Measurement 265 (2026) 120400

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Multimodal fusion of optimized LSTM for estimating dynamic oxygen uptake based on consumer wearable devices

Zhengsheng Hu [a,b], Gongcheng Xu [a], Yan Wang [a,b,c], Guanwen Qu [a], Puxiang Lai [a,b,d], Wenzhao Li [a], Jiahui Yin [a], Hui Xie [a], Huiwen Xiao [a], Tao Li [e], Ming Zhang [a,b,c,*], Qitao Tan [a,b,c,*]

[a] Department of Biomedical Engineering, Faculty of Engineering, The Hong Kong Polytechnic University, Hong Kong 999077, China
[b] Research Institute for Sports Science and Technology, The Hong Kong Polytechnic University, Hong Kong 999077, China
[c] The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen 518057, China
[d] Joint Research Center for Biosensing and Precision Theranostics, The Hong Kong Polytechnic University, Hong Kong 999077, China
[e] Key Laboratory of Traffic Safety on Track, Ministry of Education, School of Traffic & Transportation Engineering, Central South University, Changsha 410075, China

ARTICLE INFO

Keywords:
Oxygen uptake prediction
Multimodal feature fusion
Long short-term memory
Particle swarm optimization

ABSTRACT

Accurate and real-time monitoring of oxygen consumption dynamics during exercise is essential for optimizing training intensity, duration, and recovery strategies. However, existing methods often compromise between computational efficiency and predictive accuracy, either by employing large models for feature extraction or by oversimplifying input variables, which limits their applicability to consumer wearable devices requiring real-time and energy-efficient processing. To address these challenges, we propose a novel framework that integrates Multimodal Feature Fusion and Particle Swarm Optimization (PSO) based on wearable device data. The model utilizes Long Short-Term Memory (LSTM) networks to extract temporal features from anthropometric, physiological, and kinematic indicators, while PSO optimizes feature weights to enhance prediction accuracy. The proposed framework therefore presents a 14.3% reduction in root mean squared error and 24% reduction in model size for dynamic oxygen uptake (VO_2) prediction and robust performance across various moderate intensity running speeds. This multimodal fusion approach provides an effective and efficient solution for VO_2 prediction, facilitating real-time deployment on consumer wearable devices.

1. Introduction

Accurate monitoring of energy expenditure is significant for individuals, as it facilitates quantitative assessment of exercise performance, cardiopulmonary function, and postural stability [1]. Oxygen uptake (VO_2) is considered the gold standard for estimating energy expenditure during physical activities, reflecting the components of aerobic metabolisms [2]. Dynamic VO_2 data encompasses critical information regarding oxygen delivery of the cardiopulmonary system and oxygen metabolism of the musculoskeletal system [3]. Consequently, it is essential for individuals to accurately monitor oxygen consumption dynamics in real-time during exercise, enabling them to optimize training intensity, duration, and recovery protocols.

Traditional measurements of oxygen consumption rely on laboratory-based metabolic chambers utilizing the indirect calorimetry method [4], which are constrained by controlled laboratory conditions and substantial costs. In recent years, portable measurement devices such as the COSMED K5 (COSMED, Italy) reduced environment limitations, making it possible to collect data during outdoor activities like running [5,6], through breath-by-breath analysis of oxygen (O_2) and carbon dioxide (CO_2) volumes. These devices are currently employed in scientific research, clinical medicine, and professional athletics [7]. However, their widespread adoption among recreational runners remains limited due to high costs, operational complexity, and the impracticality of prolonged mask use in daily settings.

Wearable sensor-based VO_2 estimation makes it possible to overcome the above limitations [8,9]. Collected physiological indicators such as heart rate (HR) exhibit a strong correlation with oxygen uptake and have been widely utilized to predict VO_2 and maximal oxygen uptake (VO_{2max}) using regression methods [10]. Kinematic data acquired from accelerators or multi-axis inertial measurement units (IMUs) also is used as a feature to estimate VO_2, capturing rich information of

* Corresponding authors at: Department of Biomedical Engineering, Faculty of Engineering, The Hong Kong Polytechnic University, Hong Kong 999077, China.
E-mail addresses: ming.zhang@polyu.edu.hk (M. Zhang), qitao.tan@polyu.edu.hk (Q. Tan).

https://doi.org/10.1016/j.measurement.2026.120400
Received 22 September 2025; Received in revised form 15 December 2025; Accepted 9 January 2026
Available online 10 January 2026

biomechanical movement patterns [11]. Advances in machine learning have enabled more detailed and precise estimation by extracting complex coupled relationships from multi-source features [12,13]. For example, Hsiao et al. estimated the VO_2 based on photoplethysmography (PPG) using machine learning approaches like the random forest [14], which relied on a specific device with five wavelengths. Hong and Sun achieved continuous VO_2 estimation relying on HR, time, weight and other data via XGBoost algorithm [15], while the results of their study had high risk of data leakage due to the splitting method using whole dataset. To enhance the prediction performance, deep learning (DL) models were utilized. Among these, Long Short-Term Memory (LSTM) networks stand out for their ability to model temporal dependency and capture nonlinear relationships by understanding complex patterns and trends in time-series data [16], and they have been widely used in the fields of disease diagnosis [17,18], industry control [19], etc. Therefore, LSTM has the potential to model the complex relationship between VO_2 and other physiological signals. Wang et al. developed an attention-enhanced Convolutional Neural Network- Long Short-Term Memory (CNN-LSTM) model integrating resting and dynamic physiological features for exercise oxygen consumption prediction [20], but the resting oxygen consumption had to been collect as the static feature. Vidyarani et al. proposed a hybrid LSTM-transfer learning model for VO_{2max} prediction using IMU data [21], but the predicted result was a static value based on exercise tasks lasting for minutes. Davidson et al. predicted VO_2 using an LSTM model fusing HR, IMU, and inertial navigation/global positioning system (INS/GPS) data [22], but limited by extra devices and cross-participant performance. Bangaru et al. introduced a bidirectional LSTM model for VO_2 estimation by integrating IMU and electromyography (EMG)signals [23], however this method requires 4 high-frequency EMG channels of 200 Hz leading to substantial computational overhead. Although LSTM methods demonstrate robust performance in predicting oxygen uptake from HR and IMU data. Existing approaches often trade off computational efficiency for accuracy, such as using large CNN architecture or hundreds of hidden units to extract features from raw signals of several sensors or oversimplifying models with HR-only inputs. This imbalance hinders deployment on consumer wearables, such as smartwatches, where real-time processing and energy efficiency are critical.

The hybrid models combining LSTM with Particle Swarm Optimization (PSO) have demonstrated remarkable performance [24]. As a population-based heuristic search strategy, PSO effectively optimizes model parameters and feature weights relying on its ability to explore the search space, leading to performance improvement in temporal prediction tasks. Therefore, we propose introducing PSO to LSTM to reasonably distribute weights of features from different devices to improve VO_2 prediction accuracy and eliminate redundant features to balance computing cost.

We propose a multimodal fusion framework leveraging Particle Swarm Optimization (PSO) for accurate and efficient oxygen uptake prediction using consumer wearable devices. Our methodology employs LSTM networks to extract temporal features from each input indicator, including anthropometric features, physiological features, and kinematic features. These features are adaptively fused through PSO-optimized weighted integration, transforming feature selection into a constrained optimization problem. The PSO algorithm dynamically adjusts contribution weights across modalities, enhancing both prediction accuracy by prioritizing informative features and model interpretability by quantifying feature importance. This optimized fusion strategy significantly improves upon conventional approaches.

The structure of this study is organized as follows: Section 2 presents the computational framework of PSO-based multi-modal feature fusion method using wearable devices. Section 3 validates the proposed approach through cross-participant evaluation, where models trained on 22 participants demonstrate generalizable performance when tested on 4 held-out participants. This experimental paradigm achieves satisfactory VO_2 prediction accuracy and obtains optimal weights for feature fusion. Section 4 discusses the prediction performance of VO_2 dynamic characteristics and the consistent at different running paces. Section 5 concludes the entire paper.

2. Material and method

The study developed a novel multi-modal fusion method based on consumer wearable devices and Particle Swarm Optimization approach for oxygen uptake prediction during running. The approach integrates anthropometric, kinetic and physiological indicators captured by wearable sensors during exercise, employing LSTM networks for temporal feature extraction from raw sequential data, and utilizing PSO algorithm to determine optimal weighting coefficients for multi-modal feature fusion. The following section details the theoretical framework and implementation of the methodology.

2.1. Wearable device data collection

This study collected physiological and kinematic indicators by integrating multiple consumer wearable devices with the COSMED K5, as shown in Fig. 1. Kinematic data were acquired using the Stryd footpod (Stryd Inc., USA) and the GT RUNNER smartwatch (HUAWEI, China). Stryd is a validated wearable motion sensor that combines IMU, barometric, and anemometric sensors to derive running metrics including ground contact time (GCT), vertical oscillation (VO), speed, and cadence [25]. The GT RUNNER smartwatch provided left-wrist-worn triaxial acceleration, angular velocity, via its embedded six-axis IMU. As for physiological indicators, the criterion measure of oxygen uptake was obtained through breath-by-breath analysis using the COSMED K5 system and continuous heart HR was measured using the GT RUNNER smartwatch. Participants completed standardized running trials consisting of three 180-second stages: (1) easy pace, (2) marathon pace, and (3) lactate threshold pace, with individualized speeds based on historical performance data. To control fatigue effects, trials were administered in ascending intensity order, with mandatory over 10-minute recovery intervals between stages until participants achieved pre-exercise resting HR levels.

