacity while maintaining a compact architecture. A dropout rate of 0.4 and weight decay of 1×10^{-3} are used to reduce overfitting. The initial learning rate is set to 1×10^{-4} , with an adaptive learning rate scheduler that reduces the learning rate by a factor of 0.5 after four epochs without validation improvement. Model selection, learning-rate scheduling, and early stopping were governed by the validation criterion $J_{\text{sel}} = 0.7 \text{RMSE}_{\text{TAE, val}} / \sigma_{\text{TAE, train}} + 0.3 \text{RMSE}_{\text{E, val}} / \sigma_{\text{E, train}}$, where the training-set standard deviations remove the difference in physical scale between the two energy-related targets. The larger weight assigned to TAE reflects its role as the primary task-conditioned energy indicator. The physical task weights were selected to balance the contributions of the four standardized prediction targets while prioritizing TAE and future energy. The consistency weight of 0.1 is used as a regularization term so that physical consistency influences learning without dominating direct supervision. The auxiliary future-summary loss is assigned a smaller weight of 0.05 because it primarily serves as representation regularization.

Table 4. Key hyperparameters of the proposed PCMT.

Parameter	Description	Value
Prediction horizon	Duration of the future prediction interval	300 s
Historical window length	Duration of the historical operating sequence	90 s
Sliding stride	Time interval between adjacent samples	10 s
Categorical embedding dimension	Embedding dimension assigned to each categorical feature	4
LSTM hidden dimension	Hidden-state dimension of the LSTM encoder	64
Number of LSTM layers	Number of stacked LSTM layers	2
Attention hidden dimension	Hidden dimension of the temporal attention module	64
Shared hidden dimension	Hidden dimension of the shared feature layers	64
Physical-head hidden dimension	Hidden dimension of each physical prediction head	32
Dropout rate	Dropout rate applied to the LSTM and fully connected layers	0.4
Batch size	Mini-batch size used for model training	64
Initial learning rate	Initial learning rate of the optimizer	1×10^{-4}
Weight decay	Weight-decay coefficient of AdamW	1×10^{-3}
Maximum number of epochs	Maximum number of training epochs	120
Early-stopping patience	Number of epochs without validation improvement before termination	12
Huber transition parameter	Transition parameter of the Huber loss	0.5
Physical-task loss weights	Loss weights for TAE, future energy, minimum voltage, and maximum temperature	[2.0,1.5,0.1,0.1]
Auxiliary-task loss weight	Weight assigned to future-task-summary supervision	0.05
TAE reliability threshold Δs_{min}	Minimum SOC reduction required for reliable TAE supervision	0.015
Unit SOC energy weight λ_{β}	Loss weight assigned to unit SOC energy supervision	1.0
Energy consistency weight λ_{con}	Loss weight assigned to energy consistency regularization	0.1
LR scheduler factor	Learning-rate reduction factor applied by the scheduler	0.5
LR scheduler patience	Number of epochs without improvement before learning-rate reduction	4
Minimum learning rate	Lower bound of the learning rate during scheduling	1×10^{-6}

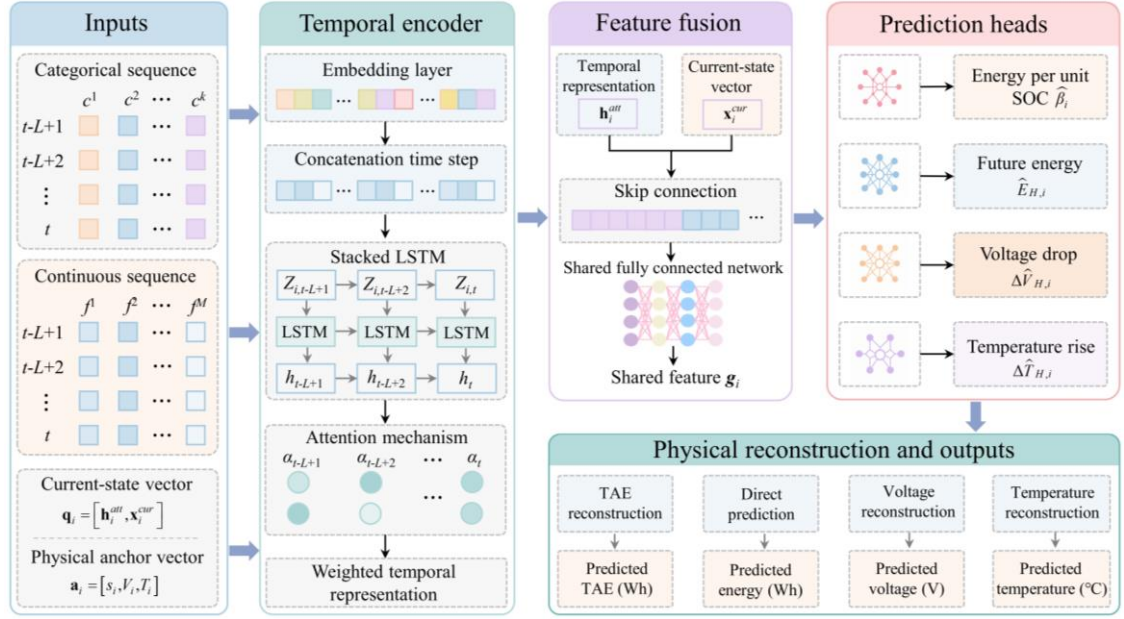


Fig. 5. Overall architecture of the proposed PCMT. Historical sequences and current-state features are encoded and fused to predict energy, voltage, and temperature quantities, which are converted into the final outputs through physical reconstruction.

3.4. Evaluation metrics

The predictive performance of PCMT is assessed using RMSE, NRMSE, R^2 , and MAPE. These metrics are calculated separately for TAE, future energy consumption, minimum voltage, and maximum battery temperature after the predicted and measured values are transformed back to their original physical units. RMSE quantifies the magnitude of the prediction error in the corresponding physical unit, while NRMSE provides a scale-independent measure for comparison across prediction targets. They are expressed by Eq. (24).

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \\ \text{NRMSE} &= \frac{\text{RMSE}}{\bar{y}} \times 100\% \end{aligned} \quad (24)$$

R^2 quantifies the proportion of variance in the measured values explained by the model predictions and is expressed by Eq. (25).

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (25)$$

MAPE quantifies the average relative deviation between the predicted and measured values and is expressed by Eq. (26).

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{\max(y_i, \varepsilon)} \quad (26)$$

where N is the number of evaluated samples for the corresponding prediction target. y_i and $\hat{y}_i$ denote the measured and predicted values of the i -th sample, respectively, while $\bar{y}$ is the mean measured value over the same evaluation subset. ε is a small positive constant introduced to avoid numerical instability when the measured value approaches zero. For TAE, N includes only the samples satisfying the reliability criterion defined in Eq. (18), whereas the other three prediction targets are evaluated using all valid samples. RMSE retains the original physical unit of each prediction target, including Wh for TAE and future energy consumption, V for minimum voltage, and °C for maximum battery temperature. The denominator used for NRMSE is the mean measured value $\bar{y}$ of the corresponding target, and NRMSE is therefore expressed as a percentage. MAPE is also expressed as a percentage. Lower RMSE, NRMSE, and MAPE values and a higher R^2 indicate better predictive performance. For the test set, the corresponding normalization denominators are 738.191 Wh for TAE, 25.116 Wh for future energy consumption, 75.602 V for minimum voltage, and 50.628 °C for maximum battery temperature.

4. Results and discussion

4.1. The performance of the proposed framework

The predictive performance of PCMT is first evaluated for the four target variables, as summarized in **Fig. 6** and **Table 5**. The predicted values are closely distributed around the identity line for TAE, future energy consumption, minimum voltage, and maximum battery temperature. The proposed framework achieves NRMSE values of 1.34%, 1.60%, 0.68%, and 0.59% for the four targets, respectively. The corresponding R^2 values range from 0.972 to 0.997, indicating that most of the variation in the observed physical states is retained by the predictions. The MAPE values remain below 1.7% for all tasks. Particularly close agreement is obtained for maximum battery temperature, with an R^2 of 0.997 and a MAPE of 0.50%. For minimum voltage, the boxplots in **Fig. 6(c)** show that the median predictions remain near the identity line across the observed voltage intervals, although the dispersion increases slightly at higher voltages. The TAE prediction provides the central result of this evaluation. TAE extends a conventional SOC description

by combining the current charge state with the task-dependent energy available per unit SOC. It therefore represents the effective energy supply under the current operating condition rather than treating identical SOC values as energetically equivalent across different tasks. Despite the additional dependence on task conditions and SOC variation, PCMT predicts TAE with an NRMSE of 1.34%, an R^2 of 0.995, and a MAPE of 1.12%. The close correspondence in **Fig. 6(a)** is maintained over an observed range of approximately 200 to 1,100 Wh. These results indicate that the proposed representation remains quantitatively stable across substantially different available-energy levels and can serve as an operational energy-state variable for subsequent safety-margin assessment.

