

RESEARCH ARTICLE

Prediction of energy expenditure and pacing optimization in long-distance running using multisource physiological feature fusion and machine learning

Zhiguang Wang^{1,*}, Yanli Bu¹, Jie Chu¹, Lan Yang¹, Yuchen Zhang²

¹College of Physical Education, Henan Agricultural University, Zhengzhou, Henan, China. ²College of Physical Education, Henan University of Technology, Zhengzhou, Henan, China.

Received: May 6, 2026; accepted: August 29, 2026.

Accurate estimation of energy expenditure (EE) and effective pacing guidance are essential for optimizing long-distance running performance and minimizing premature fatigue. However, methods based on single-modality wearable data or linear equations often provide limited accuracy during prolonged exercise. This study developed and validated a multisource feature fusion network (MSFF-Net) for minute-level EE prediction during long-distance running, compared its performance with that of conventional machine-learning models, and constructed a dynamic-programming framework for pacing optimization to mitigate late-race deceleration and perceived exertion. A retrospective analysis was conducted using training and competition data collected from 186 long-distance runners with 127 males and 59 females, aged from 18 – 60 years old between 2020 and 2024. Features were derived from cardiovascular signals including heart rate and heart-rate variability (HRV), tri-axial accelerometry, gait kinematics, and demographic and anthropometric variables. Five predictive approaches including linear regression (LR), support vector regression (SVR), random forest (RF), extreme gradient boosting (XGBoost), and the proposed MSFF-Net were evaluated using 10-fold cross-validation. A dynamic-programming pacing optimizer was then constructed using the best-performing predictor. MSFF-Net achieved the lowest prediction errors with mean absolute error (MAE) of 0.61 ± 0.08 kcal/min, root mean squared error (RMSE) of 0.82 ± 0.09 kcal/min, and the highest explained variance with R^2 of 0.94 ± 0.01 , outperforming LR ($R^2 = 0.71$) and XGBoost ($R^2 = 0.90$). SHapley Additive exPlanations (SHAP) identified heart rate, running speed, and HRV represented by the root mean square of successive differences (RMSSD) as the three most influential features. In simulation, model-guided pacing was projected to reduce finishing time by 172.6 ± 39.3 s and end-of-race ratings of perceived exertion by 1.79 ± 1.02 points. External validation yielded a Bland-Altman bias of -0.06 kcal/min and limits of agreement from -1.24 to $+1.13$ kcal/min. Multisource physiological feature fusion combined with machine learning enabled accurate and interpretable EE prediction and might support individualized pacing prescriptions for long-distance runners.

Keywords: long-distance running; energy expenditure; pacing optimization; machine learning; multisource fusion; wearable sensors; SHAP interpretability.

*Corresponding author: Zhiguang Wang, College of Physical Education, Henan Agricultural University, Zhengzhou, Henan 450046, China. Email: guangti2007@163.com.

Introduction

Long-distance running including half-marathons, marathons, and ultramarathons is among the

most widely practiced endurance sports worldwide, attracting millions of recreational and competitive participants each year. Performance over these distances depends not only on

maximal oxygen consumption (VO_2max) but also on the interaction among bioenergetic cost, thermoregulation, substrate availability, and pacing discipline during prolonged exercise [1, 2]. High-resolution quantification of energy expenditure (EE) and translation of these estimates into actionable pacing guidance may therefore improve performance, inform training prescription, and reduce injury risk. Traditionally, EE during running has been estimated using compendium-based metabolic equivalents or speed-based linear equations derived from controlled treadmill protocols [3]. These approaches assume steady-state physiology, overlook environmental conditions and interindividual variability, and may underestimate the progressive metabolic drift associated with cardiovascular drift, dehydration, and glycogen depletion late in a race [4]. Although laboratory-based indirect calorimetry remains the reference standard, its use is impractical outside a metabolic-cart setting. Commercial wearable devices offer a field-based alternative. However, independent validation studies of popular smartwatches have reported mean errors of 8 – 15% for VO_2max and lactate-threshold estimates with limits of agreement sufficiently wide to affect individual training decisions [5, 6].

Machine learning (ML) has emerged as a promising means of addressing these limitations. Activity-specific ML models trained on accelerometer and heart rate (HR) data outperform compendium-based methods for walking, cycling, and mixed daily activities [7, 8]. In running, tree-based ensembles and shallow neural networks trained against indirect calorimetry have achieved R^2 values of approximately 0.85 – 0.90 for minute-level EE prediction [9]. Recent multimodal studies integrating electrocardiography (ECG), accelerometry, and running kinematics further indicate that the relative contributions of individual features vary with exercise intensity. Accelerometer-derived features predominate at lower speeds, whereas ECG-derived features become increasingly informative above the

lactate threshold [10]. Pacing research has advanced in parallel. Large-scale analyses of marathon and half-marathon data consistently indicate that even or negative pacing is associated with better performance than positive pacing, and that more than half of recreational runners experience substantial deceleration — the so-called marathon “wall” — after 25 – 30 km [11, 12]. Models based on k-nearest neighbors and artificial neural networks can predict finishing time within approximately 2 – 6% using pre-race characteristics [13]. Subsequent studies have combined physiological monitoring with ML to detect impending fatigue before overt performance deterioration occurs [14, 15]. More recent personalization models integrate polarized or pyramidal training data to tailor pacing strategies to athlete profiles [16], while early-warning architectures based on variational autoencoders and Lyapunov-exponent dynamics have demonstrated proof of concept for anticipating glycogen-related performance collapse [17]. Nevertheless, important gaps remain. Most EE and pacing models rely on a single data modality such as HR or accelerometry and therefore underuse complementary information from HRV, gait kinematics, and environmental conditions. Moreover, many published models report only group-level performance and do not quantify the athlete heterogeneity that determines the reliability of individualized recommendations [18, 19]. Few studies have closed the translational loop from high-resolution EE prediction to an actionable, minute-by-minute pacing prescription evaluated against actual race performance within a single analytical framework.