2.2. Model development

Multi-modal feature fusion prediction represents an advanced computational approach that integrates heterogeneous data features from diverse information sources to optimize prediction model accuracy. The predictive performance of such models is critically dependent on the optimal weighting allocation among constituent features. In this section, three-level architectures of the model are introduced: the feature extraction based on LSTM, the weight optimization based on PSO, and the development of Swarm Optimization Multi-modal Feature Fusion.

2.2.1. LSTM

LSTM is a specialized recurrent neural network (RNN) method, demonstrating particular efficacy in physiological signal monitoring and sequence prediction tasks [26,27]. As illustrated in Fig. 2(a), the distinctive LSTM architecture, primarily the hidden states and cell states, enables key-feature extraction from time-series through gating mechanisms that optimally balance long-term dependencies with short-term input patterns [28]. These state vectors maintain compressed representations of the network's learned temporal dynamics, effectively encoding the historical context of input sequences. The LSTM-derived feature vectors exhibit superior capability in capturing the intrinsic characteristics of time-series physiological data compared to conventional approaches. Importantly, in trained network architectures, these intermediate feature representations provide a rich, information-dense foundation that significantly enhances downstream prediction accuracy and enables more insightful physiological analyses.

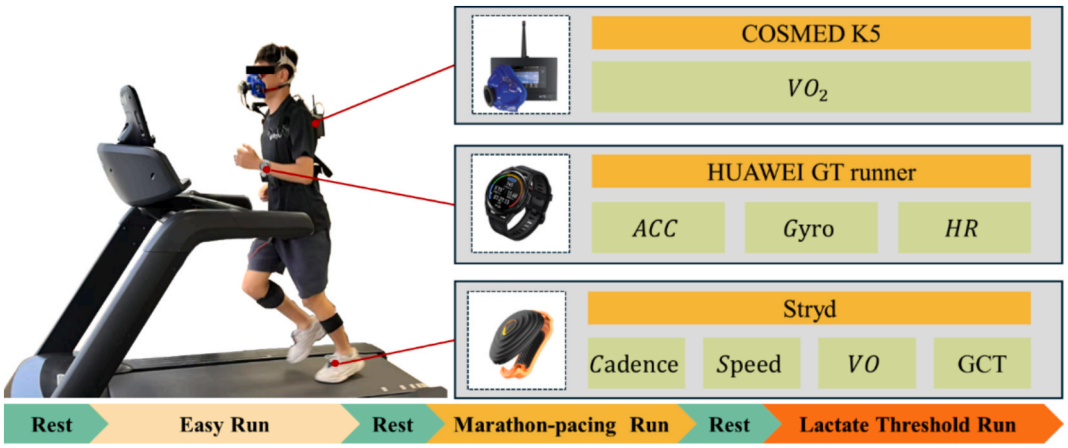

Fig. 1. Schematic of data collection during running.

2.2.2. PSO

The PSO algorithm, originally developed by psychologist James Kennedy and electrical engineer Russell Eberhart, is a metaheuristic optimization technique inspired by the collective foraging behavior of bird flocks. This approach simulates the social dynamics of particle swarms, where individual particles adjust their positions and velocities within the search space to collectively locate optimal solutions for complex non-deterministic polynomial problems [29,30]. In our framework, we employ PSO to optimize the weight distribution of multi-modal features, thereby enhancing prediction accuracy through intelligent feature fusion. The standard PSO implementation proceeds with the following steps:

Assuming a target search space of dimensionality D, composed of N particles forming a swarm, where the i^{th} particle is represented as a D-dimensional vector as shown in Eq. (1).

$$X_i = (x_{i1}, x_{i2}, \bullet\bullet\bullet, x_{iD}), i = 1, 2, \bullet\bullet\bullet, N \tag{1}$$

The flight velocity of the i^{th} particle is also a D-dimensional vector, as shown in Eq. (2).

$$V_i = (v_{i1}, v_{i2}, \bullet\bullet\bullet, v_{iD}), i = 1, 2, \bullet\bullet\bullet, N \tag{2}$$

The best position found by the i^{th} particle up to the current step is referred to as the individual best position, denoted as Eq. (3).

$$P_{best} = (p_{i1}, p_{i2}, \bullet\bullet\bullet, p_{iD}), i = 1, 2, \bullet\bullet\bullet, N \tag{3}$$

The best position found by the entire particle swarm up to the current step is referred to as the best global position, as shown in Eq. (4).

$$P_{global} = \left(p_{g1}, p_{g2}, \bullet\bullet\bullet, p_{gD}\right) \tag{4}$$

For the next iteration, all particles will update velocities and positions according to Eqs. (5) and (6).

$$v_{ij}(t+1) = \omega \bullet v_{ij}(t) + c_1 r_1(t)\left[p_{ij}(t) - x_{ij}(t)\right] + c_2 r_2(t)\left[p_{gj}(t) - x_{ij}(t)\right] \tag{5}$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \tag{6}$$

where, X_i is the position of the particle, V_i represents the speed of the particle, P_{best} is the particle's own best position, P_{global} is the best position of the particles in the swarm, $v_{ij}(t)$ and $x_{ij}(t)$ represent the velocity and value of the j^{th} dimension in the i^{th} particle, in which $j = 1, 2, \bullet\bullet\bullet, D$ and $i = 1,2,\bullet\bullet\bullet,N$; c_1 and c_2 are learning factors, also known as acceleration constants, r_1 and r_2 are uniformly distributed random numbers in the range [0,1], which are utilized to add randomness to the particle's flight, ω is inertia weight and make particle keep inertia of movement. As illustrated in Fig. 2(b), the velocity update equation (Eq. (5)) comprises three fundamental components that collectively govern particle dynamics in the optimization process. The first component is called inertial term, preserving the particle's current momentum, facilitating global exploration of the solution space. The second component is the cognitive term, attracting the particle toward its personal historical optimum position, representing individual learning capabilities. The third component is defined as social term, directing the particle toward the swarm's global optimum position, embodying collective intelligence through information sharing within the population.

2.2.3. Proposed method

The proposed model architecture aims to enhance the accuracy and robustness of dynamic VO_2 prediction using consumer wearable devices. As illustrated in Fig. 2(c), input data streams are processed by 29 independent LSTM networks to facilitate initial feature extraction. Subsequently, a feature fusion module is employed to optimize the weights of individual features. Feature fusion plays a critical role in multimodal prediction tasks by integrating data from diverse sources to construct a comprehensive feature matrix for subsequent predictive modeling. In this study, we introduce an optimized weight reallocation strategy to fuse 29 temporal features extracted from wearable sensor data using an LSTM network architecture to enhance prediction accuracy. As depicted in Fig. 2(d), each raw feature weight is treated as an optimization variable, represented as $\omega = \{\omega_1, \omega_2, \omega_3, \bullet\bullet\bullet, \omega_{29}\}$. PSO algorithm is used to optimize the fusion weights, minimizing the loss function of the final LSTM prediction model. The optimized weights enable more effective multi-modal feature fusion, improving the performance of the subsequent VO_2 prediction network performance. Relying on combining PSO's feature fusion capability with LSTM's temporal sequential modeling dynamics, our approach achieves high precision in complex prediction tasks. After weight optimization, weights are constant and multiplied by the initial feature matrix, and stream into LSTM VO_2 prediction networks again to calculate the VO_2 value. It is worth noting that the PSO will not be carried out in the test dataset due to the weight configuration process has been completed. The pseudocode of the proposed approach is presented in Table 1 below.