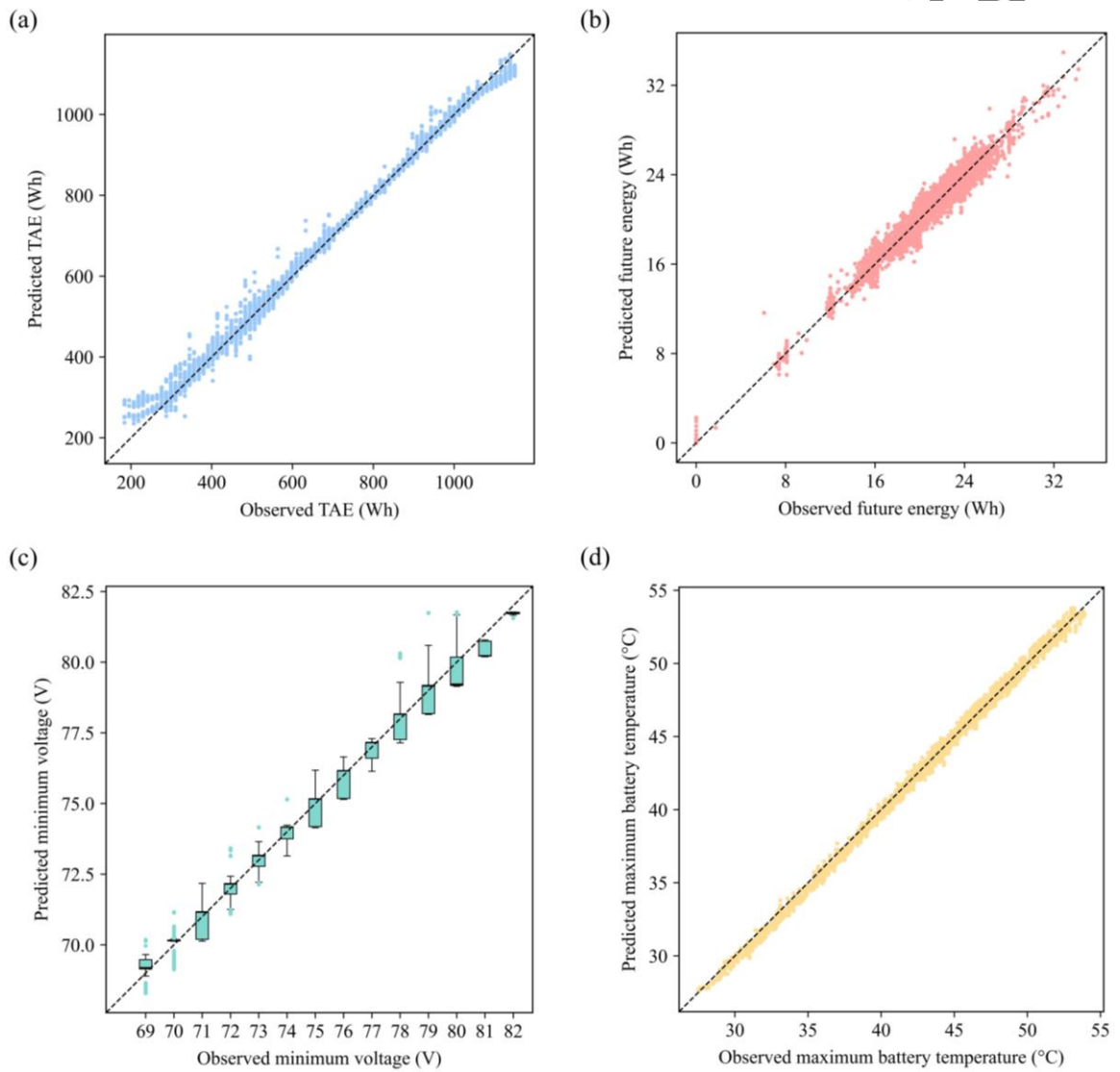


Fig. 6. Prediction results of the proposed PCMT framework for four targets. (a) Observed and predicted TAE. (b) Observed and predicted future energy consumption. (c) Observed and predicted minimum voltage, with boxplots summarizing the prediction distributions within the corresponding observed voltage intervals.

(d) Observed and predicted maximum battery temperature. The diagonal dashed line indicates ideal agreement between the observed and predicted values.

Table 5. Predictive performance of PCMT across four prediction tasks.

Prediction target	RMSE	NRMSE (%)	R ²	MAPE (%)
TAE	9.890 Wh	1.34	0.995	1.12
Future energy	0.402 Wh	1.60	0.984	1.66
Minimum voltage	0.514 V	0.68	0.972	0.35
Maximum temperature	0.299 °C	0.59	0.997	0.50

The normalized comparison in **Fig. 7** further illustrates the differences between PCMT and the baseline models. Predictions from PCMT remain concentrated near the identity line across the four target scales. Greater dispersion is observed for the vanilla LSTM, Transformer, and gated recurrent unit (GRU), particularly for TAE and future energy consumption. XGBoost provides accurate voltage and temperature predictions, although its future-energy estimates exhibit a broader deviation from the reference line. Ridge regression shows the largest systematic errors, including pronounced TAE outliers and underestimation at high future-energy values. The quantitative results in **Table 6** confirm these observations. PCMT obtains the lowest TAE NRMSE among all evaluated methods. Relative to XGBoost, which provides the strongest baseline result for TAE, the error is reduced from 4.85% to 1.34%, corresponding to a reduction of approximately 72%. A similar advantage is observed for future energy consumption, for which PCMT reduces the NRMSE from the best baseline value of 4.65% to 1.60%. XGBoost achieves slightly lower errors for minimum voltage and maximum battery temperature, whereas PCMT retains competitive accuracy for both targets while providing substantially stronger performance for the energy-related outputs. This balance is important for the proposed application because TAE, future energy demand, voltage depression, and thermal accumulation are evaluated within a unified physical framework. The results therefore demonstrate that PCMT provides a consistent multi-task representation, with its main advantage concentrated in the task-dependent characterization of available energy.

To further examine the behavior of the voltage and temperature extrema predictions under recovery and cooling conditions, a separate analysis is conducted on the independent test set. Fifteen voltage-recovery intervals are identified, accounting for 0.16% of the test samples, while 26 cooling intervals account for 0.27%. In the voltage-recovery intervals, all subsequent voltage values over the future-only interval remain above the current voltage. In the cooling intervals, all subsequent battery-temperature values remain below

the current temperature. Under the current-inclusive extrema definition in Eq. (2), the current voltage and battery temperature therefore become the corresponding extrema for these intervals, resulting in a zero reference voltage drop and a zero reference temperature rise, respectively. Although these intervals represent only a small proportion of the test set, their presence confirms that voltage recovery and battery cooling are retained in the evaluation data. More importantly, these cases show that the nonnegative voltage-drop and temperature-rise heads remain physically consistent with recovery and cooling behavior. The formulation constrains the predicted extrema relative to the current physical state rather than assuming a monotonic voltage decrease or temperature increase over the prediction horizon.

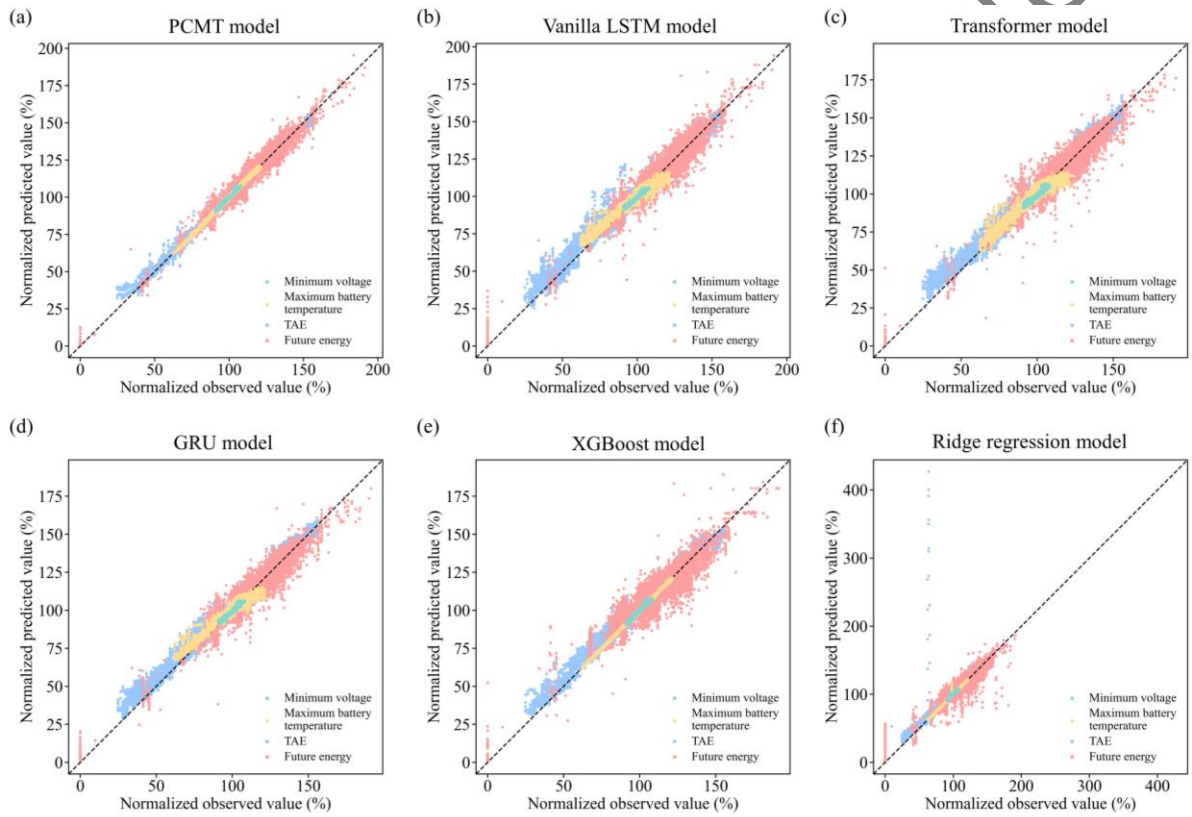


Fig. 7. Normalized prediction comparison between PCMT and the baseline models. (a) Prediction results of PCMT. (b) Prediction results of the vanilla LSTM. (c) Prediction results of the Transformer. (d) Prediction results of the GRU. (e) Prediction results of XGBoost. (f) Prediction results of ridge regression. The four prediction targets include TAE, future energy consumption, minimum voltage, and maximum battery temperature. The diagonal dashed line indicates ideal agreement between the normalized observed and predicted values.