This study developed and internally validated a multisource feature fusion network (MSFF-Net) for minute-level EE prediction during long-distance running, compared its performance with that of four conventional ML models, interpreted the model using SHapley Additive exPlanations (SHAP), and used the resulting predictor to construct and evaluate a pacing-optimization framework against athletes’ actual race performance. The proposed hybrid deep-

learning architecture combined one-dimensional convolutional layers, bidirectional long short-term memory networks, and multi-head self-attention to integrate cardiovascular, biomechanical, anthropometric, and environmental features. By linking high-resolution EE prediction with individualized pacing optimization, this work advanced multimodal physiological data integration in endurance-sport analytics and provided a methodological basis for real-time pacing guidance systems for coaches and athletes.

Materials and methods

Recruitment of participants

This study was conducted in accordance with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis plus Artificial Intelligence (TRIPOD+AI) guideline (<https://www.tripod-statement.org>) for prediction models incorporating machine-learning components [20]. A total of 186 adult long-distance runners including 127 males aged 20 – 58 years with the mean age of 39.1 ± 9.4 years and 59 females aged 18 – 56 years with the mean age of 36.8 ± 10.2 years from affiliated running clubs and the university sports medicine outpatient clinic in Zhengzhou, Henan, China between January 2020 and December 2024 were involved in this study. The inclusion criteria were age of 18 – 60 years, at least 12 months of structured endurance training before enrollment, at least three recorded running sessions of 10 km or longer with valid chest-strap electrocardiogram (ECG) and foot-mounted inertial measurement unit (IMU) data, and a reference-standard laboratory estimate of maximal oxygen consumption (VO_2max) and lactate threshold obtained within 6 months of the recorded sessions. The exclusion criteria were a cardiovascular, metabolic, or musculoskeletal condition that contraindicated vigorous exercise, incomplete or corrupted sensor recordings with more than 5% missing samples per session, self-reported acute illness or injury within 7 days before a recorded session,

or use of medication known to alter the heart-rate response such as β -blockers. The participants contributed 1,842 running sessions and 47 marathon or half-marathon races to the final analytical dataset. All procedures of this research were approved by the Institutional Review Board of Henan Agricultural University, Zhengzhou, Henan, China (Approval No. HAU-2020-038).

Demographic and anthropometric measurements

At enrollment, each runner underwent a standardized anthropometric assessment using a Seca 274 calibrated stadiometer (Seca GmbH, Hamburg, Germany) and a InBody 770 bioelectrical impedance analyzer (InBody Co., Ltd., Seoul, South Korea). Variables that were entered into the models included age, gender, height (cm), body mass (kg), body mass index (BMI) (kg/m^2), skeletal muscle mass (kg), body fat percentage, and years of structured running training. When available, self-reported personal-best times for the 10-km, half-marathon, and marathon distances were recorded as proxies for baseline endurance capacity [21].

Cardiovascular physiological monitoring

Continuous single-lead electrocardiogram (ECG) signals were recorded at 1,000 Hz using a Polar H10 chest-strap monitor (Polar Electro Oy, Kempele, Finland). R-wave peaks were identified using Pan-Tompkins algorithm obtained from PhysioNet ECGPUWAVE (<https://physionet.org/content/ecgpuwave/>). Ectopic beats were corrected using piecewise cubic-spline interpolation. From the cleaned RR-interval series, time-domain indices including mean RR interval, standard deviation of all normal-to-normal intervals (SDNN), root mean square of successive differences (RMSSD), and percentage of successive RR intervals differing by more than 50 ms (pNN50) and frequency-domain indices including low-frequency power (LF), high-frequency power (HF), and the LF/HF ratio were calculated over rolling 60-s windows in accordance with established recommendations for ambulatory athlete monitoring [22, 23].

Instantaneous heart rate (HR), HR reserve relative to the age-predicted maximum, and the HR-to-running-speed ratio used as a surrogate for cardiovascular drift were also derived. Peripheral oxygen saturation (SpO₂) defined as the proportion of oxygenated hemoglobin in total hemoglobin measured by pulse oximetry was recorded at 1-min resolution using a Garmin Forerunner 955 wrist-worn photoplethysmography sensor (Garmin International, Inc., Olathe, KS, USA).

Biomechanical and gait data acquisition

Biomechanical data were recorded at 500 Hz using a Stryd foot-mounted tri-axial inertial measurement unit (IMU) power meter (Stryd, Inc., Boulder, CO, USA). Derived gait variables included cadence (steps/min), stride length (m), ground contact time (ms), flight time (ms), vertical oscillation (cm), leg-spring stiffness (kN/m), and root mean square (RMS) acceleration along each axis. Running speed and distance were cross-validated against measurements from a Global Positioning System (GPS)-enabled Garmin Forerunner 955 watch (Garmin International, Inc., Olathe, KS, USA). When GPS drift exceeded 3%, distance was corrected using stride-integrated displacement. Ambient temperature (°C), relative humidity (%), and terrain grade (%) were synchronized at 1-min resolution using the watch's environmental sensors and, for mass-participation races, public meteorological records.