3. Experiment

3.1. Dataset construction

This study enrolled twenty-six recreational runners whose mean weekly duration is more than 3 h. Participants' essential demographic characteristics are shown in Table 2. All candidates successfully completed comprehensive health assessments to exclude individuals with contraindications to vigorous exercise, including confirmed absence of chronic conditions such as hypertension or diabetes mellitus. The study was approved by the Human Subjects Ethics Sub-committee of

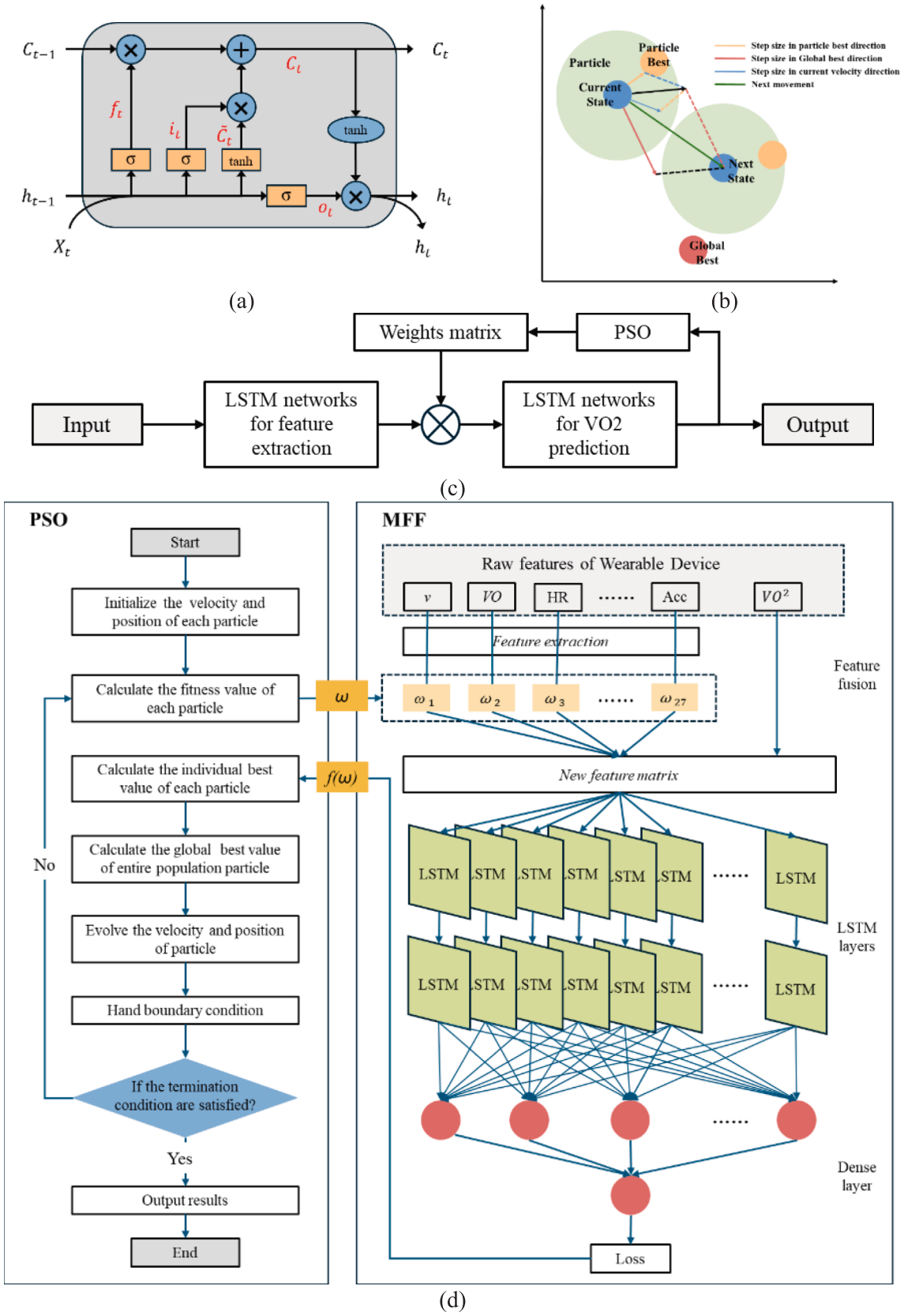

Fig. 2. Model development: (a) Diagram of LSTM network structure; (b) The update of particle velocity and position in PSO; (c) Flowchart of our proposed model; (d) The framework of the feature fusion and PSO architecture.

Table 1
Pseudocode of proposed method.

Global Parameters: Feature set size $D = 29$, Time steps $T = 5$.
PSO Parameters: Swarm size N, Maximum iterations $MaxIter$.
Phase 1: Feature Extraction (Pre-computation)
For each raw feature j in $\{1, \cdots, D\}$ do:
  Data Preprocessing:
    Load raw time-series data sequence R_j.
    Normalize R_j using Min-Max scaler to range [0,1].
    Generate sliding window samples S_j with window size T.
  Extractor Training:
    Construct a base LSTM model M_{base}.
      – LSTM layer (70 units, return sequences).
      – Dropout (0.2) → LSTM (20 units) → Activation ('swish').
      – Dense layer (Output).
    Train M_{base} using S_j.
  Extraction:
    Extract latent features L_j from the output of the first LSTM layer of M_{base}
    Save $L_j$as the extracted feature matrix for feature j.
End For
Phase 2: PSO-based Feature Fusion and Optimization
  Initialization:
    – Load all extracted feature matrices $L = \{L_1, L_2, \cdots, L_D\}$.
    – Initialize swarm particles X (representing weight vector for D features).
    – Initialize velocity V randomly.
    – Evaluate initial fitness and set P_{best} and G_{best}.
  Optimization Loop:
    For iteration $k = 1$ to $MaxIter$ do
      For each particle i in the swarm do
        Update Velocity and Position:
          $V_i = \omega \times V_i + c_1 \times r_1 \times (P_{best\_i} - X_i) + c_2 \times r_2 \times (G_{besti} - X_i)$
          $X_i = X_i + V_i(Clipweightsto[0, 1])$
        Evaluate Fitness F_i:
          a. Weighted Fusion:
            Apply weights X_i to extracted features: $L_j' = L_j \times X_{ij}$.
            Concatenate all L_j' along the feature axis to form tensor W.
          b. Deep Regression Model:
            Feed W into fusion model architecture:
            – LSTM layer (70 units, return sequences).
            – Dropout (0.2) → LSTM (20 units) → Activation ('swish').
            – Dense layer (Output).
          c. Training & Loss:
        Train the model on Training Set.
          Calculate MSE
        Update Pbest:
          If $F_i < F(P_{best\_i})$ then set $P_{best\_i} = X_i$.
      End For
      Update G_{best}:
        Set G_{best}=P_{best_i} where i has the minimum fitness in the swarm.
    End For
  Output: Optimized weights G_{best} and prediction values Y using the final trained model.

Table 2
Essential demographic characteristic.

	Male (n = 20)	Female (n = 6)
Age	34.0 ± 9.2	34.8 ± 12.8
Height (m)	1.75 ± 0.05	1.63 ± 0.07
Weight (kg)	67.5 ± 7.7	52.5 ± 6.0
BMI (kg/m^2)	22.0 ± 2.1	19.6 ± 1.0

the Hong Kong Polytechnic University (Ref. No. HSEARS20221201002). Written informed consent was obtained from all participants prior to study commencement.

The running protocol comprised three standardized intensity conditions including easy running, marathon-pacing running, and lactate threshold running, with individualized speed prescriptions based on participants' historical running records. To eliminate fatigue effects, the treadmill protocol followed an ascending intensity sequence, with mandatory recovery intervals between stages until participants achieved pre-exercise resting heart rate. Throughout the trials, the physiological and kinematic data of participants were captured using three devices: COSMED K5, WATCH GT RUNNER, and Stryd, as shown in Fig. 1. All devices were time-synchronized at initialization to ensure temporal alignment of multimodal data streams.

A preprocessing pipeline was implemented to ensure data quality for predictive modeling. For metabolic data, the COSEMED K5 was set as the breath-by-breath mode to collect VO_2 of each breath. The raw data underwent sequential preprocessing steps including standard deviation-based outlier removal, resampling to 1.0 Hz, and 10-second sliding window smoothing. The HR was collected using the smartwatch and then resampled to 1.0 Hz. The posture parts of kinematic features, including cadence (steps/min), GCT (mm), VO (cm), and speed (m/s), were sampled at 1.0 Hz via Stryds mounted on the shoes. The three-axis acceleration and gyroscope data of the left wrist were recorded at 100 Hz using the smartwatch's IMU. The maximum, minimum and Root Mean Square (RMS) were calculated using 5-second sliding windows to extract. Frequency stability (FS) metric derived from standard deviation (SD) of arm swing frequency peaks. Finally, we obtained 29 temporal raw features, as shown in Table 3. In addition, the running duration of each trial was recorded at 1 Hz as a supplementary temporal feature. In summary, a total of 13,027 data samples from 25 participants were retained. These were divided into a training set and a test set at an approximate 5:1 ratio based on randomized participant selection. The training set comprises 10,829 s of data from 21 participants (participant 1, 2, 3, 4, 6, 8, 10, 11, 13, 14, 15, 16, 17, 18, 20,21, 22, 23, 24, and 25), while the test set includes 2,198 s of data from the remaining 4 participants (Participants 5, 9, 12, and 19). This partitioning strategy was designed to rigorously evaluate model generalizability across unseen subjects while maintaining population representativeness.

For LSTM network implementation, the dataset was structured as sequential temporal segments using continue sliding window approach. As illustrated in Fig. 3(a), Raw time-series feature data was iterated into fixed-length windows, creating input features. Then, VO_2 value was used as output matched with corresponding input features, regarded as supervised learning problems, where n seconds represent the width of the sliding window, and w represents the number of input features. $W(n, w)$ is the model's input feature and $Y(VO_2)$ is the predicted output.

The selection of sliding window size influences the balance of

Table 3
Feature list.