Table 6. Performance comparison of PCMT and baseline models across four prediction tasks.

Methods	TAE NRMSE (%)	Energy NRMSE (%)	Voltage NRMSE (%)	Temperature NRMSE (%)
Proposed PCMT	1.34	1.60	0.68	0.59
Vanilla LSTM	5.43	4.68	1.07	3.43
Transformer	5.79	5.08	1.00	3.94
GRU	5.19	4.65	1.05	3.49
XGBoost	4.85	6.16	0.53	0.44
Ridge regression	10.79	7.83	0.59	0.69

To further examine the effect of future operating changes on predictions over the 300 s horizon, the test samples are grouped according to the timing of the first gait or task-phase transition after the current prediction instant. As summarized in **Table 7**, samples without a gait or task-phase change within the prediction horizon account for 6.10% of the test set. Early-transition samples, in which the first change occurs within 30 s, account for 41.33%, while intermediate-transition and late-transition samples account for 42.36% and 10.21%, respectively. This grouping provides a direct evaluation of prediction performance when the future operating condition changes shortly after the current state is observed. This analysis is particularly relevant because PCMT does not receive an explicit future mission plan during inference.

The influence of transition timing is most evident for future-energy and minimum-voltage prediction. For samples without a transition, the future-energy and minimum-voltage NRMSE values are 0.99% and 0.18%, respectively. When the first gait or task-phase change occurs within 30 s, these errors increase to 1.77% and 0.80%, indicating that an early change in operating demand introduces additional uncertainty before the new task state is reflected in the model input. As the transition occurs later in the prediction horizon, the corresponding errors decrease to 1.04% and 0.26% for the late-transition group. In contrast, the maximum-temperature NRMSE remains within a narrow range of 0.57% to 0.64% across all groups, reflecting the slower thermal response to abrupt operating changes. The TAE NRMSE also remains close to the overall test-set level across the transition groups, although the relatively small subset without a transition exhibits a higher value of 1.78%. Overall, abrupt gait or task changes mainly affect quantities that respond directly to future operating demand, particularly future energy consumption and minimum voltage. Thermal prediction remains comparatively insensitive to transition timing.

Table 7. Prediction performance with respect to the timing of future gait or task transitions.

Future-transition condition	Definition	Proportion	TAE NRMSE	Energy NRMSE	Voltage NRMSE	Temperature NRMSE
Without transition	No gait or task-phase change within 300 s	6.10%	1.78%	0.99%	0.18%	0.64%
Early transition	First gait or task-phase change occurs within 0–30 s	41.33%	1.34%	1.77%	0.80%	0.59%
Intermediate transition	First gait or task-phase change occurs within 30–120 s	42.36%	1.24%	1.62%	0.68%	0.59%
Late transition	First gait or task-phase change occurs within 120–300 s	10.21%	1.31%	1.04%	0.26%	0.57%
Overall test set	All test samples	100.00%	1.34%	1.60%	0.68%	0.59%

4.2. Ablation experiments

To clarify the contribution of different information sources and auxiliary objectives, ablation experiments are conducted at both the input and task levels. **Fig. 8(a)** compares the RMSE ratios obtained after removing each encoder-input group with those of the full PCMT model. Most ratios remain close to 1.00, demonstrating that the proposed framework maintains stable predictive performance when one source of information is unavailable. Such stability is attributed to the complementary relationships among electrical, thermal, locomotion, and operating-context variables within the shared encoder. Among the four targets, TAE exhibits the clearest dependence on terrain and motion-mode information, which agrees with its task-conditioned definition. Thermal states also contribute to maximum-temperature prediction, while the limited variations observed in other configurations indicate that correlated inputs provide compensatory information.

Further evidence is provided by the auxiliary-task analysis in **Fig. 8(b)**. Relative to the energy-only configuration, voltage supervision reduces the voltage RMSE from 0.560 to 0.516 V, while temperature supervision decreases the temperature RMSE from 0.372 to 0.299 °C. Joint electro-thermal supervision produces the lowest voltage RMSE of 0.514 V and retains the minimum temperature RMSE of 0.299 °C. These improvements correspond to reductions of 8.2% and 19.6%, respectively. The results show that task-specific supervision strengthens the representation of the associated physical process, whereas the full PCMT configuration achieves the most balanced prediction capability across energy, voltage, and thermal targets.

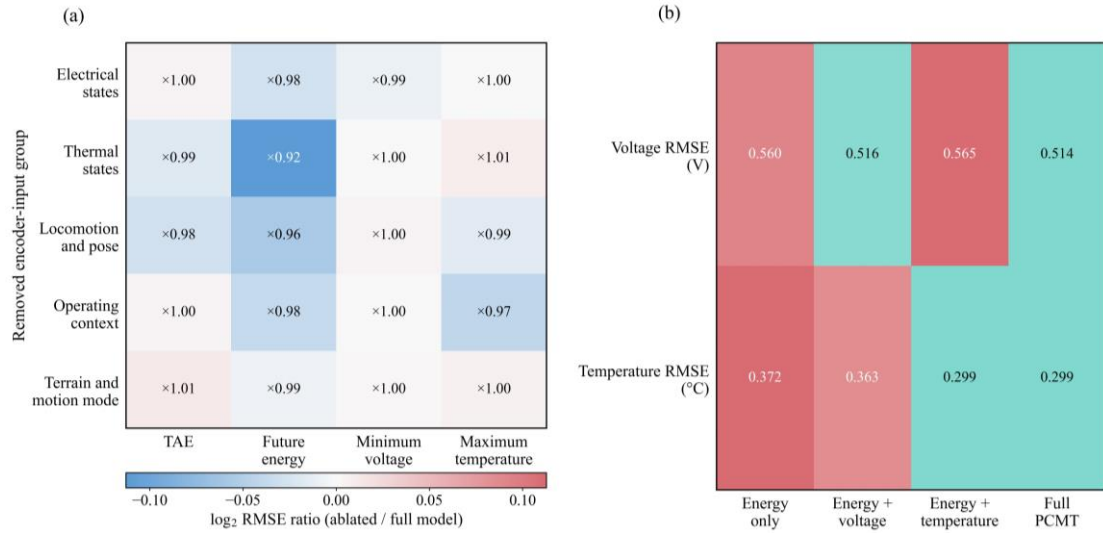


Fig. 8. Ablation of encoder-input groups and electro-thermal auxiliary tasks. (a) RMSE ratios for four prediction tasks after removing each encoder-input group relative to the full model. (b) Voltage and temperature RMSE under four electro-thermal auxiliary-task configurations.

4.3. Interpretability analysis

The interpretability of PCMT is assessed through controlled perturbation experiments and temporal attention analysis. For the perturbation experiments, 300 test windows are subjected to systematic variations in SOC, load, initial battery temperature, thermal history, and electrical history. The response distributions are summarized by the violin and box plots in **Fig. 9**, with black diamonds denoting the mean values and error bars representing the 95% cluster bootstrap confidence intervals. The SOC perturbations in **Fig. 9(a)** produce nearly symmetric changes in predicted TAE. A decrease of 5 percentage points substantially reduces TAE, whereas an equivalent increase produces a response of comparable magnitude. This relationship is consistent with the reconstruction of TAE from current SOC and predicted unit-SOC task energy.

Load sensitivity is further examined through the energy and voltage responses. Scaling the load from 0.8× to 1.2× progressively increases future energy consumption, as shown in **Fig. 9(b)**, and lowers the predicted minimum voltage in **Fig. 9(c)**. The corresponding response directions agree with the expected effects of intensified electrical demand. Thermal sensitivity exhibits a different pattern. Perturbations to the initial battery temperature induce only limited changes in predicted temperature rise in **Fig. 9(d)**, reflecting the role of current temperature as a reconstruction anchor. By comparison, amplified thermal history produces a clear increase in temperature rise in **Fig. 9(e)**, indicating dependence on accumulated thermal loading. A similar monotonic response appears for electrical history in **Fig. 9(f)**, where stronger perturbations

lead to larger voltage-drop estimates and broader response distributions. These results indicate that PCMT resolves the distinct contributions of present-state anchors and historical operating conditions across the four prediction tasks.