Reference energy expenditure measurement

Reference-standard EE was obtained from a calibration subset of 62 runners selected from the full cohort of 186 participants. This subset comprised 43 males aged 22 – 55 years with the mean age of 38.6 ± 8.9 years and 19 females aged 21 – 53 years with the mean age of 36.2 ± 9.7 years, recruited from the same affiliated running clubs and university sports medicine outpatient clinic. Each participant completed an incremental treadmill test in the laboratory, consisting of 3 minutes stages with 1 km/h speed increments, while breath-by-breath indirect COSMED K5 calorimetry was applied (COSMED Srl, Rome,

Italy). EE (kcal/min) was calculated from oxygen consumption (VO₂) and carbon dioxide production (VCO₂) using the Weir equation [24] as follows.

$$EE = 3.941 \times VO_2 \text{ (L/min)} + 1.106 \times VCO_2 \text{ (L/min)}$$

To extend the laboratory reference to field conditions, an individualized HR–VO₂ procedure was applied. Stage-level HR/VO₂ regression coefficients for each participant were applied to field HR traces to generate minute-level EE estimates as training targets for the ML models. This combined laboratory and field labeling strategy was consistent with current recommendations for scalable reference generation in wearable EE studies [7, 9].

Data preprocessing and quality control

All raw data streams were temporally aligned to a common 1 Hz sampling rate by linear resampling. Samples more than 4 standard deviations from a rolling 30 s mean were classified as outliers and imputed using cubic-spline interpolation. Sessions with more than 5% missing data after imputation were excluded. Motion-related HR artifacts were identified using a signal-quality index that combined RR-interval variance and acceleration magnitude. Contaminated epochs were masked and interpolated. Continuous variables were z-standardized within each participant to reduce between-participant scale differences while preserving within-session dynamics, an approach previously shown to improve generalization in cross-subject models [10].

Multisource feature engineering

A total of 174 candidate features were derived from the preprocessed data and grouped into four domains including cardiovascular, biomechanical, anthropometric and baseline fitness, and contextual. Cardiovascular features included instantaneous HR, the HR-to-speed ratio, SDNN, RMSSD, the LF/HF ratio, and SpO₂. Biomechanical features included cadence, stride length, ground contact time, vertical oscillation, leg stiffness, and the RMS and spectral entropy of

tri-axial acceleration. Anthropometric and baseline-fitness features included age, sex, BMI, body fat percentage, estimated VO_2max , and lactate-threshold pace. Contextual features included cumulative distance, elapsed session time, ambient temperature, humidity, and terrain grade. Feature selection was performed in two stages. A mutual-information filter removed features below the 20th percentile of ranked mutual information, while recursive feature elimination (RFE) with a random-forest base estimator retained the 48 features with the greatest cross-validated contribution to EE prediction. This filter-plus-wrapper approach was consistent with recent practice in endurance-performance modeling [25].

Baseline machine learning models

Four baseline regression models were trained using the 48 retained features. L2-regularized linear regression (LR) served as an interpretable reference model. Support vector regression (SVR) used a radial basis function kernel. Random forest (RF) comprised 500 trees with unrestricted depth. Extreme gradient boosting (XGBoost) incorporated early stopping based on a held-out validation fold. Hyperparameters were optimized over 100 iterations using Bayesian optimization within an inner 5-fold cross-validation loop, a strategy previously applied in sports-analytics research [26, 27].

Multisource feature fusion network (MSFF-Net)

The proposed MSFF-Net processed the four feature domains through parallel encoder branches before fusing their outputs at a higher representational level. Each branch comprised a one-dimensional convolutional layer with kernel size of 5 and 32 filters followed by a bidirectional long short-term memory layer with 64 units per direction to capture within-domain temporal dynamics. The branch outputs were concatenated and passed to a multi-head self-attention block with 4 heads at head dimension of 32, which modeled cross-domain dependencies such as the time-varying relationship between HR drift and cadence decline during the later stages of a long run. The

fused representation was decoded by two fully connected layers containing 128 and 64 units, respectively, with rectified linear unit (ReLU) activation and a final linear layer that generated a scalar EE estimate. The network was trained by minimizing Huber loss with the Adam optimizer with initial learning rate of 1×10^{-3} and cosine decay, a batch size of 128, early stopping with a patience of 10 epochs, and dropout of 0.3 in the dense layers. All deep-learning components were implemented in PyTorch 2.1 (<https://pytorch.org>) and trained on an NVIDIA RTX A5000 graphics processing unit.

Pacing optimization framework

Using the best-performing EE model as an internal simulator, pacing optimization was formulated as a finite-horizon constrained optimization problem. The racecourse was divided into equal 1 km segments with prescribed running speed as the decision variable for each segment to minimize total finishing time subject to three constraints including cumulative EE at each checkpoint remained below the participant-specific carbohydrate and fat oxidative capacity estimated from laboratory testing, HR reserve remained below 92% of the individual maximum after the first 10 km, and changes in speed between adjacent segments were limited to $\pm 4\%$ to represent feasible within-race adjustments. The optimization problem was solved by dynamic programming with backward induction over a discretized speed grid. The resulting pacing profile was compared segment by segment with each athlete's actual race performance.