	Feature	Name	Type	Source
1	R_1	Cadence	Kinematic	Stryd
2	R_2	Ground Contact Time	Kinematic	Stryd
3	R_3	Height	Anthropometric	Measurement
4	R_4	Weight	Anthropometric	Measurement
5	R_5	Heart Rate	Physiology	GT RUNNER
6	R_6	Acc_x_max	Kinematic	GT RUNNER
7	R_7	Acc_y_max	Kinematic	GT RUNNER
8	R_8	Acc_z_max	Kinematic	GT RUNNER
9	R_9	Gyro_x_max	Kinematic	GT RUNNER
10	R_{10}	Gyro_y_max	Kinematic	GT RUNNER
11	R_{11}	Gyro_z_max	Kinematic	GT RUNNER
12	R_{12}	Acc_x_min	Kinematic	GT RUNNER
13	R_{13}	Acc_y_min	Kinematic	GT RUNNER
14	R_{14}	Acc_z_min	Kinematic	GT RUNNER
15	R_{15}	Gyro_x_min	Kinematic	GT RUNNER
16	R_{16}	Gyro_y_min	Kinematic	GT RUNNER
17	R_{17}	Gyro_z_min	Kinematic	GT RUNNER
18	R_{18}	Acc_x_rms	Kinematic	GT RUNNER
19	R_{19}	Acc_y_rms	Kinematic	GT RUNNER
20	R_{20}	Acc_z_rms	Kinematic	GT RUNNER
21	R_{21}	Gyro_x_rms	Kinematic	GT RUNNER
22	R_{22}	Gyro_y_rms	Kinematic	GT RUNNER
23	R_{23}	Gyro_z_rms	Kinematic	GT RUNNER
24	R_{24}	Speed	Kinematic	Stryd
25	R_{25}	Frequency Stability	Kinematic	Stryd
26	R_{26}	Time	Kinematic	Stryd
27	R_{27}	Vertical Oscillation	Kinematic	Stryd
28	R_{28}	Age	Anthropometric	Measurement
29	R_{29}	Gender	Anthropometric	Measurement

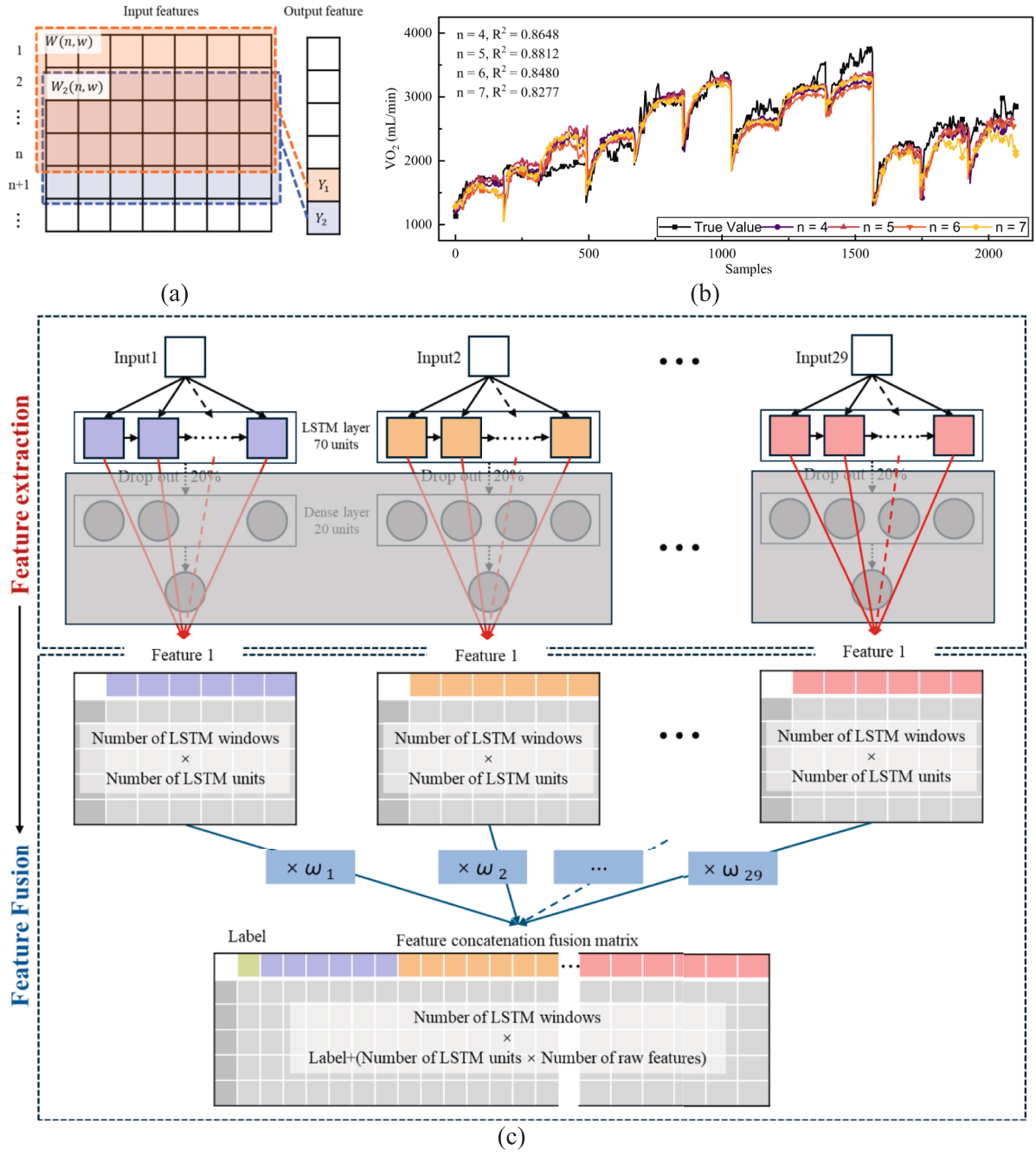

Fig. 3. Parameter setting process: (a) development of input and output features; (b) prediction performance with different sliding window size; (c) the process of extracting and fusing features.

capturing short-term patterns and longer-term trend regularities as well as the computational efficiency and generalization ability of the model. To obtain stable and dynamic VO_2 prediction, the sliding windows are fixed and below 10 s. After the parameter adjustment analysis, the sliding window size was set to 5 s with a stride of 1 s, as illustrated in Fig. 3(b). This resulted in a total of 10,457 windows in the training set and 2,126 windows in the test set.

3.2. Parameter setting

Model development was conducted using the TensorFlow framework and executed on an Intel® Core™ i7-14700F CPU @ 2.10 GHz (20 cores) with 32 GB RAM. The same LSTM network architecture was employed for both feature extraction and VO_2 prediction. Specifically, the LSTM layer comprised 70 units, followed by a fully connected layer containing 10 neurons. Following each layer, a dropout layer with 0.2 dropout rate was set to lessen the chance of overfitting [17]. The Adam optimizer was utilized in conjunction with Mean Squared Error (MSE) loss function, a batch size of 128, and 10 training epochs.

For feature extraction, raw data from wearable devices and anthropometric characteristic were processed by constructing 29 isolated LSTM networks. Each network received a distinct input S_j obtained by sliding windows from R_j. $Y(VO_2)$ was designated as the output for each network. After training, the output from the first layer of each feature extraction LSTM network was utilized as the new input feature L_j for the feature fusion LSTM model, without applying weights. After weighting $L_j \times X_{i,j}$, the initial weighted feature L'_j was developed and then were

concatenated to construct the feature concatenation fusion matrix W as the input of LSTM VO_2 prediction network. Subsequently, the PSO was carried out in this process to optimize the weight matrix $X_{i,j}$ by minimizing the MSE as the loss function. The PSO parameters were configured with a swarm size of 10, a maximum of 20 iterations, a fixed inertia weight of 0.85, both individual and social learning factors set to 1.5. The termination criterion for the PSO algorithm is determined solely by the maximum number of iterations. The upper and lower bounds for all 29 features were set to [0,1]. The above process of extracting and fusing features is described in Fig. 3(c). After termination, the weight matrix $X_{i,j}$ was fixed, considered as the best fitness weights for features, and the last LSTM VO_2 prediction network was the final network. Finally, forward pass was performed on the training and testing sets, respectively, to obtain the model's final predictions on both datasets.

3.3. Evaluation performance of model

Several essential evaluation metrics were utilized to assess the performance of the prediction model: the coefficient of determination (R^2), root mean squared error (RMSE), mean absolute error (MAE), Mean absolute percentage error (MAPE) and mean squared error (MSE). The R^2 metric quantifies the agreement between predicted values and true values, with values closer to 1 indicating superior predictive performance. MSE and MAE represent the average squared difference and the average absolute difference, respectively, between predicted and true values; lower values for both metrics correspond to higher model accuracy. RMSE, defined as the square root of MSE, offers enhanced interpretability due to its consistency with the original data scale. Additionally, Kling-Gupta efficiency (KGE) and Brier Score Function (BF) are calculated to evaluate the model accuracy. KGE was proposed by Harald Kling and Hoshin Vijai Gupta and has been widely used as a goodness-of-fit indicator in time-series prediction tasks [31]. It ranges $(-\infty, 1]$ and indicates high accuracy when close to 1. The BF reflects the accuracy of prediction and indicates better performance when close to zero. A comparison of the proposed method with the baseline LSTM network, XGBoost, and Linear Boosting in training set and test set are illustrated in Table 4.