The shared temporal attention weights are examined for three representative prediction cases in **Fig. 10**. For the high energy consumption case, substantial attention is assigned to several earlier fluctuations in power and current, as illustrated in **Fig. 10(a)**. This dispersed pattern suggests that the energy estimate is informed by sustained load variations across the historical window. A different temporal dependence is observed for the low minimum voltage case in **Fig. 10(b)**, where the attention weights become concentrated near the end of the sequence. The recent electrical variations therefore exert a stronger influence on the prediction of voltage depression. By comparison, the case with a higher future battery-temperature peak exhibits a broader attention region over the recent history, as shown in **Fig. 10(c)**. Such a distribution is consistent with the cumulative character of heat generation and delayed thermal dissipation. The displacement of the dominant attention region across the three cases indicates that the shared encoder resolves distinct temporal signatures associated with energy demand, voltage depression, and thermal accumulation.

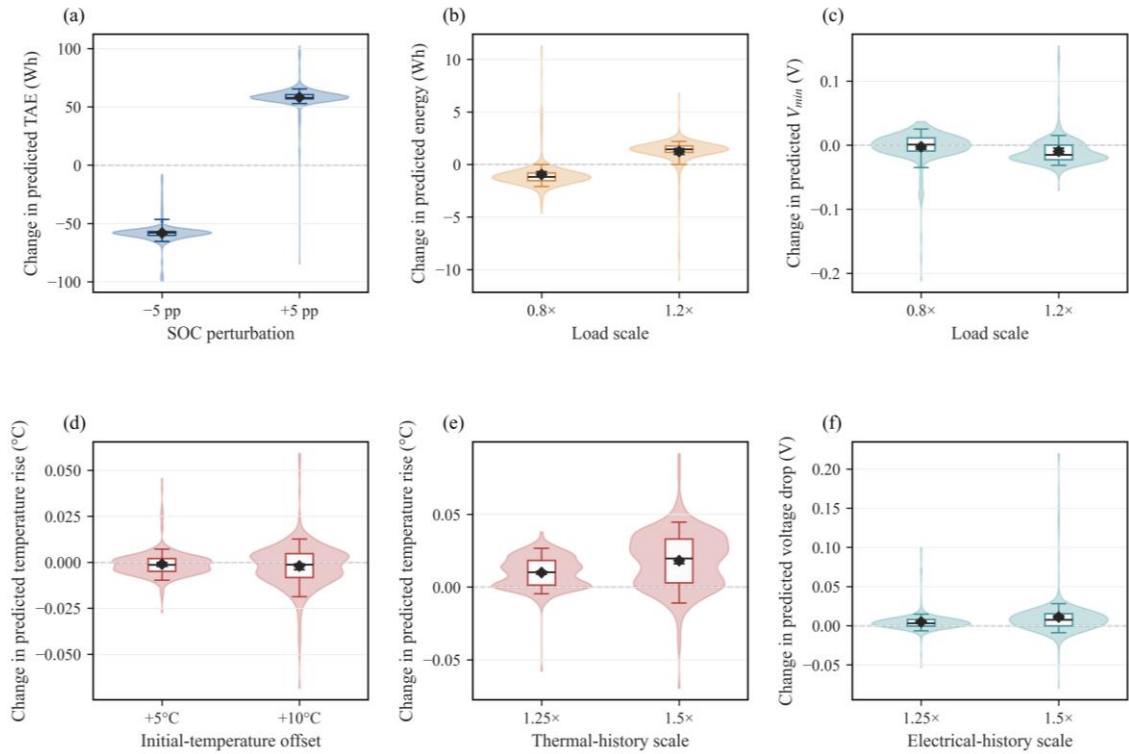


Fig. 9. Perturbation-based interpretability analysis of PCMT across 300 test windows. (a) SOC perturbation on predicted TAE, (b) load scaling on predicted future energy consumption, (c) load scaling on

predicted minimum voltage, (d) initial-temperature offset on predicted temperature rise, (e) thermal-history scaling on predicted temperature rise, and (f) electrical-history scaling on predicted voltage drop. Violin profiles represent response distributions, embedded boxplots show the median and interquartile range, black diamonds indicate mean responses, and error bars denote 95% cluster bootstrap confidence intervals.

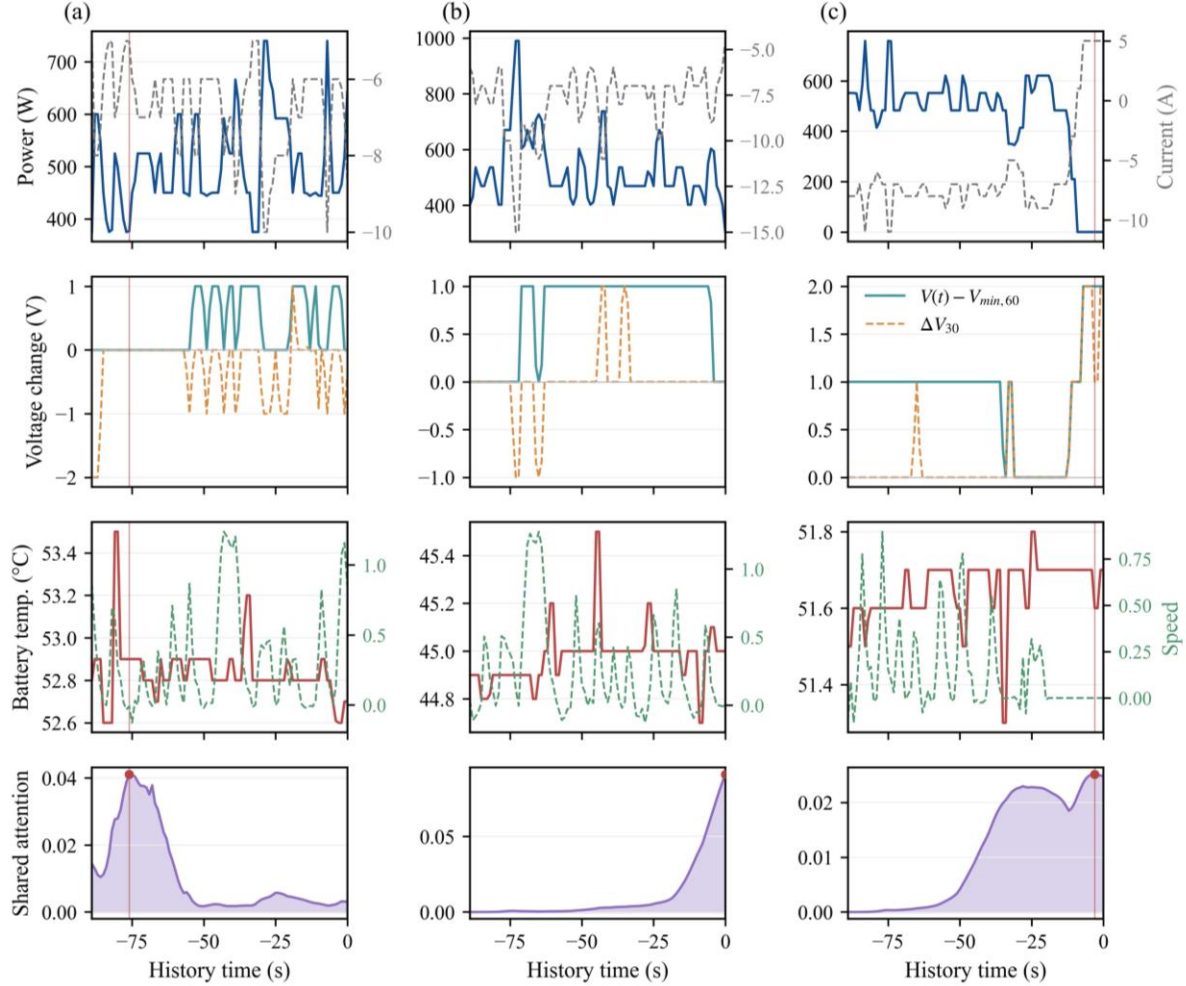


Fig. 10. Shared temporal attention for three representative prediction cases. (a) a high predicted energy consumption case; (b) a low predicted voltage case; and (c) a high predicted temperature case. For each case, the first three rows present historical power and current, voltage-related variables, and battery temperature and speed, respectively, while the fourth row presents the attention weights over the 90 historical time steps. The red vertical line and marker indicate the historical time step assigned the maximum attention weight.