Model evaluation and statistical analysis

Predictive accuracy was evaluated using subject-stratified 10-fold cross-validation and repeated five times, yielding 50 performance estimates for each model. The reported metrics were mean absolute error (MAE), root mean squared error (RMSE), and the coefficient of determination (R^2). Differences among models were assessed using repeated-measures analysis of variance (ANOVA) followed by Tukey's honestly significant difference post hoc test. The effect sizes were reported as Cohen's *d*. Agreement between

predicted and measured EE was examined using Bland-Altman analysis including bias and 95% limits of agreement. Feature-level interpretability was evaluated by applying SHAP to the dense decoder of the fusion model [28]. For the pacing analysis, paired tests were used to compare actual and optimized pace deviation, finishing time, and end-point rating of perceived exertion (RPE) with Borg of 6 – 20 scale. Wilcoxon signed-rank tests were applied when normality assumptions were not met. All tests were two-sided with $\alpha = 0.05$. Analyses were performed under Python 3.11 using scikit-learn 1.4 (<https://scikit-learn.org>), XGBoost 2.0 (<https://xgboost.readthedocs.io>), PyTorch 2.1 (<https://pytorch.org>), and SHAP 0.44 (<https://shap.readthedocs.io>) in R 4.3 (<https://www.r-project.org>).

Results

Participant characteristics

Among 186 runners, males constituted 68.3% of the cohort, which primarily comprised experienced recreational runners with median training history of 6 years and interquartile range of 4 – 9 years. The mean height was 172.5 ± 8.3 cm with males of 176.9 ± 6.1 cm and females of 163.2 ± 5.9 cm. The mean body mass was 67.2 ± 10.1 kg with males of 71.4 ± 8.7 kg and females of 58.3 ± 7.2 kg, while the mean BMI was 22.5 ± 2.1 kg/m² with males of 22.8 ± 2.0 kg/m² and females of 21.9 ± 2.2 kg/m². The mean body fat percentage was $17.4 \pm 5.2\%$ with males of $14.6 \pm 3.8\%$ and females of $23.5 \pm 4.1\%$, and the mean training history was 6.8 ± 4.0 years with males of 7.1 ± 4.0 years and females of 6.1 ± 3.8 years. The mean laboratory VO_2max was 51.7 ± 6.3 mL/kg·min overall with males of 54.2 ± 5.4 mL/kg·min and females of 46.3 ± 5.1 mL/kg·min, consistent with values reported for trained endurance populations [4, 21]. The mean lactate-threshold pace was $4:32 \pm 0:28$ min/km with males of $4:21 \pm 0:24$ min/km and females of $4:55 \pm 0:22$ min/km, and the mean personal-best marathon time was $3:24 \pm 0:31$ h:min with males of $3:16 \pm 0:27$ h:min and females of $3:41 \pm 0:33$

h:min. Each runner contributed a mean of 9.9 ± 3.2 sessions with males of 10.1 ± 3.1 sessions and females of 9.5 ± 3.4 sessions. Although females represented a smaller proportion of the cohort, the number of long-distance sessions contributed per participant was comparable between genders, supporting balanced exposure in the cross-validation partitions.

Predictive performance across models

Repeated subject-stratified 10-fold cross-validation (5 repeats, 50 folds) produced a consistent ranking across all evaluation metrics. MSFF-Net achieved the lowest errors with MAE of 0.61 ± 0.08 kcal/min and RMSE of 0.82 ± 0.09 kcal/min and the highest explained variance with R^2 of 0.94 ± 0.01 . Performance was next best for XGBoost with MAE of 0.78 ± 0.09 kcal/min, RMSE of 1.03 ± 0.10 kcal/min, and R^2 of 0.90 ± 0.02 followed by RF with MAE of 0.94 ± 0.10 kcal/min, RMSE of 1.22 ± 0.11 kcal/min, and R^2 of 0.86 ± 0.02 , while SVR showed MAE of 1.22 ± 0.12 kcal/min, RMSE of 1.58 ± 0.14 kcal/min, R^2 of 0.79 ± 0.03 , and LR showed MAE of 1.48 ± 0.14 kcal/min, RMSE of 1.92 ± 0.17 kcal/min, R^2 of 0.71 ± 0.03 . Differences among models were statistically significant ($P < 0.001$). Post hoc comparisons showed that MSFF-Net outperformed XGBoost (Cohen's $d = 1.63$) and LR ($d = 6.45$). The fold-level distributions using combined violin and box plots demonstrated that, relative to XGBoost, MSFF-Net reduced MAE by 21.8% and increased R^2 by 0.04 in the strongest conventional baseline (Figure 1).

Feature importance and interpretability

SHAP analysis of the MSFF-Net decoder identified instantaneous HR and running speed as the two most influential predictors of EE. Together, they accounted for 38.2% of the total mean absolute SHAP magnitude. HRV, represented by RMSSD, was the third most influential feature, consistent with its physiological relationship to autonomic load during prolonged exercise [22]. Vertical-acceleration RMS and cadence were the leading biomechanical predictors, supporting meta-analytic evidence that these kinematic variables were consistently associated with running

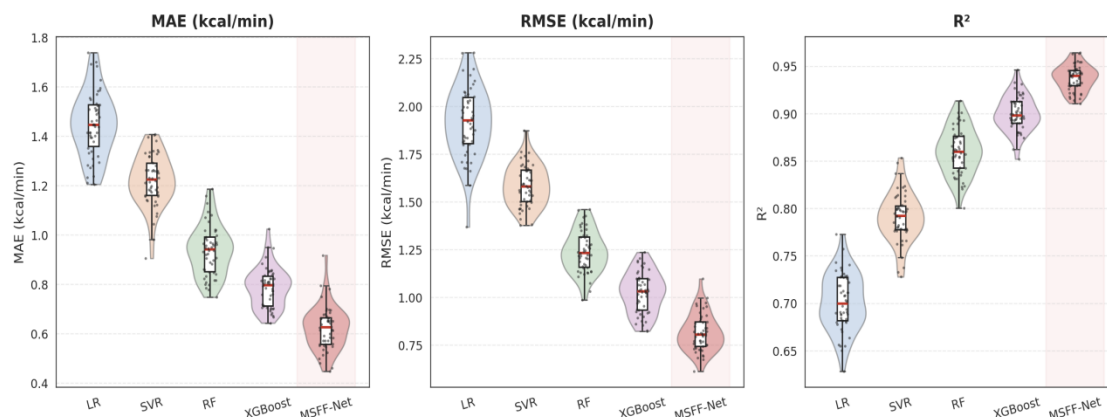


Figure 1. Distribution of 10-fold cross-validated performance for the five candidate models. Violin bodies showed the density of 50 fold-level estimates. White boxes displayed the interquartile range with red median lines. Individual dots were single folds. The proposed MSFF-Net (highlighted) achieved the lowest errors and the highest R^2 .