The proposed model demonstrated the most robust generalization capability among all evaluated architectures. Our model achieved a consistent performance between the training ($R^2 = 0.909$) and test set ($R^2 = 0.913$). In contrast, the XGBoost model exhibited severe overfitting. While XGBoost attained near-perfect accuracy on the training set ($R^2 = 0.987$, MAPE = 2.048%), its performance deteriorated sharply on the testing set, with R^2 dropping to 0.650 and MAPE increasing to 11.575%, which suggests that tree-based models, without specific temporal adaptation, may fail to extrapolate the complex dynamics of VO_2 data effectively. Conversely, Linear Boosting showed signs of underfitting, characterized by consistently high errors (RMSE > 289) and low R^2 values across both datasets.

In terms of distributional consistency, the proposed model achieved the highest KGE on the test set (0.874), surpassing the LSTM (0.855). This indicates that the predicted VO2 trajectories not only minimize point-wise errors but also preserve the variance and correlation structure of the physiological ground truth. Additionally, the low BF (0.0019) confirms that the model is free from systematic overestimation or underestimation, ensuring reliability for practical physiological monitoring.

The time-series plots comparing different models with the actual VO_2 curves are shown in Fig. 4(a), demonstrating that our model better reflects the fluctuations in oxygen uptake across three running paces among different participants. The linear correlation between measured and predicted VO_2 values in the test set is presented in Fig. 4(b), indicating a strong positive linear relationship, with $R^2 = 0.921$. The Bland-Altman analysis, shown in Fig. 4(c), evaluates the agreement and error distribution between predicted and actual values. The predicted values exhibit a slight underestimation, with a mean difference of –22.70, and demonstrate good agreement and consistent variability, as evidenced by the concentrated and evenly distributed differences. Finally, 6-fold cross-validation was carried out to evaluate generalization ability. As shown in Table 5, the proposed model presents consistent performance in every fold.

In summary, these results demonstrate that the proposed architecture effectively captures the non-linear and temporal characteristics of VO_2 kinetics, providing a robust solution for accurate physiological estimation.

3.4. Optimal weights

The optimal weights derived from PSO optimization for VO_2 estimation are presented in Fig. 4(d). Acc_x_rms, Gyro_z_min, FS, and Time exhibit the most prominent feature contributions, each with a weight of nearly 1.000. In descending order of weight, other highly predictive features include HR, Acc_x_min, Gender, Gyro_x_min, and Height, all with weights greater than 0.8. Additionally, Acc_y_max, Gyro_y_max, Gyro_z_max, Gyro_x_rms, Gyro_y_rms, and Gyro_z_rms display minimal feature selection effect, with weights below 0.010.

Anthropometric parameters, including age, gender, weight, and height, have been widely used in VO_{2max} estimation [32], as they help constrain the fluctuation range of oxygen uptake. The work of Akay et al. [33] demonstrated that neural networks can capture the correlation between VO_{2max} and anthropometric parameters. The distribution of the optimized weights assigned to anthropometric parameters in our study confirms that our model captures this non-linear relationship and quantifies its effects. Similarly, Wang et al. [3,34] reported that Heart Rate (HR) played a critical role in predicting VO_2 as running intensity increased in their XGBoost model. This aligns with our findings, where a high weight was assigned to HR (0.985).

Our results also indicate that gait parameters such as FS (0.995), GCT (0.764), Cadence (0.473), and Speed (0.174) contribute significantly to predicting dynamic VO_2. This is likely because significant energy is consumed by the leading leg during ground contact and gait transition [35]. The correlation between kinematic foot parameters and VO_2 has also been modeled using an LSTM network with shoe-mounted IMUs by Vidyarani. While there is limited research on estimating real-time oxygen uptake, related studies have shown that physical activity intensity can be measured using wrist-worn IMUs [36]. Consequently, our results offer guidance for feature selection in model development. For instance, x-axis parameters warrant particular attention due to their consistently

Table 4
Accuracy metrics of each model.

Model	Dataset	MAPE	R^2	MSE	RMSE	MAE	KGE	BF
Our Model	Training	6.591	0.909	28345.816	168.362	128.596	0.901	0.0018
	Test	7.250	0.913	30420.113	174.414	139.703	0.874	0.0019
LSTM	Training	5.73	0.907	28929.913	170.088	134.547	0.951	0.0017
	Test	7.40	0.881	41462.733	203.624	205.103	0.855	0.0026
XGBoost	Training	2.048	0.987	4286.247	65.469	46.825	0.979	0.0003
	Test	11.575	0.650	124355.552	352.641	269.958	0.771	0.0077
Linear	Training	9.951	0.748	84072.163	289.952	208.592	0.787	0.0052
Boosting	Test	11.491	0.671	116685.672	341.593	271.342	0.744	0.0072

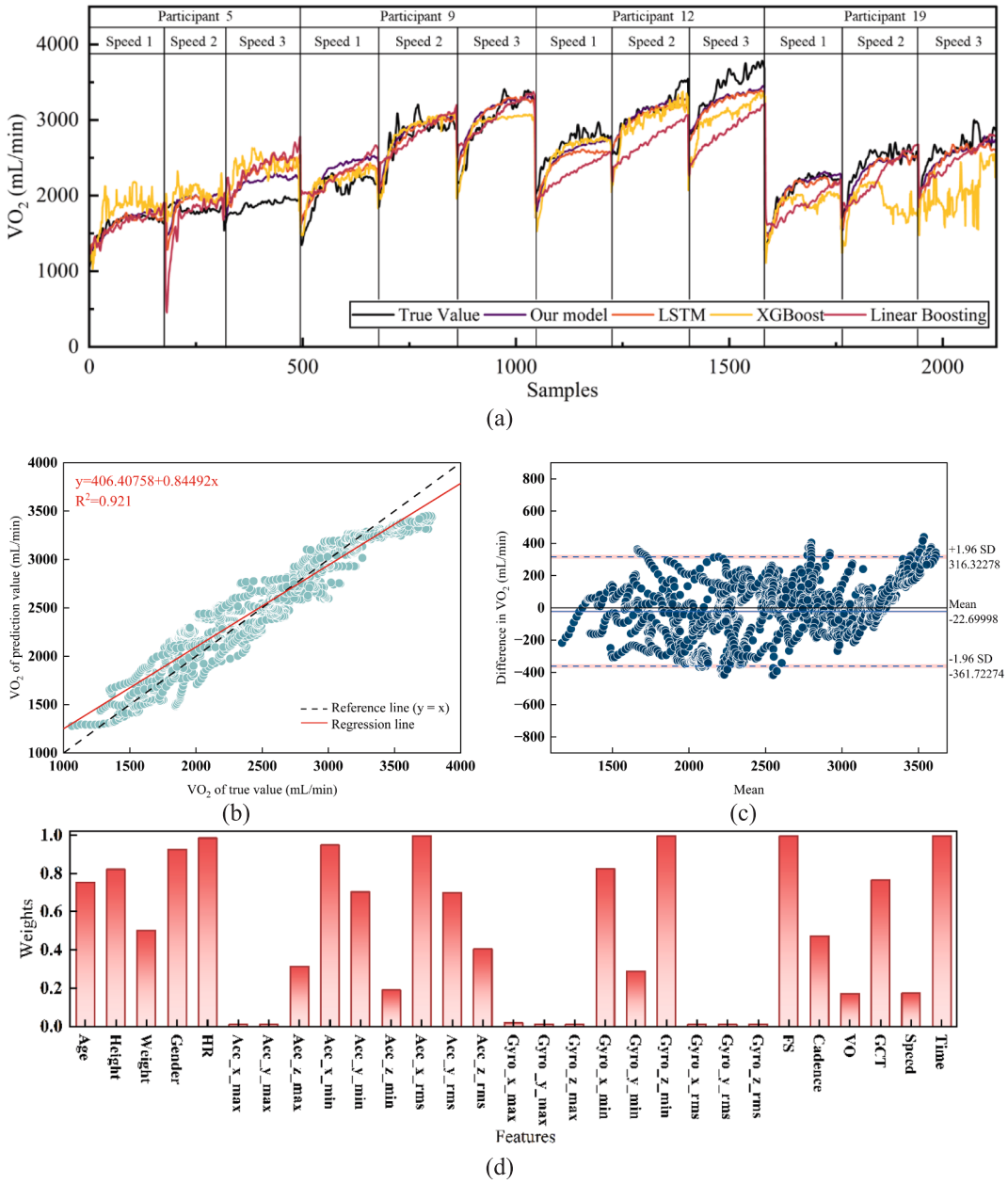

Fig. 4. Prediction performance of our model: (a) comparison between the proposed model and other models; (b) Scatter-regression plot of predicted vs. true VO_2 on the test set; (c) Bland-Altman plot predicted vs. true VO_2 on the test set (d) optimized weights.

Table 5
6-fold cross validation.