4.4. Predictive risk margin and decision support

Beyond the prediction of individual battery variables, practical quadruped-robot operation requires a mechanism that converts heterogeneous predictions into interpretable operational guidance. Accordingly,

the data-driven relative risk margin introduced in Section 3.2 is applied to the independent test set to integrate TAE, minimum voltage, and maximum battery temperature into a unified assessment. The reference margin M_H^{ref} is calculated from the measured future labels, whereas the predictive margin M_H is derived from the corresponding PCMT outputs. Both margins are evaluated using the calibration-set boundaries fixed before test-set analysis, thereby enabling a direct examination of whether the multi-task predictions preserve the resulting operational risk level. Quantitative results in **Table 8** further confirm the accuracy of the integrated assessment. The predictive relative risk margin achieves an RMSE of 0.187 and an R^2 of 0.993, indicating a close correspondence between the predicted and reference margins. After the continuous margin is converted into the three operational risk levels, an agreement of 94.01% is obtained. The optimistic misclassification rate is limited to 5.61%, showing that only a relatively small proportion of samples are assigned a less severe predicted risk level than their reference level. Because this type of error may delay energy-saving, return, or protective actions, its explicit evaluation provides a more operationally meaningful assessment than overall agreement alone. In addition, the dominant-risk-factor accuracy reaches 93.30%, demonstrating that the proposed framework can identify whether TAE, minimum voltage, or maximum battery temperature governs the final risk assessment.

Table 8. Quantitative performance of PCMT for data-driven relative risk assessment and operational decision support on the test set.

Metric	Value
Relative-margin RMSE	0.187
Relative-margin R^2	0.993
Risk-level agreement	94.01%
Optimistic misclassification rate	5.61%
Dominant-risk-factor accuracy	93.30%

To further examine the distribution and severity of optimistic misclassifications, **Fig. 11** presents the relationship between the reference and predicted margins, the class-level confusion matrix, and the dependence of optimistic errors on proximity to the decision boundaries. As shown in **Fig. 11(a)**, most predicted margins remain close to their corresponding reference values, while the highlighted optimistic errors are mainly located near the boundaries at $M_H = 0$ and $M_H = -1$. **Fig. 11(b)** further shows the classification transitions among the Normal, Elevated, and Extreme risk classes. The class-specific results are further quantified in **Table 9**. Among the 1088 reference Elevated samples, 294 are classified as Normal,

corresponding to an optimistic error rate of 27.02%. Among the 723 reference Extreme samples, 258 are classified as Elevated, corresponding to an optimistic error rate of 35.68%. No Extreme-to-Normal transition is observed. Therefore, all 552 optimistic errors correspond to one-level risk underestimation, with no two-level optimistic misclassification observed in the test set.

Table 9. Optimistic misclassification analysis across reference risk classes.

Reference class	Number of samples	Correct classification	One-level optimistic	Two-level optimistic	Optimistic error rate (%)
Normal	8026	7996	N/A*	N/A*	N/A*
Elevated	1088	787	294	0	27.02
Extreme	723	465	258	0	35.68
Overall	9837	9248	552	0	5.61

* Optimistic misclassification is undefined for the reference Normal class because no lower risk category exists.

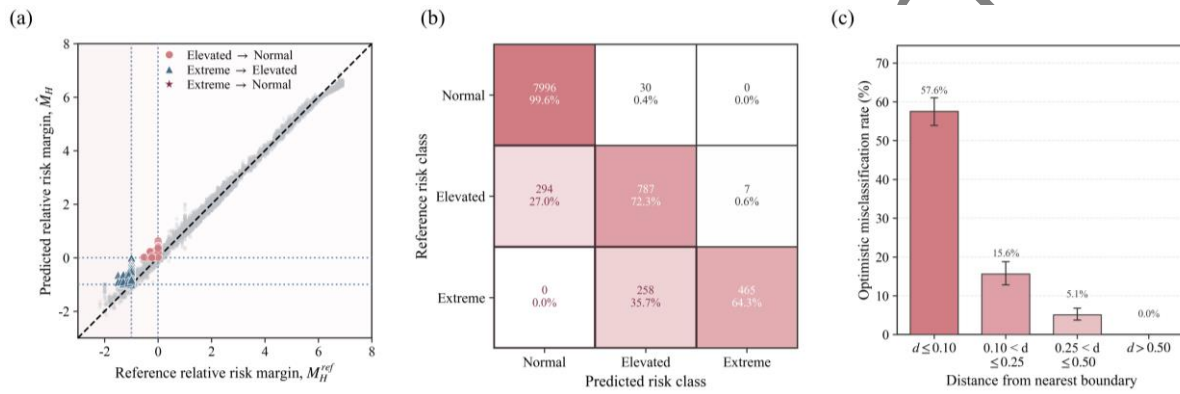


Fig. 11. Optimistic misclassification analysis of the predictive relative risk margin on the test set. (a) Reference and predicted relative risk margins with optimistic errors highlighted by risk class transition. (b) Confusion matrix of the three risk classes. (c) Optimistic misclassification rate versus distance from the nearest decision boundary. Error bars indicate 95% cluster bootstrap confidence intervals.

To evaluate whether the optimistic errors are concentrated near the decision boundaries, the distance of each test sample from the nearest data-defined boundary is expressed by Eq. (27).

$$d_i = \min(|M_{H,i}^{ref}|, |M_{H,i}^{ref} + 1|) \quad (27)$$

where d_i denotes the distance of the i -th test sample from the nearest decision boundary, and the two data-defined decision boundaries are located at $M_H = 0$ and $M_H = -1$. A smaller d_i indicates that the reference margin lies closer to a risk class boundary. The relationship between optimistic misclassification and proximity to the decision boundary is summarized in **Fig. 11(c)** and **Table 10**. The optimistic error rate is 57.55% for samples located within 0.10 normalized margin units of the nearest boundary, and this region contains 76.63% of all optimistic errors. The rate decreases to 15.61% for $0.10 < d \leq 0.25$ and to 5.08% for

$0.25 < d \leq 0.50$. No optimistic misclassification is observed for $d > 0.50$. These results indicate that the optimistic errors are strongly concentrated near the data-defined decision boundaries and decrease progressively as the reference margin moves farther from the nearest boundary.

Table 10. Optimistic misclassification with respect to the distance from the nearest decision boundary.

Distance from nearest boundary	Total samples	Optimistic errors	Optimistic error rate (%)	Share of all optimistic errors (%)
$d \leq 0.10$	735	423	57.55	76.63
$0.10 < d \leq 0.25$	570	89	15.61	16.12
$0.25 < d \leq 0.50$	788	40	5.08	7.25
$d > 0.50$	7744	0	0	0

Representative prediction checkpoints are summarized in **Table 11** to illustrate how the continuous margin is translated into specific operational recommendations. For Case 1, the reference and predicted margins are 0.500 and 0.574, respectively. Both values remain above zero and therefore correspond to the normal operating range, allowing the current task to continue. Case 2 yields reference and predicted margins of -0.529 and -0.347 , respectively. Although a numerical difference is present, both margins remain within the elevated-risk interval, and maximum battery temperature is identified as the dominant risk factor. An energy-saving operation is consequently recommended to reduce thermal and electrical loading. For Case 3, the predicted margin of -1.001 closely matches the reference value of -1.000 , indicating an extreme operating condition dominated by minimum voltage. This condition triggers a return operation together with protective measures. These examples show that the proposed decision layer maintains consistent operational interpretations even when moderate numerical deviations occur between the predicted and reference margins.

Table 11. Data-driven operational risk assessment of PCMT at representative prediction checkpoints.

Prediction case	Reference M_H^{ref}	Predicted M_H	Risk level	Dominant risk factor	Recommended action
Case 1	0.500	0.574	Normal operating range	TAE	Continue the current task
Case 2	-0.529	-0.347	Elevated operational risk	Maximum temperature	Apply energy-saving operation
Case 3	-1.000	-1.001	Extreme operating condition	Minimum voltage	Initiate return and activate protective measures

To further evaluate the dependence of the relative risk assessment on the selected quantile boundaries, two alternative boundary configurations are examined in addition to the baseline setting. As shown in **Fig. 12(a)**, the less conservative setting moves the warning and extreme boundaries toward the distribution tails, whereas the more conservative setting shifts them toward the central region. The same boundary adjustment

is applied consistently to TAE and minimum voltage in the lower tail and to maximum battery temperature in the upper tail. As shown in **Fig. 12(b)**, the relative risk assessment remains stable under the two alternative boundary settings. Risk level agreement varies from 94.01% at the baseline to 95.98% and 96.10%, while the optimistic misclassification rate changes from 5.61% to 3.58% under both alternative settings. Dominant factor accuracy remains within a narrow range from 93.30% to 94.71%. These variations indicate that the main decision outcomes are insensitive to reasonable changes in the quantile boundaries. The baseline setting is retained as the predefined nominal configuration rather than being selected through post hoc optimization of the test set metrics.