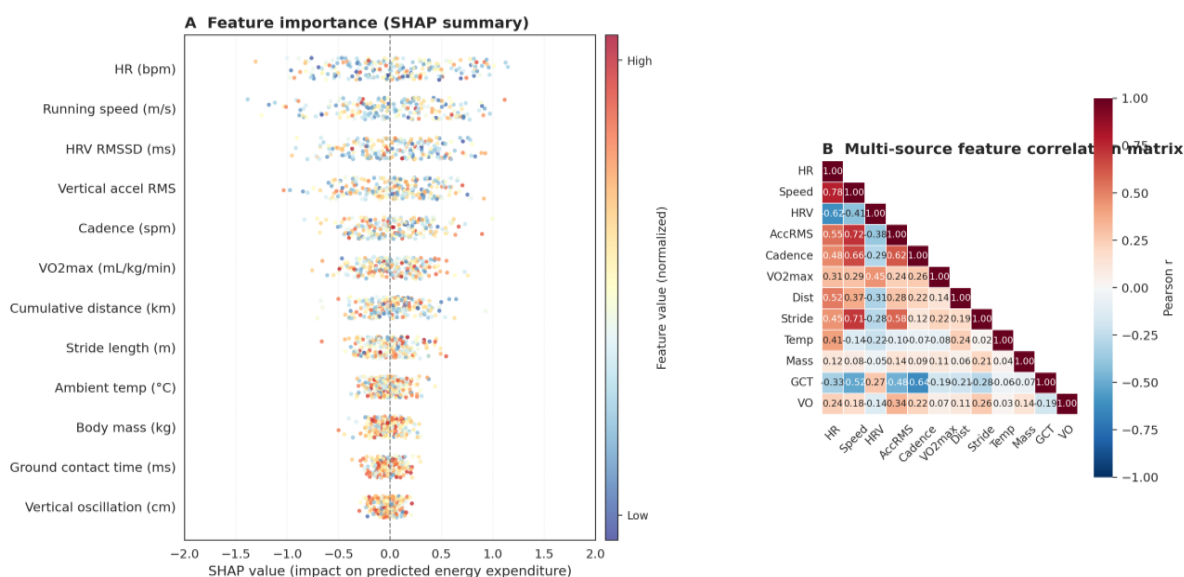


Figure 2. Interpretability of the MSFF-Net. **A.** SHAP summary plot of the twelve most influential features with color indicating normalized feature value. **B.** Lower-triangular Pearson correlation matrix of the same feature set.

economy [29]. The SHAP beeswarm plot showed that higher HR and speed values consistently increased predicted EE, whereas higher HRV values reduced the prediction (Figure 2A). The correlation matrix for the 12 displayed features showed strong correlations between speed and stride length ($r = 0.71$) and between speed and acceleration RMS ($r = 0.72$), as well as the expected inverse correlation between HRV and HR ($r = -0.62$) (Figure 2B).

Pacing optimization performance

The model-guided pacing optimizer was applied to 47 long-distance races. Comparison of the prescribed and actual paces showed that athletes began the first 5 km at 8.1 ± 6.1 s/km faster than the model-recommended target, whereas the optimized deviation was -2.0 ± 3.1 s/km ($P < 0.001$). From 15 km onward, actual pace increasingly lagged behind the target. In the 15 – 20 km segment, the actual pace deviation was $+5.3 \pm 7.2$ s/km, compared with $+1.2 \pm 3.1$ s/km

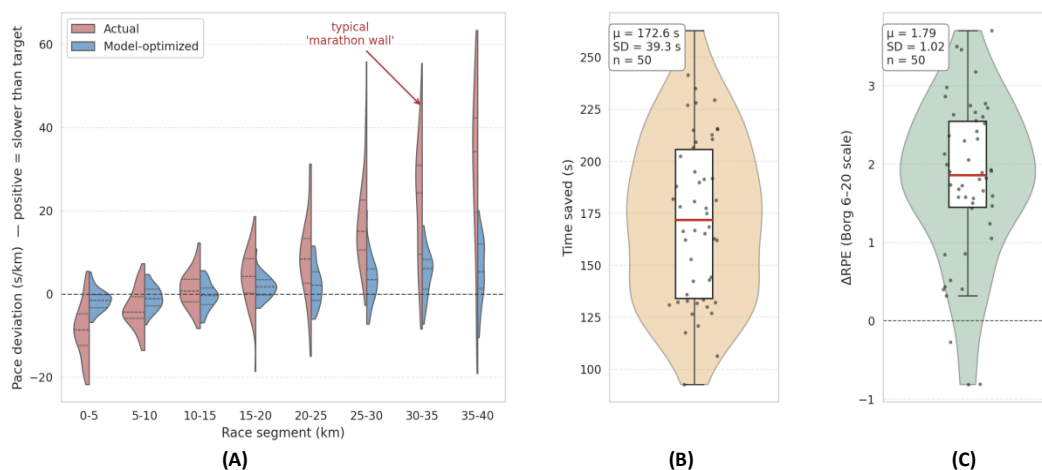


Figure 3. Split-violin plot of segment-wise pace deviation from the model-optimal target under actual versus optimized pacing (A). Finishing-time gain (B) and end-of-race RPE reduction (C) achieved under model-guided pacing ($n = 47$).