Model	Dataset	MAPE	R^2	RMSE	MAE
Our Model	Training	6.591	0.909	168.362	128.596
	Test	7.250	0.913	174.414	139.703
CV 2	Training	6.735	0.907	172.335	132.689
	Test	7.088	0.910	170.512	135.712
CV 3	Training	7.122	0.913	170.816	146.265
	Test	6.688	0.901	185.532	150.298
CV 4	Training	4.160	0.940	136.299	102.904
	Test	6.224	0.904	180.190	140.921
CV 5	Training	5.812	0.931	160.065	140.624
	Test	7.036	0.897	189.085	170.327
CV 6	Training	5.732	0.951	170.088	134.847
	Test	6.975	0.893	198.283	163.309

high weights, such as Acc_x_min (0.947), Acc_x_rms (0.995), and Gyro_x_min (0.824).

4. Discussion

4.1. Model performance in dynamics characterization

Oxygen uptake kinetics contain complexity and critical information, reflecting the integrated performance of pulmonary, cardiovascular, and muscle-control mechanisms during exercise [37]. To evaluate the dynamic prediction performance of the proposed our model, a comparison between the true values and the predicted values, calculated as the mean value at each time point, was conducted and is presented in Fig. 5(a). The mean response curves demonstrate substantial consistency, and confidence intervals show uniform width, and the confidence intervals exhibit uniform width, indicating both the accuracy and stability of the time-sequence VO_2 predictions. As shown in Fig. 5(b), VO_2 displays a transient phase during each test, rather than rising immediately to a

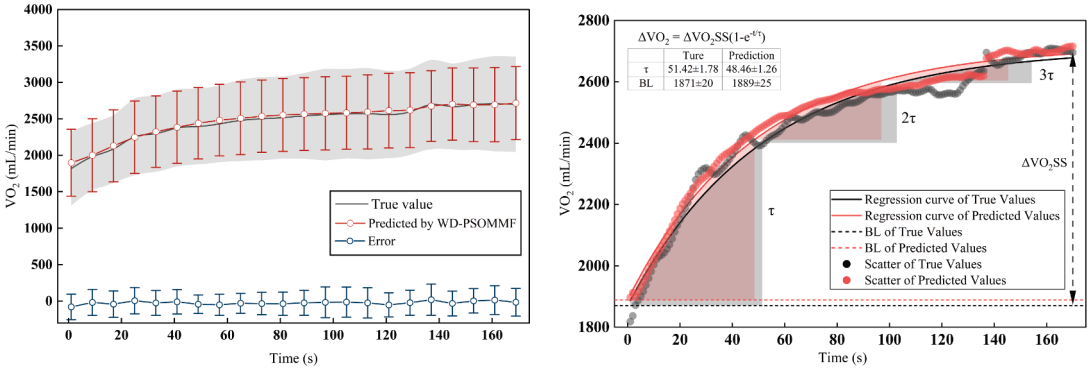

Fig. 5. VO2 kinetics response: (a) Mean response and SD of predicted, true VO_2, and error (b) VO_2 kinetics curve regression.

steady state, which is attributable to the limited nonoxidative muscle energy stores [38]. For moderate exercise below the lactate threshold, the exponential process can be described by Eq. (7) [39]

$$\Delta VO_2(t) = \Delta VO_2ss(1 - e^{-\frac{t}{\tau}}) \tag{7}$$

where, $\Delta VO_2(t)$ and ΔVO_2ss represents the increments of VO_2 at time t and a steady state VO_2, respectively, relative to the baseline. The parameter τ is a time constant that quantifies the rate of increase in VO_2, typically ranging from 10 to 100 s and is independent of exercise intensity. For each multiple of τ, VO_2 increases by 63% of the difference between its value at the previous τ and the required steady state. As depicted in Fig. 5(b), the τ calculated from the predicted values is 48.46, which is close to 51.42 s obtained from the true values and approximates the result (τ = 40) reported in reference [39] for moderate-intensity cycle ergometer exercise. For true values and predicted values, VO_2 reaches 2,400.7/2,411.6 mL/min after τ, 2,595.6/2,603.8 mL/min after 2τ, and 2,667.3/2,674.6 mL/min after 2τ, showing the robust performance of our model in predicting oxygen uptake dynamics characteristics. It is noteworthy that fitted baselines are 1871 mL/min for the true values and 1889 mL/min for the predictions, both higher than the typical resting value of approximately 250–1000 mL/min. This discrepancy is attributed to the fact that data collection commenced when the treadmill reached a stable speed, rather than from a resting state. For constant-work-rate moderate intensity exercise, VO_2 of participants begins to increase within the first breath [39,40], when the speed of treadmill is increasing. After spending about 30 s on reaching the target speed, VO_2 of participants was still in the exponential increase

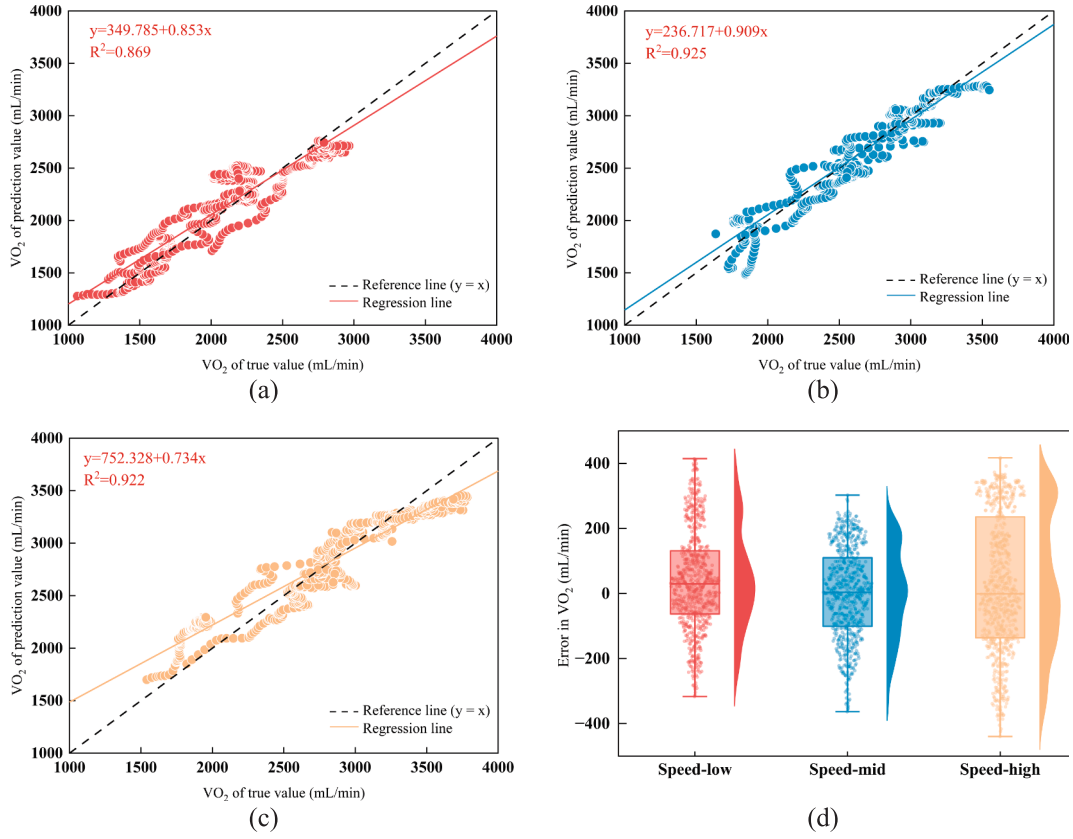

Fig. 6. Model performance at different speeds on the test set: (a) Scatter-regression plot at speed1; (b) scatter-regression plot at speed2; (c) scatter-regression plot at speed3 (d) the box and violin plot of errors at different speeds.

stage until to steady state after few minutes. Therefore, when the COSMED K5 started to record the breath data at the target speed, the participant was in running state with higher VO2 instead of rest, leading to the difference of baseline.

4.2. Model performance at different speeds

Our proposed model demonstrated strong agreement in predicting VO_2 across all three speed groups. It is consistent with the physiological characterization of breath system because the mode of VO_2 response at the range of constant moderate intensity exercise is similar could be divided into the exponential increase stage and the relative stable stage [39]. The proposed model, especially LSTM networks, is good at capturing this potential time-series characterization from multimodal features, leading to the robust cross-pace performance. As shown in Fig. 6(a)–(c), linear regression analysis between predicted and true values yielded slopes of 0.853, 0.909, and 0.734 for easy running, marathon-pace running, and lactate-threshold running, respectively, indicating slight compression effects. A consistent deviation pattern was observed across all three speeds, with interceptions of 349.789, 236.717, and 752.328 mL/min, respectively, suggesting a systematic overestimation. Comparable R^2 values of 0.869, 0.925, and 0.922 were obtained for the three speeds, demonstrating consistent goodness-of-fit. As depicted in Fig. 6, boxplot analysis of error values revealed similar error distribution patterns across speed groups, with comparable medians (29.7, 3.84, and −1.02 mL/min), however the third quartile (Q3) of lactate-threshold running (235.3 mL/min) was wider than the other speeds (131.5 and 110.1 mL/min), indicating the overestimation tendency showed slight amplifications.