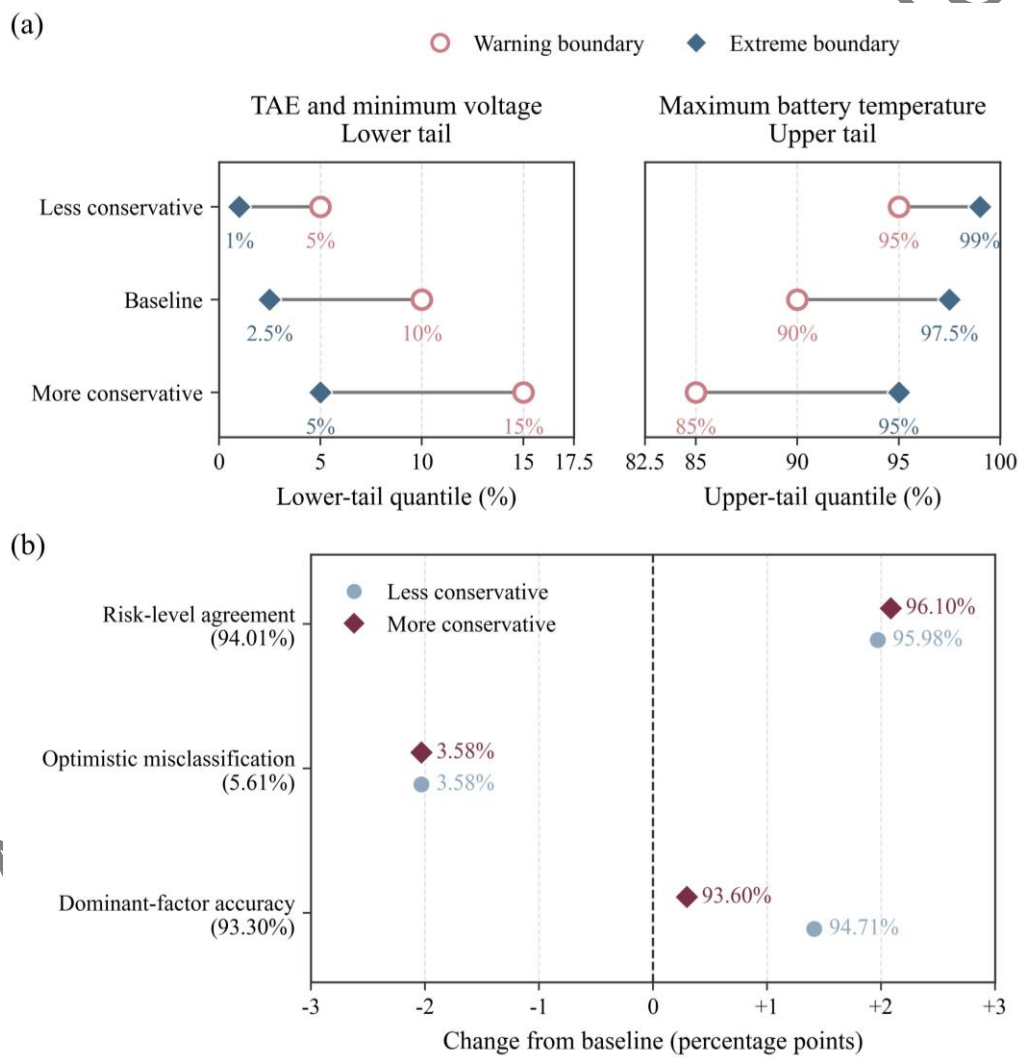


Fig. 12. Sensitivity of the relative risk assessment to alternative quantile boundary settings. (a) Warning and extreme quantiles for TAE and minimum voltage in the lower tail and maximum battery temperature in the upper tail under less conservative, baseline, and more conservative settings. (b) Changes in risk-level

agreement, optimistic misclassification rate, and dominant-factor accuracy relative to the baseline setting. Values next to the markers indicate the corresponding absolute metric values.

A central contribution of this study lies in the direct coupling of multi-task physical prediction with interpretable operational decision support. Rather than using a separate black-box risk classifier, the proposed approach derives each relative risk level from the predicted TAE, minimum voltage, and maximum battery temperature through explicit normalization and minimum margin operations. The dominant physical factor associated with each recommendation is retained throughout this process, allowing each operational action to be traced to a specific predicted battery constraint. The warning and extreme boundaries used in the present study are derived from the calibration distribution because platform-certified protection thresholds are unavailable. Accordingly, the resulting margin represents relative operational risk within the observed operating distribution rather than an absolute hardware safety limit. The agreement between predicted and reference risk classes therefore reflects consistency with the data defined reference system rather than demonstrated hardware safety. When certified voltage, temperature, and reserve energy limits become available, the same decision structure can incorporate these engineering boundaries without changing the PCMT prediction architecture.

4.5. Validation and reliability assessment of TAE

The relationship between TAE and physically delivered energy is further examined over a defined discharge interval. A complementary discharge-limit analysis is performed using complete discharge episodes with a common terminal SOC of 20%. For each episode, the reference energy is obtained by directly integrating the measured battery power from the selected anchor point until the SOC reaches 20%. This provides a physically measured discharge-limit reference that is separate from the 300 s window used to estimate the local energy-to-SOC relationship. The corresponding TAE is referenced to the same 20% SOC limit using the local energy-to-SOC-change ratio. Four progressively stricter anchor-SOC requirements, namely at least 30%, 50%, 60%, and 80%, are considered to examine the effect of increasing extrapolation distance. These requirements yield 57, 47, 42, and 31 valid discharge episodes, respectively. The comparison between TAE and the measured delivered energy is shown in **Fig. 13**, and the quantitative results are summarized in **Table 12**.

As shown in **Fig. 13**, the estimated TAE values remain closely distributed around the identity line for all four anchor-SOC requirements, with most episodes located within the $\pm 5\%$ error band. **Table 12** shows that the mean TAE values remain very close to the corresponding mean measured energies. For anchor SOC values of at least 30%, 50%, and 60%, the NRMSE values are 1.90%, 1.90%, and 1.87%, respectively, while the corresponding R^2 values remain between 0.990 and 0.994. The mean relative biases are also small, ranging from +0.045% to +0.069%. When the anchor-SOC requirement increases to at least 80%, the extrapolation interval becomes longer, and the number of available episodes decreases to 31. Under this more demanding condition, the RMSE increases to 15.95 Wh and R^2 decreases to 0.946, whereas the NRMSE remains limited to 1.94% and the mean bias remains only +0.174%. The limited variation in NRMSE and bias across the progressively stricter anchor conditions indicates that the local energy-to-SOC relationship remains sufficiently stable for estimating the energy delivered to the defined 20% SOC limit. This agreement provides an additional physical validation of TAE beyond comparison with its supervision label constructed over the 300 s horizon and supports its interpretation as a task-conditioned estimate of available energy.

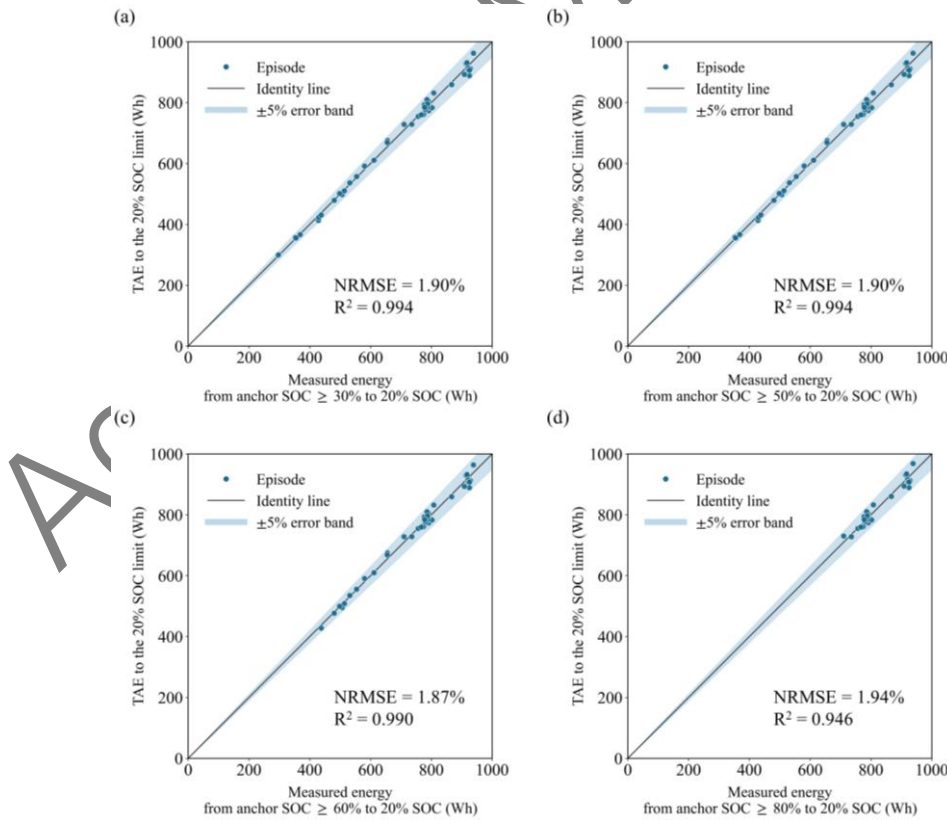


Fig. 13. Validation of TAE against measured energy delivered from the anchor SOC to the 20% SOC discharge limit under different anchor-SOC conditions. (a) Anchor SOC $\geq 30\%$; (b) anchor SOC $\geq 50\%$;

(c) anchor SOC $\geq 60\%$; and (d) anchor SOC $\geq 80\%$. The solid diagonal line denotes ideal agreement, and the shaded region indicates the $\pm 5\%$ error band.

Table 12. Discharge-limit validation of TAE under different anchor-SOC conditions.