under the optimized strategy ($P < 0.001$). The largest differences occurred in the “marathon wall” zone. In the 25 – 30 km segment, actual and optimized pace deviations were $+16.8 \pm 12.1$ and $+3.4 \pm 5.0$ s/km, respectively, with a concurrent HR change of -4.6 ± 4.1 bpm ($P < 0.001$). In the 30 – 35 km segment, the corresponding deviations were $+24.2 \pm 15.3$ and $+5.1 \pm 6.1$ s/km with an HR change of -6.8 ± 4.6 bpm ($P < 0.001$). In the final 35 – 40 km segment, the actual deviation reached $+30.6 \pm 18.4$ s/km, compared with $+7.2 \pm 7.1$ s/km under the optimized condition (Δ HR of -8.1 ± 5.2 bpm) ($P < 0.001$). Simulated deviations under the optimized strategy remained closely centered on zero throughout the race. These patterns were illustrated using split-violin plots. The optimized distributions were less skewed than the actual distributions, whose positive tail widened progressively (Figure 3A). The optimized strategy was associated with a simulated reduction in finishing time of 172.6 ± 39.3 s or 2.88 ± 0.66 min ($P < 0.001$) (Figure 3B) and a reduction in end-of-race RPE of 1.79 ± 1.02 points on the Borg 6 – 20 scale ($P < 0.001$) (Figure 3C).

External validation of energy expenditure prediction

On an independent hold-out set comprising 340 session-level records from 38 runners who were not included in model training, MSFF-Net showed

strong agreement with the reference EE values. The correlation between predicted and measured EE showed $r = 0.985$ ($R^2 = 0.970$), and the regression relationship was close to the line of identity ($y = 1.04x - 0.50$). Bland-Altman analysis yielded a bias of -0.058 kcal/min and 95% limits of agreement from -1.24 to $+1.13$ kcal/min, which were narrower than those reported for commercial smartwatches in comparable populations [5]. Residuals were centered near zero, unimodal, and largely confined within ± 1 kcal/min, whereas residuals from the baseline linear regression model were more dispersed and had heavier tails (Figure 4). The results supported the generalizability of the multisource fusion model to unseen participants.

Discussion

This study found that integrating cardiovascular, biomechanical, anthropometric, and environmental features within a hybrid convolutional-recurrent-attentional architecture improved minute-level EE prediction in long-distance running relative to conventional ML models. The resulting predictor supported a dynamic-programming pacing optimizer that, in simulation, reduced finishing time and end-of-race perceived exertion. MSFF-Net reduced MAE by 21.8% relative to XGBoost and by 58.8%

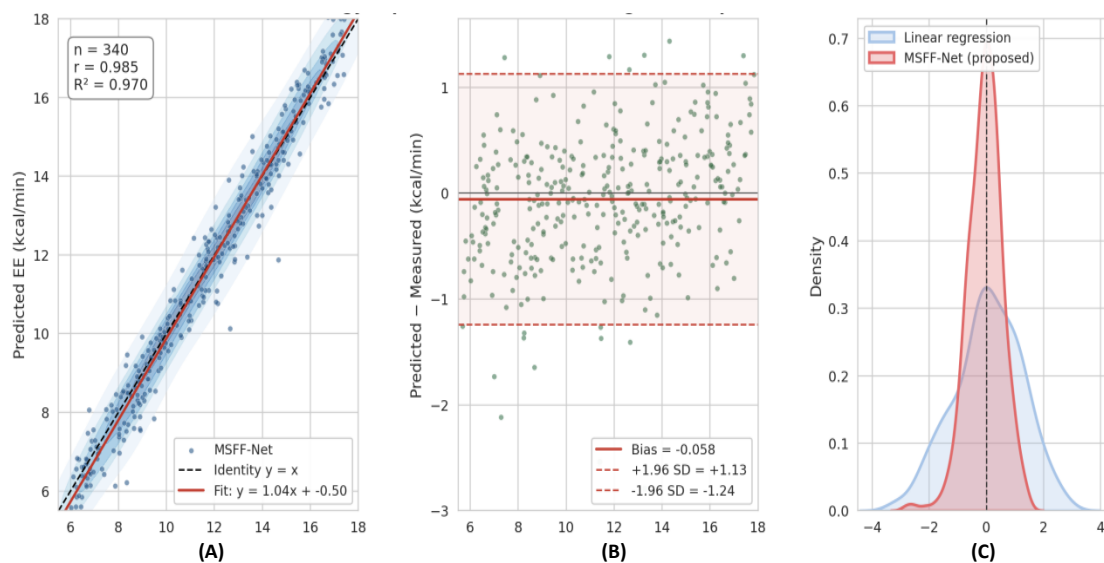


Figure 4. External validation of the MSFF-Net. **A.** Density scatter of predicted vs. measured EE with identity and fitted regression lines. **B.** Bland-Altman plot showing bias and 95 % limits of agreement. **C.** Residual-density comparison against a linear-regression baseline.