In summary, the proposed model exhibited consistent and stable VO_2 prediction performance across different moderate running paces, while performance was relatively weaker at lactate-threshold running speeds. This phenomenon may be attributed to the slow component of oxygen uptake [41], which emerges during exercise above the lactate threshold and continues to rise for several minutes until a delayed but higher steady state is reached [42]. Although the running dataset was limited to moderate intensity, running speeds were calculated based on each participant's historical records, and the blood lactate curve may have been influenced by individual training or glycogen stores, resulting in rightward or leftward shifts [43], These factors could contribute to abnormalities in the oxygen uptake fluctuation pattern at certain moments or paces, thereby affecting the temporal feature relationships within the overall dataset.

4.3. Efficiency analysis

According to the difference in devices and feature selections, it is hard to compare the performance of our model with others. In order to discuss the balance between accuracy and efficiency of the model, we developed a popular CNN-LSTM model and configured using similar research methods [13,21]. 29 CNN networks (filters = 32) were used to extract features from raw features and stream to a two-layer LSTM networks (units of layer 1 = 128, units of layer 2 = 20) to estimate the oxygen uptake. Besides, after pruning the feature with weight <0.01, a pruned model was developed to present the final performance of our model. The results of comparison in the test set are shown in Table 6.

Table 6
Efficiency analysis of different models.

Model	Input Features	Parameters	Latency (ms)	RMSE	R2
Proposed model	29	595,581	47.227	174.414	0.913
Pruned model	22	458,381	46.536	170.325	0.915
CNN-LSTM	29	16,798,819	189.663	210.400	0.878

The pruned model demonstrated superior performance across all evaluated metrics, achieving the highest prediction accuracy (R^2 = 0.915, RMSE = 170.325) and the lowest inference latency (46.54 ms). Compared to the raw model, the pruning strategy reduced the input feature dimension by approximately 24% (from 2030 to 1540) and yielded a slight improvement in goodness-of-fit. It suggests that the excluded features likely constituted noise or redundant information that hindered the baseline model's convergence. In contrast, the CNN-LSTM exhibited suboptimal performance. Despite possessing a massive parameter count of 16.8 million, approximately 36 times larger than the proposed Pruned model, its predictive capability collapsed, yielding an R^2 of only 0.878, which cannot reach the R^2 = 0.9460 in their dataset [13]. For one hand, the accuracy of their model may depend on the features of rest VO_2 and rest HR, which do not include in our model considering the inconvenience in daily configuration. On the other hand, it can be attributed to severe over-parameterization. Given the finite size of the physiological dataset, such a massive model is highly prone to overfitting, where the network memorizes noise training rather than learning generalizable physiological dynamics. Furthermore, the high computational complexity of the parallel convolution architecture resulted in a four-fold increase in inference latency (189.66 ms). In comparison, our Pruned model not only reduced latency by 75.4% compared to the CNN-LSTM but also achieved the highest accuracy.

5. Conclusion

The primary contribution of this study is the development of our model, which integrates feature weight optimization with an LSTM network to predict dynamic oxygen uptake during running. By leveraging the proposed architecture, the model achieves a balance between accuracy and computational cost, as evidenced by high R^2 and KGE, as well as low RMSE, MAPE, MSE, MAE, and BF values. This enables runners to monitor real-time energy expenditure during exercise using existing wearable consumer devices. Performance analysis demonstrates that the model accurately predicts the dynamic characteristics of oxygen uptake and performs robustly across different speed groups within moderate intensity. The optimized feature weights indicate that the proposed weighting process enhances the predictive accuracy of the model. Specifically, Acc_x_rms, Gyro_z_min, FS, duration time, HR, Acc_x_min, Gender, Gyro_x_min, and Height exhibit dominant feature contributions, while Acc_y_max, Gyro_y_max, Gyro_z_max, Gyro_x_rms, Gyro_y_rms, and Gyro_z_rms can be excluded, thereby further reducing model complexity and computational cost.

6. Limitation

While the proposed model demonstrates superior accuracy and computational efficiency for VO_2 estimation, several limitations inherent to the current study design must be acknowledged to guide future research.

First, the development of the model was achieved using data primarily from healthy, recreational runners. Consequently, the generalization of these findings to elite athletes or clinical populations remains limited, who often exhibit distinct physiological characters and VO_2 kinetics. Future studies should expand the dataset to include these cohorts to validate and recalibrate the model for broader demographic applicability.

Second, the data collection was conducted under control laboratory conditions using treadmills. While this ensures data consistency, it introduces a discrepancy between treadmill and outdoor running. Outdoor environments involve external factors such as air resistance, uneven terrain, and ambient temperature fluctuations, which can alter running biomechanics and energy expenditure. Future work will focus on validating the model using wearable sensors in outdoor environments to ensure robustness against real-world variability.

Third, the current model operates as a static inference engine trained

on a fixed dataset. However, deploying a real-time calculation model presents a complex engineering challenge [44,45]. This requires a comprehensive performance evaluation across diverse scenarios, including extreme conditions, to balance accuracy, robustness, efficiency, and cost. In the future, we will transform and enhance the existing model to achieve robust real-world deployment.

CRediT authorship contribution statement

Zhengsheng Hu: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Gongcheng Xu:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Yan Wang:** Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Guanwen Qu:** Software, Methodology, Formal analysis, Data curation. **Puxiang Lai:** Supervision, Resources, Methodology, Conceptualization. **Wenzhao Li:** Resources, Conceptualization. **Jiahui Yin:** Formal analysis, Data curation. **Hui Xie:** Formal analysis, Data curation, Conceptualization. **Huiwen Xiao:** Data curation. **Tao Li:** Software, Methodology. **Ming Zhang:** Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Qitao Tan:** Writing – original draft, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

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.

Acknowledgements

This work was supported by the Research Institute for Sports Science and Technology at The Hong Kong Polytechnic University [grant number P0051374]; National Natural Science Foundation of China [grant number 82330061]; and the Shenzhen Research Fund [grant numbers JCYJ20230807140414029].

Data availability

Data will be made available on request.