Minimum starting SOC	Episodes	Mean measured energy	Mean TAE	RMSE	NRMSE	R ²	Mean bias
$\geq 80\%$	31	824.00 Wh	824.17 Wh	15.95 Wh	1.94%	0.946	+0.174%
$\geq 60\%$	42	749.89 Wh	749.98 Wh	14.06 Wh	1.87%	0.990	+0.045%
$\geq 50\%$	47	710.39 Wh	710.48 Wh	13.48 Wh	1.90%	0.994	+0.058%
$\geq 30\%$	57	701.58 Wh	701.65 Wh	13.33 Wh	1.90%	0.994	+0.069%

The reliability of the constructed TAE labels is evaluated by examining the activation of the TAE reliability mask under different operating conditions. As summarized in **Table 13**, 6,207 of the 61,178 task fragments are excluded from TAE and unit-SOC-energy supervision, corresponding to an overall exclusion rate of 10.15%, while the remaining 89.85% provide valid TAE supervision. The exclusion rate varies across gait and SOC conditions, although the percentages for rough terrain and slope are based on only 2 and 8 fragments and therefore have limited statistical representativeness. A clearer trend appears across the load quartiles. The exclusion rate decreases from 40.03% in Q1 to 0.54%, 0.01%, and 0.00% in Q2, Q3, and Q4, respectively. This behavior is consistent with the TAE formulation because lower loads generally produce smaller SOC reductions over the fixed 300 s horizon, increasing the sensitivity of the energy-to-SOC-change ratio to SOC quantization.

The influence of SOC quantization is further examined in **Table 14**. The original 1% SOC resolution is used as the baseline, while 2% and 5% quantization are introduced as progressively coarser stress-test settings. Under 2% quantization, 91.39% of the fragments remain valid, while the median and 95th-percentile absolute TAE changes reach 23.17% and 50.00%, respectively. With 5% quantization, the valid proportion decreases to 48.06%, and the corresponding TAE changes increase to 40.00% and 59.55%. These results confirm that coarse SOC quantization can substantially perturb the constructed TAE label, thereby supporting the use of reliability-aware supervision and a minimum SOC decrease criterion. The sensitivity to the minimum accepted SOC decrease is summarized in **Table 15**. Under the 1% SOC resolution and the strict criterion $\Delta_{SH} > \Delta_{Smin}$, $\Delta_{Smin} = 0.010$ and 0.015 both correspond to an effective minimum accepted SOC decrease of 2% and retain 89.85% of the fragments. Increasing the threshold to 0.020 raises the effective minimum decrease to 3% and reduces the retained proportion to 53.29%. The adopted value of 0.015

therefore provides a practical balance between limiting the influence of SOC quantization and preserving sufficient task fragments for TAE supervision.

Table 13. Exclusion rates of the TAE reliability mask across operating conditions.

Condition	Group	Total fragments	Excluded fragments	Exclusion rate (%)
Overall	All	61178	6207	10.15
Gait	Walking	60739	6093	10.03
	Rough terrain	2	1	50.00
	Slope	8	8	100.00
	Stair sensing	57	2	3.51
	Continuous stairs	372	103	27.69
SOC	0 to 20%	781	683	87.45
	20 to 40%	11822	1449	12.26
	40 to 60%	14238	707	4.97
	60 to 80%	16791	848	5.05
	80 to 100%	17546	2520	14.36
Load*	Q1	15295	6123	40.03
	Q2	15294	83	0.54
	Q3	15296	1	0.01
	Q4	15293	0	0.00

*Load quartiles are defined according to the mean discharge power over the 300 s prediction horizon: Q1 $\leq$ 328.33 W, Q2 328.33 to 361.65 W, Q3 361.65 to 385.97 W, and Q4 $>$ 385.97 W. SOC groups are defined using the current SOC at the prediction instant. The gait specific percentages for rough terrain and slope should be interpreted cautiously because of the limited number of corresponding task fragments.

Table 14. Sensitivity of TAE labels to SOC quantization.

SOC quantization	Valid TAE fragments	Valid proportion (%)	Common valid fragments	Median absolute TAE change (%)	percentile absolute TAE change (%)
1% baseline	54971	89.85	54971	0.0	0.00
2%	55912	91.39	54971	23.17	50.00
5%	29405	48.06	29137	40.00	59.55

Table 15. Sensitivity of TAE sample retention to the minimum accepted SOC decrease.

ΔS_{min}	Effective minimum SOC decrease	Retained TAE fragments	Retained proportion (%)	Excluded fragments	Exclusion rate (%)	Median TAE (Wh)
0.010	2%	54971	89.85	6207	10.15	739.79
0.015 baseline	2%	54971	89.85	6207	10.15	739.79
0.020	3%	32604	53.29	28574	46.71	572.47

4.6. Discussion and limitations

The results obtained in this study demonstrate that the proposed PCMT framework maintains stable predictive performance across a broad range of operating conditions within the investigated quadruped robot platform. The 85-day dataset covers multiple locomotion and terrain conditions, including walking, rough-terrain traversal, slope locomotion, and stair-related tasks, together with substantial variations in electrical

load, SOC, battery voltage, and thermal states. These conditions introduce heterogeneous electro-thermal and motion demands and therefore provide a meaningful basis for evaluating the robustness of task-conditioned energy prediction and relative risk assessment under practical operation. The consistent performance observed across these diverse operating states indicates that the proposed framework captures relationships that extend beyond a single fixed task or operating regime. Although the operating conditions are diverse, all data are collected from the same quadruped robot platform, with unchanged hardware, battery configuration, sensing architecture, and control system. The present results therefore support within-platform robustness across heterogeneous operating conditions, while cross-platform transferability remains to be established. In addition, several sources of domain variation cannot be independently isolated in the current dataset. Controlled payload mass is not explicitly recorded, ambient environmental variables such as external temperature and humidity are not independently measured, and battery-aging states are not characterized using capacity-fade, cycle-count, or state-of-health indicators.

An additional applicability consideration concerns a newly commanded gait or task within the prediction horizon of 300 s. PCMT does not use an explicit future mission plan during inference, and the prediction therefore represents a context-conditioned forecast based on the recent operating history and current observed state. The early gait or task-phase changes mainly increase the prediction uncertainty of future energy consumption and minimum voltage, while the maximum-temperature prediction remains comparatively stable. Accordingly, the prediction over the 300 s horizon is most directly applicable when subsequent operation remains consistent with the current operating context or with transition patterns represented in the training data. If a new command changes the operating state after the prediction instant, the previous forecast reflects the preceding context and can be updated once the new gait, task phase, or operating state becomes observable.

These considerations define the current applicability boundary of the proposed framework rather than reducing its value for practical robot energy management. The present study establishes a unified task-conditioned prediction and decision-support framework under realistic and diverse operating conditions, while broader validation represents the next stage toward deployment across heterogeneous robot systems. Future studies can extend the current framework using multi-platform datasets, controlled payload experiments, explicit environmental measurements, and long-term battery-health tracking. These extensions

can support systematic evaluation of platform-level and condition-level domain shifts and transferability under broader deployment conditions.

5. Conclusions

This study proposes a task-conditioned energy assessment framework for battery-powered mobile robots by introducing TAE and a predictive relative risk margin. TAE characterizes the effective energy reserve available under the current operating condition, extending conventional SOC-based assessment toward task-level energy availability. The proposed PCMT jointly predicts TAE, future energy consumption, minimum voltage, and maximum battery temperature through physics-constrained multi-task learning. Across the four prediction targets, the NRMSE values range from 0.59% to 1.60%, with R^2 values between 0.972 and 0.997. Compared with the strongest corresponding baselines, the NRMSE values for TAE and future energy consumption are reduced by 72.4% and 65.6%, respectively. The predictive relative risk margin further integrates energy, voltage, and thermal constraints, achieving an R^2 of 0.993, a risk-level agreement of 94.01%, and a dominant-risk-factor accuracy of 93.30%. These results demonstrate that task-conditioned energy availability can be predicted accurately and converted into interpretable operational guidance for mobile robots.

Despite these promising results, several limitations remain. Validation is currently based on operating data from one quadruped robot platform, and broader investigation is required across robot configurations, payloads, battery aging states, and environmental conditions. Future research can further integrate battery condition, task-dependent energy demand, environmental interaction, and operational risk into unified energy management architectures, allowing energy information to influence mission scheduling, task allocation, and autonomous decision-making. This broader direction can move mobile robot energy management beyond battery state monitoring toward task-aware and decision-oriented energy intelligence for long-duration operation in complex environments.

Acknowledgement

This work is supported by the National Key Research and Development Program of China (2024YFE0115800), the Introduced Innovative R&D Team of Guangdong (2023ZT10L145), the Guangdong Basic and Applied Basic Research Foundation (2025A1515140131), and the Fundamental

Research Funds for the Central Universities. All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of sponsors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this manuscript, OpenAI ChatGPT was used solely to improve the readability and language of the text. The manuscript was subsequently reviewed and edited by the authors, who take full responsibility for its content.

Data availability

The dataset is confidential, and the authors do not have permission to share it.

Code availability

The code related to the modeling framework is available at <https://github.com/Translab-SCUT/Energy-Consumption-Prediction-for-Mobile-Robots> or from the corresponding author upon reasonable request.