relative to linear regression, while producing narrower Bland-Altman limits of agreement than those commonly reported for commercial wearables evaluated against indirect calorimetry [5, 6]. The consistency of these improvements across three performance metrics and five repeated cross-validation runs suggested that the observed advantage was robust to data partitioning. The performance of MSFF-Net was physiologically plausible. At rest and at lower running speeds, accelerometer-derived features could characterize mechanical work because metabolic demand was scaled with external power. As exercise intensity approached the lactate threshold, oxygen uptake kinetics exhibited non-linear characteristics. SHAP analysis of data from incremental treadmill tests by Song *et al.* demonstrated that cardiovascular responses provided increasingly important information during this process [10]. The attention-based fusion block could learn this intensity-dependent reweighting of modalities, and the SHAP results indicated that HR and HRV contributed increasingly to prediction as cumulative distance increased. Traditional machine learning pipelines that merely concatenated input features did not explicitly model such cross-domain interactions, which might partly explain why their performance

plateaued at an R^2 of approximately 0.90. These findings complemented two established areas of research. First, studies of running economy had shown that biomechanical variables explained approximately 4 – 12% of the between-individual variation in oxygen cost [29]. The present findings confirmed the relevance of vertical oscillation, cadence, and stride length and suggested that their predictive value increased when they were integrated with intensity-sensitive cardiovascular signals in a learned representation. Second, large-scale studies of pacing consistently showed that runners who started too aggressively incur disproportionate performance losses during the final third of a race [11, 12]. The dynamic-programming optimizer formalized this principle by generating a pacing profile that approximated a modest negative split, consistent with patterns observed among elite marathon runners [16]. In simulation, the model-guided strategy reduced pace deviation in the 25 – 35 km “marathon wall” zone by approximately 80%. This improvement was greater than that reported in simulation studies that used variational autoencoders solely to anticipate glycogen-related performance collapse [17], possibly because the present framework combined prediction with prescriptive pacing guidance. The 1.79 point

decrease in end-of-race RPE might also be practically meaningful, corresponding approximately to a shift from “very hard” to “hard” on the Borg 6–20 scale. Previous research had associated lower exertional strain with reduced risks of post-race injury, dehydration, and syncope [15, 18]. The mean reduction in finishing time was approximately 2%, smaller in absolute terms than the 11 min improvement reported after polarized training interventions [16], but notable because it was achieved in simulation without changes to training volume, intensity distribution, or equipment. Interpretability was an explicit priority of the study design. Previous XGBoost-based athlete performance models showed that SHAP explanations could identify actionable features with clear physiological relevance to coaches and practitioners [26, 27]. In the present analysis, the prominence of the HR-to-speed ratio and HRV (RMSSD) was consistent with recommendations to integrate short-term markers of autonomic balance into real-time training-load monitoring [23, 30]. The feature ranking also included potentially modifiable gait variables of cadence, vertical oscillation, and leg stiffness, thereby linking model outputs to possible gait-retraining interventions.

Several limitations should be considered. The retrospective design of this research remained susceptible to residual confounding despite the stringent eligibility criteria and did not permit causal inference. Reference EE values were generated by extrapolating individualized HR–VO₂ regressions from laboratory tests to field recordings, which was subject to error, particularly during the final hour of prolonged running, when cardiovascular drift and dehydration could alter the HR–VO₂ relationship [4]. Although the cohort covered a broad range of fitness levels, it consisted primarily of recreational runners and included more males than females. Separate validation was therefore required in elite athletes and ultra-endurance populations. The pacing optimizer was evaluated in silico against actual race performance rather than in a prospective field trial. The feasibility of

maintaining closely prescribed real-time paces, particularly in mass-participation events affected by pack dynamics, remained uncertain. Further, although environmental variables were included, the model did not explicitly incorporate the full course-elevation profile or nutritional intake during races, both of which might influence late-race fatigue. Future research should prioritize prospective validation and real-time deployment on wearable or mobile devices. Incorporating sweat-based lactate sensing and continuous glucose monitoring may further improve EE prediction during the late stages of a race where the current model showed its largest residuals. The attention-based fusion framework was not specific to running and should also be evaluated in cycling, triathlon, and team-sport settings, particularly in comparison with emerging hybrid Transformer–LSTM models of athlete performance.

This retrospective study showed that the proposed MSFF-Net predicted minute-level EE with an MAE of 0.61 kcal/min and explained 94% of the variance, outperforming four conventional ML models. When coupled with a dynamic-programming optimizer, the model generated a simulated pacing strategy that reduced finishing time and end-of-race RPE, while attenuating late-race deceleration in the marathon-wall zone. SHAP-based interpretation identified HR, running speed, HRV, and gait kinematics as important contributors to predictions and provided physiologically interpretable information for coaches and athletes. Prospective, on-device validation was required before the framework could be translated into individualized real-time pacing guidance.

Acknowledgements

This work was supported by the Henan Department of Science and Technology (Grant No. 262102320291).