References

[1] C.E. Garber, B. Blissmer, M.R. Deschenes, et al., Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise, Med. Sci. Sports Exerc. 43 (7) (2001) 1334–1359, https://doi.org/10.7916/D8CR5T2R.
[2] J.B. Whipp, A.S. Ward, Gas exchange dynamics and the tolerance to muscular exercise: Effects of fitness and training, Ann. Physiol. Anthropol. 11 (3) (1992) 207–214, https://doi.org/10.2114/ahs1983.11.207.
[3] A.S. Divakaruni, M. Jastroch, A practical guide for the analysis, standardization and interpretation of oxygen consumption measurements, Nat. Metab. 4 (2022) 978–994, https://doi.org/10.1038/s42255-022-00619-4.
[4] S. Chen, E. Wohlers, E. Ruud, et al., Improving temporal accuracy of human metabolic chambers for dynamic metabolic studies, PLoS One 13 (4) (2018) e0193467, https://doi.org/10.1371/journal.pone.0193467.
[5] K. Winkert, J. Kirsten, R. Kamnig, et al., Differences in VO2max measurements between breath-by-breath and mixing-chamber mode in the COSMED K5, Int. J. Sports Physiol. Perform. 16 (9) (2021) 1335–1340, https://doi.org/10.1123/ijspp.2020-0634.
[6] A. Toulouse, D. Joubert, G. Oden, et al., Comparison of the VO2 Master Pro and Cosmed K5 during walking, jogging, and running, J. Sci. Sport Exer. 4 (2) (2022) 119–127, https://doi.org/10.1007/s42978-021-00146-w.
[7] F. Gao, Y. Chan-Yu, Z. Li, et al., Comparison of two portable metabolic systems for measuring energy expenditure at rest and during exercise in untrained women, Front. Physiol. 16 (2025) 1583703, https://doi.org/10.3389/fphys.2025.1583703.
[8] M. Altini, J. Penders, O. Amft, Estimating oxygen uptake during nonsteady-state activities and transitions using wearable sensors, IEEE J. Biomed. Health Inform. 20 (2) (2015) 469–475, https://doi.org/10.1109/JBHI.2015.2390493.
[9] M.M.H. Shandhi, W.H. Bartlett, J.A. Heller, et al., Estimation of instantaneous oxygen uptake during exercise and daily activities using a wearable cardio-electromechanical and environmental sensor, IEEE J. Biomed. Health Inform. 25 (3) (2020) 634–646, https://doi.org/10.1109/JBHI.2020.3009903.
[10] J. Lounana, F. Campion, T.D. Noakes, et al., Relationship between% HRmax,% HR reserve,% VO2max, and% VO2 reserve in elite cyclists, Med. Sci. Sports Exerc. 39 (2) (2007) 350–357, https://doi.org/10.1249/01.mss.0000246996.63976.5f.
[11] M. O'Reilly, B. Caulfield, T. Ward, et al., Wearable inertial sensor systems for lower limb exercise detection and evaluation: a systematic review, Sports Med. 48 (5) (2018) 1221–1246, https://doi.org/10.1007/s40279-018-0878-4.
[12] T. Li, X. Lou, Z. Yang, et al., Clarifying the impact of engine operating parameters of heavy-duty diesel vehicles on NOx and CO2 emissions using multimodal fusion methods, Sci. Total Environ. 954 (2024) 176598, https://doi.org/10.1016/j.scitotenv.2024.176598.
[13] L. Jiang, C. Wang, X. Shang, et al., Two-step event-triggered data driven model predictive control for trajectory tracking of unmanned surface vessel under environmental disturbances, IEEE Trans. Autom. Sci. Eng. 22 (2025) 16801–16813, https://doi.org/10.1109/TASE.2025.3579396.
[14] C.T. Hsiao, C. Tong, G.L. Coté, Machine Learning-based VO2 estimation using a wearable multiwavelength photoplethysmography device, Biosensors 15 (4) (2025) 208, https://doi.org/10.3390/bios15040208.
[15] D. Hong, S. Sun, Machine learning regressors to estimate continuous oxygen uptakes (VO2), Appl. Sci. 14 (17) (2024) 7888, https://doi.org/10.3390/app14177888.
[16] H.C. Kilinc, S. Apak, F. Ozkan, et al., Multimodal fusion of optimized GRU–LSTM with self-attention layer for hydrological time series forecasting, Water Resour. Manag. 38 (15) (2024) 6045–6062, https://doi.org/10.1007/s11269-024-03943-4.
[17] T. Fu, Ensemble learning of attention-based BiLSTM networks for ADHD detection from EEG signals code, Comput. Methods Biomech. Biomed. Eng. (2025) 1–13, https://doi.org/10.1080/10255842.2025.2576625.
[18] Y. Gong, J. Zhang, Y. Lv, et al., Cross-LUSleepNet: a U-shaped sleep staging method based on cross-layer connection modules and Bi-LSTM, IAENG Int. J. Comput. Sci. 52 (11) (2025).
[19] C. Wang, H. Liao, K. Xie, et al., Research on control methods for gas-liquid separators based on UKF-LSTM hybrid observation and sliding mode control, J. Process Control 155 (2025) 103573, https://doi.org/10.1016/j.jprocont.2025.103573.
[20] Z. Wang, Y. Song, L. Pang, et al., Attention-enhanced CNN-LSTM model for exercise oxygen consumption prediction with multi-source temporal features, Sensors 25 (13) (2025) 4062, https://doi.org/10.3390/s25134062.
[21] K.R. Vidyarani, V. Talasila, R. Umar, et al., Inertial sensor-based heel strike and energy expenditure prediction using a hybrid machine learning approach, 20552076251333375, Digital Health 11 (2025), https://doi.org/10.1177/20552076251333375.
[22] P. Davidson, H. Trinh, S. Vekki, et al., Surrogate modelling for oxygen uptake prediction using LSTM neural network, Sensors 23 (4) (2023) 2249, https://doi.org/10.3390/s23042249.
[23] S.S. Bangaru, C. Wang, F. Aghazadeh, et al., Oxygen uptake prediction for timely construction worker fatigue monitoring through wearable sensing data fusion, Sensors 25 (10) (2025) 3204, https://doi.org/10.3390/s25103204.
[24] H.C. Kilinc, S. Apak, M.E. Ergin, et al., Enhancing hydrological time series forecasting with a hybrid Bayesian-ConvLSTM model optimized by particle swarm optimization, Acta Geophys. (2025) 1–18, https://doi.org/10.1007/s11600-025-01570-0.
[25] C.R. van Rassel, O.O. Ajayi, K.M. Sales, et al., Is running power a useful metric? Quantifying training intensity and aerobic fitness using stryd running power near the maximal lactate steady state, Sensors 23 (21) (2023) 8729, https://doi.org/10.3390/s23218729.
[26] D. Xu, H. Zhou, W. Quan, et al., Accurately and effectively predict the ACL force: Utilizing biomechanical landing pattern before and after-fatigue, Comput. Methods Programs Biomed. 241 (2023) 107761, https://doi.org/10.1016/j.cmpb.2023.107761.
[27] U.M. Butt, S. Letchmunan, M. Ali, et al., Machine learning based diabetes classification and prediction for healthcare applications, J. Healthcare Eng. 1 (2021) 9930985, https://doi.org/10.1155/2021/9930985.
[28] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.
[29] R. Eberhart, J. Kennedy, A new optimizer using particle swarm theory, in: MHS'95. Proceedings of the sixth international symposium on Micro machine and human science. Presented at the MHS'95, in: Proceedings of the Sixth International Symposium on Micro Machine and Human Science, 1995, pp. 39–43, https://doi.org/10.1109/MHS.1995.494215.
[30] Y. Peng, T. Li, C. Bao, et al., Performance analysis and multi-objective optimization of bionic dendritic furcal energy-absorbing structures for trains, Int. J. Mech. Sci. 246 (2023) 108145, https://doi.org/10.1016/j.ijmecsci.2023.108145.
[31] H.C. Kilinc, S. Apak, F. Ozkan, et al., AH-GCAN-LSTM: adaptive hybrid-graph convolution attention network with LSTM and genetic optimization for hydrological time series forecasting, Water Resour. Manag. (2025) 1–20, https://doi.org/10.1007/s11269-025-04343-y.
[32] D.I. Bradshaw, J.D. George, A. Hyde, et al., An accurate VO2max nonexercise regression model for 18–65-year-old adults, Res. Q. Exerc. Sport 76 (4) (2005) 426–432, https://doi.org/10.1080/02701367.2005.10599315.
[33] M.F. Akay, C. Inan, D.I. Bradshaw, et al., Support vector regression and multilayer feed forward neural networks for non-exercise prediction of VO2max, Expert Syst. Appl. 36 (6) (2009) 10112–10119, https://doi.org/10.1016/j.eswa.2009.01.009.

[34] Z. Wang, Q. Zhang, K. Lan, et al., Enhancing instantaneous oxygen uptake estimation by non-linear model using cardio-pulmonary physiological and motion signals, Front. Physiol. 13 (2022) 897412, https://doi.org/10.3389/fphys.2022.897412.
[35] K.R. Vidyarani, V. Talasila, N. Megharjun, et al., An inertial sensing mechanism for measuring gait parameters and gait energy expenditure, Biomed. Signal Process. Control 70 (2021) 103056, https://doi.org/10.1016/j.bspc.2021.103056.
[36] F. Marin, K. Lepetit, L. Fradet, et al., Using accelerations of single inertial measurement units to determine the intensity level of light-moderate-vigorous physical activities: technical and mathematical considerations, J. Biomech. 107 (2020) 109834, https://doi.org/10.1016/j.jbiomech.2020.109834.
[37] M. Buchheit, P.B. Laursen, High-intensity interval training, solutions to the programming puzzle: Part I: cardiopulmonary emphasis, Sports Med. 43 (5) (2013) 313–338, https://doi.org/10.1007/s40279-013-0029-x.
[38] A.V. Hill, C.N.H. Long, H. Lupton, Muscular exercise, lactic acid, and the supply and utilisation of oxygen, Proceed. R. Soc. Lond. Series B, Containing Papers Biol. Character 96 (679) (1924) 438–475, https://doi.org/10.1098/rspb.1924.0037.
[39] D.C. Poole, A.M. Jones, Oxygen uptake kinetics, Compr. Physiol. 2 (2) (2012) 933–996, https://doi.org/10.1002/j.2040-4603.2012.tb00419.x.
[40] D.W. Hill, D.C. Poole, J.C. Smith, The relationship between power and the time to achieve VO2max, Med. Sci. Sports Exerc. 34 (4) (2002) 709–714.
[41] A.M. Jones, H. Carter, J.H. Doust, A disproportionate increase in VO2 coincident with lactate threshold during treadmill exercise, Med. Sci. Sports Exerc. 31 (9) (1999) 1299–1306, https://doi.org/10.1097/00005768-199909000-00011.
[42] T.J. Barstow, P.A. Molé, Linear and nonlinear characteristics of oxygen uptake kinetics during heavy exercise, J. Appl. Physiol. 71 (6) (1991) 2099–2106, https://doi.org/10.1152/jappl.1991.71.6.2099.
[43] O. Faude, W. Kindermann, T. Meyer, Lactate threshold concepts: how valid are they? Sports Med. 39 (6) (2009) 469–490, https://doi.org/10.2165/00007256-200939060-00003.
[44] A. Guan, S. Zhou, W. Gu, et al., Dynamic simulation and parameter calibration-based experimental digital twin platform for heat-electric coupled system, IEEE Trans. Sustainable Energy (2025), https://doi.org/10.1109/TSTE.2025.3609042.
[45] P. Liu, S. Li, H. Jin, et al., Shape parameterization method and hydrodynamic noise characteristics of low-noise toroidal propeller, Ocean Eng. 328 (2025) 121088, https://doi.org/10.1016/j.oceaneng.2025.121088.