References

- [1] Mon-Williams, R., Li, G., Long, R., Du, W., & Lucas, C. G. (2025). Embodied large language models enable robots to complete complex tasks in unpredictable environments. *Nature Machine Intelligence*, 7(4), 592-601.
- [2] Gupta, A., Savarese, S., Ganguli, S., & Fei-Fei, L. (2021). Embodied intelligence via learning and evolution. *Nature Communications*, 12, 5721.
- [3] Lee, J., Bjelonic, M., Reske, A., Wellhausen, L., Miki, T., & Hutter, M. (2024). Learning robust autonomous navigation and locomotion for wheeled-legged robots. *Science Robotics*, 9(89), eadi9641.
- [4] Kang, J.-G., Park, J., Song, T.-G., Kim, J.-H., Hong, S., & Park, H.-W. (2026). Agile perceptive multiskill locomotion for quadrupedal robots in the wild. *Science Robotics*, 11(116), eadz7397.

-
- [5] Tu, H., Borah, M., Moura, S., Wang, Y., & Fang, H. (2024). Remaining discharge energy prediction for lithium-ion batteries over broad current ranges: A machine learning approach. *Applied Energy*, 376, 124086.
- [6] Jing, H., Ou, S., Lv, Z., Guo, H., Burke, A. F., & Zhao, J. (2026). Battery Safety: Mechanisms, Monitoring, and Machine Intelligence. *Advanced Energy Materials*, e71093.
- [7] Zheng, L., Zhu, J., Wang, G., He, T., & Wei, Y. (2016). Novel methods for estimating lithium-ion battery state of energy and maximum available energy. *Applied Energy*, 178, 1-8.
- [8] Xu, Y., Chen, J., Zeng, T., Peng, W., Guo, R., Tian, J., & Chung, C. Y. (2026). Machine learning-assisted health management of retired batteries. *Journal of Energy Chemistry*, 123, 436-459.
- [9] Zhang, X., Wang, Y., Wu, J., & Chen, Z. (2018). A novel method for lithium-ion battery state of energy and state of power estimation based on multi-time-scale filter. *Applied Energy*, 216, 442-451.
- [10] Lai, X., Huang, Y., Gu, H., Han, X., Feng, X., Dai, H., Zheng, Y., & Ouyang, M. (2022). Remaining discharge energy estimation for lithium-ion batteries based on future load prediction considering temperature and ageing effects. *Energy*, 238, 121754.
- [11] Miki, T., Lee, J., Hwangbo, J., Wellhausen, L., Koltun, V., & Hutter, M. (2022). Learning robust perceptive locomotion for quadrupedal robots in the wild. *Science Robotics*, 7(62), eabk2822.
- [12] Hwangbo, J., Lee, J., Dosovitskiy, A., Bellicoso, C. D., Tsounis, V., Koltun, V., & Hutter, M. (2019). Learning agile and dynamic motor skills for legged robots. *Science Robotics*, 4(26), eaau5872.
- [13] Lee, J., Hwangbo, J., Wellhausen, L., Koltun, V., & Hutter, M. (2020). Learning quadrupedal locomotion over challenging terrain. *Science Robotics*, 5(47), eabc5986.
- [14] Yang, C., Yuan, K., Zhu, Q., Yu, W., & Li, Z. (2020). Multi-expert learning of adaptive legged locomotion. *Science Robotics*, 5(49), eabb2174.
- [15] Jin, Y., Liu, X., Shao, Y., Wang, H., & Yang, W. (2022). High-speed quadrupedal locomotion by imitation-relaxation reinforcement learning. *Nature Machine Intelligence*, 4, 1198-1208.
- [16] Choi, S., Ji, G., Park, J., Kim, H., Mun, J., Lee, J. H., & Hwangbo, J. (2023). Learning quadrupedal locomotion on deformable terrain. *Science Robotics*, 8(74), eade2256.
- [17] Hoeller, D., Rudin, N., Sako, D., & Hutter, M. (2024). ANYmal parkour: Learning agile navigation for quadrupedal robots. *Science Robotics*, 9(88), eadi7566.

-
- [18] Jenelten, F., He, J., Farshidian, F., & Hutter, M. (2024). DTC: Deep tracking control. *Science Robotics*, 9(86), eadh5401.
- [19] Kim, H., Oh, H., Park, J., Kim, Y., Youm, D., Jung, M., Lee, M., & Hwangbo, J. (2025). High-speed control and navigation for quadrupedal robots on complex and discrete terrain. *Science Robotics*, 10(102), eads6192.
- [20] Ruppert, F., & Badri-Spröwitz, A. (2022). Learning plastic matching of robot dynamics in closed-loop central pattern generators. *Nature Machine Intelligence*, 4, 652-660.
- [21] Aractingi, M., Léziart, P.-A., Flayols, T., Perez, J., Silander, T., & Souères, P. (2023). Controlling the Solo12 quadruped robot with deep reinforcement learning. *Scientific Reports*, 13, 11945.
- [22] Han, L., Zhu, Q., Sheng, J., Zhang, C., Li, T., Zhang, Y., ... & Zhang, Z. (2024). Lifelike agility and play in quadrupedal robots using reinforcement learning and generative pre-trained models. *Nature Machine Intelligence*, 6(7), 787-798.
- [23] Shafiee, M., Bellegarda, G., & Ijspeert, A. J. (2024). Viability leads to the emergence of gait transitions in learning agile quadrupedal locomotion on challenging terrains. *Nature Communications*, 15, 3073.
- [24] Humphreys, J., & Zhou, C. (2025). Learning to adapt through bio-inspired gait strategies for versatile quadruped locomotion. *Nature Machine Intelligence*, 7, 1141-1153.
- [25] Buchner, T. J., Fukushima, T., Kazemipour, A., Gravert, S. D., Prairie, M., Romanescu, P., ... & Katzschmann, R. K. (2024). Electrohydraulic musculoskeletal robotic leg for agile, adaptive, yet energy-efficient locomotion. *Nature Communications*, 15(1), 7634.
- [26] Stella, F., Achkar, M. M., Della Santina, C., & Hughes, J. (2025). Synergy-based robotic quadruped leveraging passivity for natural intelligence and behavioural diversity. *Nature Machine Intelligence*, 7, 386-399.
- [27] Sihite, E., Kalantari, A., Nemovi, R., Ramezani, A., & Gharib, M. (2023). Multi-Modal Mobility Morphobot with appendage repurposing for locomotion plasticity enhancement. *Nature Communications*, 14, 3323.
- [28] Dogru, S., & Marques, L. (2018). A physics-based power model for skid-steered wheeled mobile robots. *IEEE Transactions on Robotics*, 34(2), 421-433.
- [29] Sadrpour, A., Jin, J., & Ulsoy, A. G. (2013). Mission energy prediction for unmanned ground vehicles using real-time measurements and prior knowledge. *Journal of Field Robotics*, 30(3), 399-414.

-
- [30] Pentzer, J., Brennan, S., & Reichard, K. (2022). On-line estimation of power model parameters for skid-steer robots with applications in mission energy use prediction. *Journal of Field Robotics*, 39(6), 763-782.
- [31] Poskart, B., Iskierka, G., Krot, K., Burduk, R., Gwizdal, P., & Gola, A. (2022). Multi-parameter predictive model of mobile robot's battery discharge for intelligent mission planning in multi-robot systems. *Sensors*, 22(24), 9861.
- [32] Shafaei, S. M., & Mousazadeh, H. (2023). Characterization of motion power loss of off-road wheeled robot in a slippery terrain. *Journal of Field Robotics*, 40(1), 57-72.
- [33] Mohamadi, P. H. A., Khorasani, A., Verstraten, T., & Vanderborght, B. (2025). A hybrid parameters estimation approach for power consumption modeling of ground mobile robots with unknown payload. *Journal of Field Robotics*, 42(5), 1840-1860.
- [34] Shahsavari, S., Haghbayan, H., Miele, A., Immonen, E., & Plosila, J. (2025). A coordinated approach to control mechanical and computing resources in mobile robots. *IEEE Transactions on Robotics*, 41, 347-363.
- [35] Wang, S., Fu, Y., Zhang, W., Fernandez, C., & Blaabjerg, F. (2026). Review on improved neural network algorithms for battery state of energy estimation in smart grids. *Renewable and Sustainable Energy Reviews*, 231, 116797.
- [36] Wang, S., Zhang, S., Wen, S., & Fernandez, C. (2024). An accurate state-of-charge estimation of lithium-ion batteries based on improved particle swarm optimization-adaptive square root cubature Kalman filter. *Journal of Power Sources*, 624, 235594.
- [37] Wang, S., Wu, Y., Zhou, H., Zhang, Q., Fernandez, C., & Blaabjerg, F. (2025). Improved particle swarm optimization-adaptive dual extended Kalman filtering for accurate battery state of charge and state of energy joint estimation with efficient core factor feedback correction. *Energy*, 322, 135686.
- [38] Lin, Z., Huang, Z., Yang, S., Wu, C., Fang, S., Liu, Z., Yang, M., Xie, Z., Li, D., & Zou, Y. (2024). Survey on task-centric robot battery management: A neural network framework. *Journal of Power Sources*, 610, 234674.