References

1. Muñoz-Pérez I, Varela-Sanz A, Lago-Fuentes C, Mecías-Calvo M. 2024. Predictive performance models in marathon based on half-marathon, age group and pacing behavior. *Sport Sci Health*. 20(3):797-810.
2. Sha J, Yi Q, Jiang X, Wang Z, Cao H, Jiang S. 2024. Pacing strategies in marathons: A systematic review. *Heliyon*. 10(17):e36760.
3. Keogh A, Smyth B, Caulfield B, Lawlor A, Berndsen J, Doherty C. 2019. Prediction equations for marathon performance: A systematic review. *Int J Sports Physiol Perform*. 14(9):1159-1169.
4. Unhjem RJ, Nordrehaug AE, Øfsteng SJ, Støa EM, Sandbakk Ø, Rønnestad BR. 2024. Changes in running economy and attainable maximal oxygen consumption in response to prolonged running: The impact of training status. *Scand J Med Sci Sports*. 34(5):e14637.
5. Carrier B, Helm MM, Cruz K, Barrios B, Navalta JW. 2023. Validation of aerobic capacity (VO₂max) and lactate threshold in wearable technology for athletic populations. *Technologies*. 11(3):71.
6. Caserman P, Yum S, Göbel S, Reif A, Matura S. 2024. Assessing the accuracy of smartwatch-based estimation of maximum oxygen uptake using the Apple Watch Series 7: Validation study. *JMIR Biomed Eng*. 9:e59459.
7. O'Driscoll R, Turicchi J, Hopkins M, Gibbons C, Larsen SC, Palmeira AL, *et al*. 2021. Comparison of the validity and generalizability of machine learning algorithms for the prediction of energy expenditure: validation study. *JMIR Mhealth Uhealth*. 9(8):e23938.
8. O'Driscoll R, Turicchi J, Hopkins M, Horgan GW, Finlayson G, Stubbs JR. 2020. Improving energy expenditure estimates from wearable devices: A machine learning approach. *J Sports Sci*. 38(13):1496-1505.
9. Liu F, Wang H, Zang W, Li Y, Sun F. 2024. Influencing factors mining and modeling of energy expenditure in running based on wearable sensors. In: *Proceedings of the 2024 International Conference on Sports Technology and Performance Analysis*. New York: ACM. Pp. 1-8.
10. Song Y, Wang Z, Wang H, Sun G, Gong B, Zhang F. 2025. Interpretable deep learning for personalized energy expenditure prediction using ECG and acceleration signals in incremental exercise. *Sci Rep*. 15(1):36277.
11. Lerebourg L, Saboul D, Cléménçon M, Coquart JB. 2023. Prediction of marathon performance using artificial intelligence. *Int J Sports Med*. 44(5):352-360.
12. Thuany M, Weiss K, Villiger E, Scheer V, Gomes TN, Cuk I, *et al*. 2025. Pacing in New York City Marathon: A tracking approach. *Sport Sci Health*. 21(4):3349-3358.
13. Miyazaki Y, Matsui H, Zushi K, Fukui T. 2025. Early marathon running metrics from inertial measurement units predict significant pace reduction. *Front Sports Act Living*. 7:1681444.
14. Allik A, Pilt K, Viigimäe M, Fridolin I, Jervan G. 2022. A novel physical fatigue assessment method utilizing heart rate variability and pulse arrival time towards personalized feedback with wearable sensors. *Sensors*. 22(4):1680.
15. Kim JE, Kim NH, Choi SK, Lee JY, Lee K, Han JS. 2025. Machine learning-based fatigue classification using heart rate variability and cortisol: A multimodal approach to wearable health monitoring. *Digit Health*. 11:20552076251395570.
16. Qin G, Lee S, Kim S. 2025. Machine learning-based personalized training models for optimizing marathon performance through pyramidal and polarized training intensity distributions. *Sci Rep*. 15(1):41516.
17. El Dandachi MM, Billat V, Palacin F, Vigneron V. 2025. Early detection of the marathon wall to improve pacing strategies in recreational marathoners. *AI*. 6(6):130.
18. Barreira J, Marques EA, Nakamura FY, Brito J, Figueiredo P. 2025. Exploring the physiological mechanisms of sleep's influence on athletic performance and recovery: A narrative review. *Sleep Breath*. 29(6):354.
19. Dash S. 2024. Win your race goal: A generalized approach to prediction of running performance. *Sports Med Int Open*. 8:a24016234.
20. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, *et al*. 2024. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 385:e078378.
21. Wiecha S, Kasiak PS, Cieśliński I, Maciejczyk M, Mamcarz A, Śliż D. 2022. Modeling physiological predictors of running velocity for endurance athletes. *J Clin Med*. 11(22):6688.
22. Grosicki GJ, Carter JR, Laursen PB, Plews D, Altini M, Galpin A, *et al*. 2026. Heart rate variability coefficient of variation during sleep as a digital biomarker that reflects behavior and varies by age and sex. *Am J Physiol Heart Circ Physiol*. 330(1):H187-H199.
23. Esco MR, Fields AD, Mohammadnabi MA, Kliszczewicz BM. 2025. Monitoring training adaptation and recovery status in athletes using heart rate variability *via* mobile devices: A narrative review. *Sensors*. 26(1):3.
24. Weir JB de V. 1949. New methods for calculating metabolic rate with special reference to protein metabolism. *J Physiol*. 109(1-2):1-9.
25. Peyré-Tartaruga LA, Machado E, Guimarães P, Borba E, Tartaruga MP, Buzzachera CF, *et al*. 2024. Biomechanical, physiological and anthropometrical predictors of performance in recreational runners. *Peer J*. 12:e16940.
26. Ouyang Y, Li X, Zhou W, Hong W, Zheng W, Qi F, *et al*. 2024. Integration of machine learning XGBoost and SHAP models for NBA game outcome prediction and quantitative analysis methodology. *PLoS One*. 19(7):e0307478.
27. Chen Y. 2025. Hybrid Transformer-LSTM model for athlete performance prediction in sports training management. *Informatica*. 49(2):185-196.
28. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, *et al*. 2020. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2(1):56-67.
29. Van Hooren B, Jukic I, Cox M, Frenken KG, Bautista I, Moore IS. 2024. The relationship between running biomechanics and running economy: A systematic review and meta-analysis of observational studies. *Sports Med*. 54(5):1269-1316.
30. Li K, Cardoso C, Moctezuma-Ramirez A, Elgalad A, Perin E. 2023. Heart rate variability measurement through a smart wearable device: Another breakthrough for personal health monitoring? *Int J Environ Res Public Health*. 20(24):7146.